PAPER *Special Section on Technology Trials and Proof-of-Concept Activities for 5G and Beyond*

Channel Estimation and Achievable Rate of Massive MU-MIMO Systems with IQ Imbalance*

Nana ZHANG[†], **Huarui YIN**[†], **Weidong WANG**[†], ***Nonmembers, and*** **Suhua TANG**[††a)], ***Member***

SUMMARY In-phase and quadrature-phase imbalance (IQI) at transceivers is one of the serious hardware impairments degrading system performance. In this paper, we study the overall performance of massive multi-user multi-input multi-output (MU-MIMO) systems with IQI at both the base station (BS) and user equipments (UEs), including the estimation of channel state information, required at the BS for the precoding design. We also adopt a widely-linear precoding based on the real-valued channel model to make better use of the image components of the received signal created by IQI. Of particular importance, we propose estimators of the real-valued channel and derive the closed-form expression of the achievable downlink rate. Both the analytical and simulation results show that IQI at the UEs limits the dowlink rate to finite ceilings even when an infinite number of BS antennas is available, and the results also prove that the widely-linear precoding based on the proposed channel estimation method can improve the overall performance of massive MU-MIMO systems with IQI.

key words: *massive MIMO, in-phase and quadrature-phase imbalance (IQI), achievable rate, channel estimation*

1. Introduction

Massive multi-input multi-output (MIMO) systems can achieve high spectral and energy efficiency [1], [2], and simple linear precoding approaches such as zero-forcing (ZF) precoding and matched-filter (MF) precoding are available [2]–[4]. It is a promising technique for future wireless communication systems [5], [6]. The success of massive MIMO systems relies heavily on the availability of low-cost and power-efficient hardware, due to the very large array dimensions at the base station (BS). However, the low-cost radio frequency (RF) components cause more serious hardware impairments than the conventional circuits of RF chains. Thus, the hardware impairments at transceivers should be taken into consideration when behaviors of massive MIMO systems are investigated.

Recently, massive MIMO systems with hardware impairments have been studied in many works [7]–[9]. These works use additive or multiplicative stochastic models to characterize hardware impairments, and show that the existence of hardware impairments limits the channel estimation accuracy, the achievable rate of the system, and so on. Direct-conversion has been a major scheme to achieve simplified and low-cost transceiver design [10]. In-phase and quadrature-phase imbalance (IQI) caused by the limited accuracy of analog hardware as well as temperature variation and aging, is one important factor of hardware impairments closely associated with direct-conversion. Previous works have proved that IQI leads to performance degradation, since it twists the modulation constellations directly [11]. Consequently, user experience is deteriorated due to IQI [12].

To design a precoding matrix, channel state information is required at the BS. Most massive MIMO systems use the time-division duplex (TDD) mode and acquire the channel state information in the uplink. For the uplink of massive MIMO systems with IQI, some methods to estimate the uplink equivalent channel, and to eliminate IQI influence have been proposed [12]–[15]. A pilot-based joint estimator of the channel and a low-complexity IQI compensation scheme based on IQI coefficients estimation were proposed in [13]. The RF-aware digital beamforming method developed in [12], named the widely-linear minimum variance distortionless response beamformer, can provide efficient beamsteering and interference suppression. The low resolution problem is considered in [14], and it is assumed that a high precision IQI detector is adopted at the BS and that IQI parameters of the circuits of the transmitter (TX) and receiver (RX) are the same. In [12]–[14], only the IQI at the BS is considered. Behaviors of the achievable rate for uplink MU-MIMO system with IQI at both the BS and user equipments (UEs) were investigated in [15], which proposed an IQI aware widely-linear channel estimation and data detection scheme, while taking mirror signals as an interference for a complex model. It reveals that in the uplink the IQI at the UEs has more serious impact than the IQI at the BS.

However, IQI leads to mismatch between the uplink and downlink channels. More attention needs to be paid to the impact of IQI on the downlink of massive MIMO, which is affected by TX IQI at the BS and RX IQI at the UEs, as well as TX IQI at the UEs and RX IQI at the BS. The downlink of the massive MIMO systems with IQI was investigated in [16]–[20]. The effective channel estimation and the sum rate analysis for the downlink were developed in [16], [17], which treat the image signal (i.e., the conjugate of the original signal for single carrier systems) as interference. Actually, the

Manuscript received September 1, 2018.
Manuscript revised December 18, 2018.
Manuscript publicized February 20, 2019.

[†]The authors are with the Dept. of Electronic and Information Science, University of Science and Technology of China, Hefei, Anhui, China.
[††]The author is with the Dept. of Computer and Network Engineering, University of Electro-Communications, Tokyo, 182-8585 Japan.

*This work is supported by National Science and Technology Major Project of China MIIT (Grant No. 2018ZX03001001-003) and National Natural Science Foundation of China (Grant No. 61571412).

a) E-mail: shtang@uec.ac.jp

DOI: 10.1587/transcom.2018TTP0001

Copyright © 2019 The Institute of Electronics, Information and Communication Engineers

image signal still contains the transmitted information. IQI only at the BS is studied in [18], where an IQI-aware widely-linear regularized ZF precoding method was proposed to reduce the impact of IQI. A compensation scheme of channel reciprocity was proposed in [19], and some widely-linear precoding methods were proposed for massive MU-MIMO systems with IQI in [20], assuming perfect channel information of the downlink. The channel estimation is affected by IQI, and the estimation of the downlink equivalent channel is hard and very costly due to the channel mismatch between the uplink and the downlink. However, the downlink of massive MIMO systems with IQI at both the BS and UEs has not been well investigated yet.

Motivated by aforementioned discussion, we focus on the downlink of massive MIMO systems with IQI at both the BS and UEs, and we adopt the widely-linear precoding method as it makes good use of image signals. The main contributions of this paper are summarized as follows:

- We investigate the impact of IQI on the estimation of the uplink real-valued equivalent channel with the least squares (LS) criterion. We adopt training sequences with a length that is twice the number of users, which can achieve a similar performance as in the hardware without IQI [17].
- We propose a joint estimation method for the wireless channel and IQI parameters to eliminate the impact of the TX IQI at the UEs and RX IQI at the BS in the uplink. The wireless channel and the parameters of RX IQI at the BS and TX IQI at the UEs are estimated in the proposed algorithm.
- We evaluate the achievable rate of the downlink system with widely-linear precoding including the estimated channel when IQI exists. Results show that IQI at the UEs limits the rate to a finite ceiling even when a large number of BS antennas is available. Results also reveal that the precoding based on the proposed channel estimation methods can improve the system performance. Moreover, for IQI at the BS, the performance is more sensitive to the amplitude imbalance, and for IQI at the UEs, the performance is more sensitive to the phase imbalance.

The remaining part of this paper is organized as follows. The augmented real-valued system model is presented in Sect. 2. Section 3 provides the uplink equivalent channel estimation and the joint estimation of wireless channel and IQI parameters. Section 4 derives the large-number-approximation expression of the achievable rate of systems with IQI at TX and RX. Simulations results are shown in Sect. 5, and finally Sect. 6 concludes this paper.

Notation: Let the superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^{-1}$, $\|\cdot\|_F$ denote the conjugate, transpose, inverse and Frobenius norm of a matrix, respectively. Let $[\mathbf{A}]_{s,:}$, $[\mathbf{A}]_{:,l}$ and $[\mathbf{A}]_{s,l}$ denote the sth row, the lth column and the element on the sth row and the lth column of the matrix $\mathbf{A}$, respectively. $\mathbf{I}_N$ and $\mathbf{0}_N$ denote the $N \times N$ identity matrix and zero matrix, respectively, and $\tilde{\mathbf{I}}_{2N}$ is a $2N$ by $2N$ block diagonal matrix in which the diagonal blocks are respectively $\mathbf{I}_N$ and $-\mathbf{I}_N$. diag$(\mathbf{a})$ returns a diagonal matrix with elements of the vector $\mathbf{a}$ as the diagonal elements. tr$(\cdot)$ and $\mathbb{E}\{\cdot\}$ stand for the trace and expectation operators, respectively. A complex Gaussian stochastic vector $\mathbf{x}$ is denoted as $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\theta},\Sigma)$, and a real Gaussian stochastic vector $\mathbf{x}$ is denoted $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\theta},\Sigma)$, where $\boldsymbol{\theta}$ and Σ are the the mean vector and the covariance matrix of vector $\mathbf{x}$, respectively.

2. System Model

Consider a single cell large-scale antenna systems with M single-antenna users and one base station equipped with N antennas, $N \gg M$. We assume that the system operates in the TDD mode, and the channel state information is acquired via uplink training sequences at the BS. Here we consider a single-carrier system, and we believe that the multi-carrier transmission, such as OFDM, can be analyzed similarly, which will be left as future work due to the limited space of the article.

The transmitted and received signals with IQI are generally expressed based on the primary signal and its conjugate (i.e., image signal) [17], [20]. In order to make the best use of the image component caused by IQI of the received signal, some augmented signal models have been represented in the related works [12], [20], [21]. Referring to these models, we reformulate the real-valued signal model for the received signal by separating it into signals' real and imaginary parts. For the sake of analysis simplicity, we present the signal model based on the equivalent channel matrices.

Firstly, the complex to real (C-R) mapping functions are given below

$$\mathbf{a} = \begin{bmatrix} \Re(\mathbf{a}_c) \\ \Im(\mathbf{a}_c) \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \Re(\mathbf{A}_c) & -\Im(\mathbf{A}_c) \\ \Im(\mathbf{A}_c) & \Re(\mathbf{A}_c) \end{bmatrix}, \tag{1}$$

where $\mathbf{a} \in \mathbb{R}^{2N\times 1}$ and $\mathbf{A} \in \mathbb{R}^{2N\times 2M}$ are the functions with respect to vector $\mathbf{a}_c \in \mathbb{C}^{N\times 1}$ and matrix $\mathbf{A}_c \in \mathbb{C}^{N\times M}$, respectively.

We assume the channel independence among different antennas, and adopt a Rayleigh fading channel model. Let $\mathbf{h}_{c,m} \in \mathbb{C}^{N\times 1}$ denote the complex channel response vector of the wireless propagation channel from the mth UE to the BS, whose entries are independent and identically distributed (i.i.d) random variables and follow $\mathcal{CN}(0,1)$. Let $\mathbf{H}_c = [\mathbf{h}_{c,1}, \cdots, \mathbf{h}_{c,M}]$ denote the channel matrix between the BS and UEs. Note that $\mathbf{H}_c$ only describe the wireless channel, while the equivalent channel contains the wireless channel and the transceiver RF channels.

In the downlink, let $d_{c,m}$ denote the symbol transmitted from the BS to the mth user, where $d_{c,m}$ for $1 \le m \le M$ are i.i.d random variables following $\mathcal{CN}(0, \frac{P_{\mathrm{BS}}}{M})$, and let $\mathbf{d}_c = [d_1, \cdots, d_M]^T \in \mathbb{C}^{M\times 1}$ denote the symbol vector transmitted by BS. The BS applies a precoding technique to $\mathbf{d}$ to yield $\mathbf{x} = \mathbf{W}\mathbf{d} \in \mathbb{R}^{2N\times 1}$, where $\mathbf{W} \in \mathbb{R}^{2N\times 2M}$ is the precoding matrix and $\mathbf{d} \in \mathbb{R}^{2M\times 1}$ is the C-R mapping of $\mathbf{d}_c$.

Denoting the C-R mapping of $\mathbf{H}_c$ as $\mathbf{H} \in \mathbb{R}^{2N\times 2M}$, the

vector $\mathbf{y}_{\mathrm{D}} \in \mathbb{R}^{2M\times 1}$ of real-valued received signals of all UEs in the downlink can be written as

$$\mathbf{y}_{\mathrm{D}} = \underbrace{\mathbf{K}_{\mathrm{D},r}\tilde{\mathbf{I}}_{2M}\mathbf{H}^T\tilde{\mathbf{I}}_{2N}\mathbf{K}_{\mathrm{D},t}}_{\mathbf{H}_{\mathrm{D,eq}}}\mathbf{x} + \mathbf{K}_{\mathrm{D},r}\mathbf{v}_{\mathrm{D}}, \tag{2}$$

where $\mathbf{H}_{\mathrm{D,eq}} \in \mathbb{R}^{2M\times 2N}$ denotes the equivalent channel of the downlink, which contains the RF channel at the transmitter, the wireless channel and the RF channel at the receiver. $\tilde{\mathbf{I}}_{2M}$ and $\tilde{\mathbf{I}}_{2N}$ are applied to ensure the transposition relationship of the complex wireless channels between the downlink and uplink. Here $\mathbf{v}_{\mathrm{D}} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I}_{2M})$ represents the real additive white Gaussian noise of $2M$ branches at all UEs, and $\mathbf{K}_{\mathrm{D},r} \in \mathbb{R}^{2M\times 2M}$, $\mathbf{K}_{\mathrm{D},t} \in \mathbb{R}^{2N\times 2N}$ stand for matrices of the RX IQI at the UEs and the TX IQI at the BS, which are defined as

$$\mathbf{K}_{\mathrm{D},t} = \begin{bmatrix} \mathbf{K}^1_{\mathrm{D},t} & \mathbf{K}^2_{\mathrm{D},t} \\ \mathbf{K}^3_{\mathrm{D},t} & \mathbf{K}^4_{\mathrm{D},t} \end{bmatrix}, \mathbf{K}_{\mathrm{D},r} = \begin{bmatrix} \mathbf{K}^1_{\mathrm{D},r} & \mathbf{K}^2_{\mathrm{D},r} \\ \mathbf{K}^3_{\mathrm{D},r} & \mathbf{K}^4_{\mathrm{D},r} \end{bmatrix}. \tag{3}$$

In (3), $\mathbf{K}^i_{\mathrm{D},t} \in \mathbb{R}^{N\times N}$ and $\mathbf{K}^i_{\mathrm{D},r} \in \mathbb{R}^{M\times M}$, $i = 1, 2, 3, 4$ are diagonal matrices, and their entries are given by

$$\begin{aligned}
[\mathbf{K}^1_{\mathrm{D},t}]_{nn} &= (1+\epsilon^{\mathrm{BS}}_{t,n})\cos(\theta^{\mathrm{BS}}_{t,n}/2)/\sqrt{1+(\epsilon^{\mathrm{BS}}_{t,n})^2}, \\
[\mathbf{K}^2_{\mathrm{D},t}]_{nn} &= (1-\epsilon^{\mathrm{BS}}_{t,n})\sin(\theta^{\mathrm{BS}}_{t,n}/2)/\sqrt{1+(\epsilon^{\mathrm{BS}}_{t,n})^2}, \\
[\mathbf{K}^3_{\mathrm{D},t}]_{nn} &= (1+\epsilon^{\mathrm{BS}}_{t,n})\sin(\theta^{\mathrm{BS}}_{t,n}/2)/\sqrt{1+(\epsilon^{\mathrm{BS}}_{t,n})^2}, \\
[\mathbf{K}^4_{\mathrm{D},t}]_{nn} &= (1-\epsilon^{\mathrm{BS}}_{t,n})\cos(\theta^{\mathrm{BS}}_{t,n}/2)/\sqrt{1+(\epsilon^{\mathrm{BS}}_{t,n})^2},
\end{aligned} \tag{4}$$

and

$$\begin{aligned}
[\mathbf{K}^1_{\mathrm{D},r}]_{mm} &= (1+\epsilon^{\mathrm{UE}}_{r,m})\cos(\theta^{\mathrm{UE}}_{r,m}/2)/\sqrt{(1+(\epsilon^{\mathrm{UE}}_{r,m})^2}, \\
[\mathbf{K}^2_{\mathrm{D},r}]_{mm} &= (1+\epsilon^{\mathrm{UE}}_{r,m})\sin(\theta^{\mathrm{UE}}_{r,m}/2)/\sqrt{(1+(\epsilon^{\mathrm{UE}}_{r,m})^2}, \\
[\mathbf{K}^3_{\mathrm{D},r}]_{mm} &= (1-\epsilon^{\mathrm{UE}}_{r,m})\sin(\theta^{\mathrm{UE}}_{r,m}/2)/\sqrt{(1+(\epsilon^{\mathrm{UE}}_{r,m})^2}, \\
[\mathbf{K}^4_{\mathrm{D},r}]_{mm} &= (1-\epsilon^{\mathrm{UE}}_{r,m})\cos(\theta^{\mathrm{UE}}_{r,m}/2)/\sqrt{(1+(\epsilon^{\mathrm{UE}}_{r,m})^2}.
\end{aligned} \tag{5}$$

The IQI model above is symmetric, and $1/\sqrt{1+\epsilon^{\cdot 2}}$ is applied to ensure that the signal power is same with the ideal branches case. Here $\epsilon^{\mathrm{UE}}_{r,m}$, $\theta^{\mathrm{UE}}_{r,m}$ represent the amplitude and the phase imbalance of RX RF chain at the mth UE, and $\epsilon^{\mathrm{BS}}_{t,n}$, $\theta^{\mathrm{BS}}_{t,n}$ denote the amplitude and the phase imbalance of the nth TX RF chain at the BS. In the ideal branches case where no IQI exists at the BS and UEs, parameters of amplitude and phase imbalance are equal to 0, i.e. $\mathbf{K}_{\mathrm{D},t} = \mathbf{I}_{2N}$ and $\mathbf{K}_{\mathrm{D},r} = \mathbf{I}_{2M}$.

In the uplink, let $\mathbf{s}_c = [s_1, \cdots, s_M]^T \in \mathbb{C}^{M\times 1}$ denote the transmitted data symbols of M UEs, and assume that each UE has the same transmit power P_{UE}, i.e. $\mathbb{E}\{|s_i|^2\} = P_{\mathrm{UE}}$. Denoting the C-R mapping of $\mathbf{s}_c$ as $\mathbf{s} \in \mathbb{R}^{2M\times 1}$, the real-valued received signal $\mathbf{y}_{\mathrm{U}} \in \mathbb{R}^{2N\times 1}$ in the uplink can be expressed as

$$\mathbf{y}_{\mathrm{U}} = \underbrace{\mathbf{K}_{\mathrm{U},r}\mathbf{H}\mathbf{K}_{\mathrm{U},t}}_{\mathbf{H}_{\mathrm{U,eq}}}\mathbf{s} + \mathbf{K}_{\mathrm{U},r}\mathbf{v}_{\mathrm{U}}, \tag{6}$$

where $\mathbf{H}_{\mathrm{U,eq}} \in \mathbb{R}^{2N\times 2M}$ denotes the uplink equivalent channel, and $\mathbf{v}_{\mathrm{U}} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I}_{2N})$ represents the real additive white Gaussian noise of all branches at the BS. Here $\epsilon^{\mathrm{BS}}_{r,n}$, $\theta^{\mathrm{BS}}_{r,n}$ and $\epsilon^{\mathrm{UE}}_{t,m}$, $\theta^{\mathrm{UE}}_{t,m}$ represent parameters of the RX IQI at the BS and the TX IQI at the UEs, respectively. Similar to (3), $\mathbf{K}^i_{\mathrm{U},r}$ and $\mathbf{K}^i_{\mathrm{U},t}$, $i = 1, 2, 3, 4$ are the block matrices of $\mathbf{K}_{\mathrm{U},r}$ and $\mathbf{K}_{\mathrm{U},t}$, which have the same forms as (4) and (5), except that $\epsilon^{\mathrm{BS}}_{r,n}$, $\theta^{\mathrm{BS}}_{r,n}$ and $\epsilon^{\mathrm{UE}}_{t,m}$, $\theta^{\mathrm{UE}}_{t,m}$ substitute $\epsilon^{\mathrm{UE}}_{r,m}$, $\theta^{\mathrm{UE}}_{r,m}$ and $\epsilon^{\mathrm{BS}}_{t,n}$, $\theta^{\mathrm{BS}}_{t,n}$, respectively.

Remark: By splitting each complex channel into I and Q real channels, the real-valued model can be treated as $2M \times 2N$ IQI-related MIMO. In the real-valued model, receivers can extract the transmit information from the image signal, which is treated as interference in the complex-value model. Using the image signal will help us to improve the performance of massive MIMO systems.

3. Channel Estimation

The uplink equivalent channel state information acquired by the BS is used to design precoding matrices for the downlink transmission. In this section, we provide an equivalent channel estimation scheme with the LS criterion. At first, IQI compensation is not considered. In order to mitigate the effect of the TX and RX IQI in the uplink, we further propose a joint estimation method for the wireless channel and IQI parameters.

Let $\mathbf{S}_{\mathrm{tr},c} \in \mathbb{C}^{M\times\tau}$ denote the matrix of training sequences, where τ is the length of the training sequences for each user. With $\mathbf{S}_{\mathrm{tr}} = [\Re(\mathbf{S}^T_{\mathrm{tr},c}) \quad \Im(\mathbf{S}^T_{\mathrm{tr},c})]^T$, the real-valued received signal $\mathbf{Y}_{\mathrm{U,tr}} \in \mathbb{R}^{2N\times\tau}$ for training sequences is written as

$$\mathbf{Y}_{\mathrm{U,tr}} = \mathbf{H}_{\mathrm{U,eq}}\mathbf{S}_{\mathrm{tr}} + \mathbf{K}_{\mathrm{U},r}\mathbf{V}_{\mathrm{U,tr}}, \tag{7}$$

where $\mathbf{V}_{\mathrm{U,tr}} \in \mathbb{R}^{2N\times\tau}$ represents the noise matrix, whose entries follow the distribution $\mathcal{N}(0, \frac{\sigma^2}{2})$.

3.1 Equivalent Channel Estimation

In this subsection, we assume that the TX and RX IQI is unknown. We derive the equivalent channel estimation method with the aid of training sequences and assume that $\tau \geq 2M$, since $2M$ equivalent channels need to be estimated. The LS estimation of equivalent channel $\mathbf{H}_{\mathrm{U,eq}}$ is expressed as

$$\hat{\mathbf{H}}_{\mathrm{U,eq}} = \mathbf{Y}_{\mathrm{U,tr}}\mathbf{S}^T_{\mathrm{tr}}(\mathbf{S}_{\mathrm{tr}}\mathbf{S}^T_{\mathrm{tr}})^{-1}. \tag{10}$$

The mean squared error (MSE) of the channel estimation results is given by

$$\begin{aligned}
\mathrm{MSE}_{\mathbf{H}_{\mathrm{U,eq}}} &= \mathbb{E}\{\|\mathbf{\Delta}_{\mathrm{eq}}\|^2_F\} = \mathrm{tr}(\mathbb{E}\{\mathbf{\Delta}^T_{\mathrm{eq}}\mathbf{\Delta}_{\mathrm{eq}}\}) \\
&= \mathrm{tr}\left((\mathbf{S}_{\mathrm{tr}}\mathbf{S}^T_{\mathrm{tr}})^{-1}\mathbf{S}_{\mathrm{tr}}\mathbb{E}\{\mathbf{V}^T_{\mathrm{U,tr}}\mathbf{K}^T_{\mathrm{U},r}\mathbf{K}_{\mathrm{U},r}\mathbf{V}_{\mathrm{U,tr}}\}\mathbf{S}^T_{\mathrm{tr}}(\mathbf{S}_{\mathrm{tr}}\mathbf{S}^T_{\mathrm{tr}})^{-1}\right),
\end{aligned} \tag{11}$$

where $\mathbf{\Delta}_{\mathrm{eq}} \triangleq \hat{\mathbf{H}}_{\mathrm{U,eq}} - \mathbf{H}_{\mathrm{U,eq}}$ denotes the estimation error

$$\left(\frac{1+\hat{\epsilon}_{t,m}^{\mathrm{UE}}}{1-\hat{\epsilon}_{t,m}^{\mathrm{UE}}}\right)^2=\frac{\sum_{n=1}^{N}([\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m}^2)}{\sum_{n=1}^{N}([\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m'}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m'}^2)},\ \sin\hat{\theta}_{t,m}^{\mathrm{UE}}=\frac{\sum\limits_{n=1}^{N}([\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m}[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m'}+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m}[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m'})}{\sqrt{\sum\limits_{n=1}^{N}([\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m}^2)\sum\limits_{n=1}^{N}([\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m'}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m'}^2)}},\tag{8}$$

$$\left(\frac{1+\hat{\epsilon}_{r,n}^{\mathrm{BS}}}{1-\hat{\epsilon}_{r,n}^{\mathrm{BS}}}\right)^2=\frac{\sum_{m=1}^{M}[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m'}^2}{\sum_{m=1}^{M}[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m'}^2},\ \sin\hat{\theta}_{r,n}^{\mathrm{BS}}=\frac{\sum\limits_{m=1}^{M}[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m}[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m}+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m'}[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m'}}{\sqrt{\sum\limits_{m=1}^{M}([\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n,m'}^2)\sum\limits_{m=1}^{M}([\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m}^2+[\hat{\mathbf{H}}_{\mathrm{U,eq}}]_{n',m'}^2)}}.\tag{9}$$

matrix.

Proposition 1: In a special case where $\mathbf{S}_{\mathrm{tr,c}}\in\mathbb{C}^{M\times\tau}$ satisfies $\mathbf{S}_{\mathrm{tr,c}}\mathbf{S}_{\mathrm{tr,c}}^H=P_{\mathrm{UE}}\mathbf{I}_M,\mathbf{S}_{\mathrm{tr,c}}\mathbf{S}_{\mathrm{tr,c}}^T=\mathbf{0}$, when $N\to\infty$, the average estimation error approaches the one of the ideal branches case, whose variance results in

$$\sigma_{\mathrm{eq}}^2=\frac{\sigma^2}{P_{\mathrm{UE}}}.\tag{12}$$

Proof 1: See Appendix B.1.

Remark: In our proposed estimator, it is assumed that the length of training sequences satisfies $\tau\geq 2M$ in order to distinguish $2M$ equivalent channels. In contrast, in the complex-valued system model, the real-valued training sequences with $\tau=M$ is appropriate for the channel estimation. Since complex-valued training sequences with $\tau=M$ can be extended to $2M\times 2M$ dimensions, training sequences with length M can be applied for the estimation of real-valued channels which needs further study. Due to the limited article space, designs of optimal training sequences will be left for future work.

3.2 Wireless Channel and IQI Parameters Estimation

As mentioned above, parameters of the TX and RX IQI at the UEs and the TX and RX IQI at the BS are independent. Intuitively, elimination of the TX and RX IQI in the uplink can improve the system performance. Thus, we develop a joint estimator for the wireless channel and the parameters of the TX at the UEs and RX at the BS, based on the estimated equivalent channel in Sect. 3.1. Then, the estimation of the wireless channel is used in the precoding design for the downlink transmission.

3.2.1 With IQI at the UEs only

Assume that the perfect equivalent channel is available at the BS. When IQI only exists at the UEs, we can easily obtain the expressions of the IQI parameters from the expression of the equivalent channel in (2) as

$$\frac{1+\epsilon_{t,m}^{\mathrm{UE}}}{1-\epsilon_{t,m}^{\mathrm{UE}}}=\sqrt{\frac{[\mathbf{H}_{\mathrm{U,eq}}]_{n,m}^2+[\mathbf{H}_{\mathrm{U,eq}}]_{n',m}^2}{[\mathbf{H}_{\mathrm{U,eq}}]_{n,m'}^2+[\mathbf{H}_{\mathrm{U,eq}}]_{n',m'}^2}},$$
$$\sin\theta_{t,m}^{\mathrm{UE}}=\frac{[\mathbf{H}_{\mathrm{U,eq}}]_{n,m}[\mathbf{H}_{\mathrm{U,eq}}]_{n,m'}+[\mathbf{H}_{\mathrm{U,eq}}]_{n',m}[\mathbf{H}_{\mathrm{U,eq}}]_{n',m'}}{\sqrt{([\mathbf{H}_{\mathrm{U,eq}}]_{n,m}^2+[\mathbf{H}_{\mathrm{U,eq}}]_{n',m}^2)([\mathbf{H}_{\mathrm{U,eq}}]_{n,m'}^2+[\mathbf{H}_{\mathrm{U,eq}}]_{n',m'}^2)}},$$

where (n,n') stands for the index pairs of the I and Q branches at an RF chain, and n' is defined as

$$n'=\begin{cases}n+N, & 1\leq n\leq N,\\ n-N, & N+1\leq n\leq 2N,\end{cases}\tag{13}$$

and the operator '′' can be applied to other index, such as m.

When the estimated equivalent channel is applied at the BS, referring to the expressions of parameters of TX IQI at the UEs above, the estimation for IQI parameters from the estimated equivalent channel in the uplink can be given in (8).

When the parameter of IQI at the UEs are estimated, the wireless channel can be acquired by IQI compensation, which is expressed as

$$\hat{\mathbf{H}}=\hat{\mathbf{H}}_{\mathrm{U,eq}}\hat{\mathbf{K}}_{\mathrm{U},t}^{-1},$$

where $\hat{\mathbf{K}}_{\mathrm{U},t}$ is obtained by substituting $(\epsilon_{t,m}^{\mathrm{UE}},\theta_{t,m}^{\mathrm{UE}})$ with $(\hat{\epsilon}_{t,m}^{\mathrm{UE}},\hat{\theta}_{t,m}^{\mathrm{UE}})$ in $\mathbf{K}_{\mathrm{U},t}$.

3.2.2 With IQI at the BS only

Similarly, when IQI only exists at the BS, based on the expression of the equivalent channel in (2), the estimation for IQI parameters based on the estimated equivalent channel can be written as (9). Since the number of users is limited and much smaller than that of BS antennas, the estimation error of IQI parameters is closely related to the estimation error of the equivalent channel.

The wireless channel can be acquired from the estimated IQI parameters as

$$\hat{\mathbf{H}}=\hat{\mathbf{K}}_{\mathrm{U},r}^{-1}\hat{\mathbf{H}}_{\mathrm{U,eq}},$$

where $\hat{\mathbf{K}}_{\mathrm{U},r}$ is obtained by substituting $(\epsilon_{r,n}^{\mathrm{BS}},\theta_{r,n}^{\mathrm{BS}})$ with $(\hat{\epsilon}_{r,n}^{\mathrm{BS}},\hat{\theta}_{r,n}^{\mathrm{BS}})$ in $\mathbf{K}_{\mathrm{U},r}$.

Algorithm 1 : Wireless Channel and IQI Parameters Estimation

Input:
Estimation results of the uplink equivalent channel $\hat{\mathbf{H}}_{\mathrm{U,eq}}$ as shown in (10), and a given threshold σ_0^2.

Output:
Estimation of the wireless channel as $\hat{\mathbf{H}} = \hat{\mathbf{H}}_{\mathrm{U,eq}}^{r,(i-1)}\hat{\mathbf{K}}_{\mathrm{U},t}^{(i-1)^{-1}}$.

1: $\hat{\mathbf{H}}_{\mathrm{U,eq}}^{r,(0)} = \hat{\mathbf{H}}_{\mathrm{U,eq}}$, $i = 1$.
2: **while** $\sigma^2(i) > \sigma_0^2$ **do**
3: Compensation for TX IQI at the UEs: From (8), BS estimates $\{\epsilon_{t,m}^{\mathrm{UE}}, \theta_{t,m}^{\mathrm{UE}},\}_{m=1}^{M}$ based on $\hat{\mathbf{H}}_{\mathrm{U,eq}}^{r,(i-1)}$, and gets the estimation matrix $\hat{\mathbf{K}}_{\mathrm{U},t}^{(i)}$. Then, BS acquires the equivalent channel after TX IQI compensation as $\hat{\mathbf{H}}_{\mathrm{U,eq}}^{t,(i)} = \hat{\mathbf{H}}_{\mathrm{U,eq}}\hat{\mathbf{K}}_{\mathrm{U},t}^{(i)^{-1}}$.
4: Compensation for RX IQI at the BS: From (9), BS estimates $\{\epsilon_{t,n}^{\mathrm{UE}}, \theta_{t,n}^{\mathrm{UE}},\}_{n=1}^{N}$ based on $\hat{\mathbf{H}}_{\mathrm{U,eq}}^{t,(i)}$, and gets the estimation matrix $\hat{\mathbf{K}}_{\mathrm{U},r}^{(i)}$. Then, BS acquires the equivalent channel after RX IQI compensation as $\hat{\mathbf{H}}_{\mathrm{U,eq}}^{r,(i)} = \hat{\mathbf{K}}_{\mathrm{U},r}^{(i)^{-1}}\hat{\mathbf{H}}_{\mathrm{U,eq}}$.
5: Error Calculation: $\sigma^2(i) = \frac{\|\hat{\mathbf{H}}_{\mathrm{U,eq}}^{r,(i)} - \hat{\mathbf{H}}_{\mathrm{U,eq}}^{r,(i-1)}\|_F^2}{4MN}$, $i = i+1$.
6: **end while**

3.2.3 With IQI at Both the BS and UEs

When IQI exists at both the BS and UEs, we estimate parameters of the TX IQI at the UEs and RX IQI at the BS iteratively, and then obtain the wireless channel state information. The wireless channel can be acquired from the estimated IQI parameters as

$$\hat{\mathbf{H}} = \hat{\mathbf{K}}_{\mathrm{U},r}^{-1}\hat{\mathbf{H}}_{\mathrm{U,eq}}\hat{\mathbf{K}}_{\mathrm{U},t}^{-1}.$$

The processing procedure is shown in Algorithm 1. Here the first step of the algorithm is estimating the parameters of TX IQI, instead of RX IQI, because the parameters estimation of IQI at the UEs is more accurate than that of IQI at the BS, since the number of BS antennas is much larger than that of users. Note that the average squared error of the $2M \times 2N$ elements between neighboring iterations is calculated as the termination condition of the algorithm.

When the transmission power of training sequences goes to infinity, we can obtain that $\sigma_{\mathrm{eq}}^2 \to 0$. Then, especially when $N \to \infty$, $\hat{\mathbf{K}}_{\mathrm{U},t} \to \mathbf{K}_{\mathrm{U},t}$. Further, we can get that $\hat{\mathbf{K}}_{\mathrm{U},r} \to \mathbf{K}_{\mathrm{U},r}$ after the compensation of TX IQI at the UEs. The estimation of IQI parameters will be more accurate with more iteration times. However, the computational cost increases with the iteration times. There is a tradeoff between the computational cost and the estimation accuracy. From simulation results in Sect. 5, we find out that one iteration can provide an acceptable accuracy of the estimated wireless channel.

The variance of the estimation error of the wireless channel is expressed as

$$\sigma_{\mathrm{H}}^2 = \mathbb{E}\{\|\hat{\mathbf{H}} - \mathbf{H}\|_F^2\}. \tag{14}$$

Since the estimation of the wireless channel is corresponding to a number of coefficients, such as four coefficients of IQI and estimation error of the uplink equivalent channel, the analytical results of the estimation error is too complicated to be obtained, which will be evaluated by numerical simulation in Sect. 5.

4. Achievable Downlink Rate Analysis

In this section, we investigate the achievable rate of the downlink for the systems with IQI. To facilitate the analysis, the approximation of the average rate is derived with ZF precoding techniques. More precisely, we obtain the asymptotic expression for the extreme case where $N \to \infty$ or $P_{\mathrm{BS}} \to \infty$. The transmit precoding matrices for the downlink are designed based on the estimated equivalent channel and estimated wireless channel from the uplink as shown in Sect. 3.

The average rate for a single cell is given by

$$R = \frac{1}{M}\sum_{m=1}^{2M}\mathbb{E}\left\{\frac{1}{2}\log_2(1 + \mathrm{SINR}_m)\right\}, \tag{15}$$

where the expectation is taken with respect to the wireless propagation channel, and SINR_m represents the received signal-to-interference-plus-noise rate (SINR) for the mth branch. Here we treat each I or Q branch at all UEs as a link, and the real-valued model of the received signal is applied. Thus, the expression of achievable rate in (15) has $2M$ SINR and a factor $\frac{1}{2}$.

4.1 Precoder Based on Uplink Equivalent Channel

Based on the estimated equivalent channel $\hat{\mathbf{H}}_{\mathrm{U,eq}}$ in Sect. 3.1, the precoding matrix $\mathbf{W}$ for the downlink is expressed as

$$\mathbf{W} = \sqrt{\beta_{\mathrm{eq}}}\tilde{\mathbf{I}}_{2N}\hat{\mathbf{H}}_{\mathrm{U,eq}}(\hat{\mathbf{H}}_{\mathrm{U,eq}}^T\hat{\mathbf{H}}_{\mathrm{U,eq}})^{-1}\tilde{\mathbf{I}}_{2M}, \tag{16}$$

where β_{eq} is the power normalization factor which ensures the transmit power constraint $\mathbb{E}\{\|\mathbf{x}\|_F^2\} = \mathbb{E}\{\|\mathbf{d}\|_F^2\}$, i.e., $\mathrm{tr}(\mathbf{W}\mathbf{W}^T) = 2M$. Thus, the normalization factor is given by

$$\beta_{\mathrm{eq}} = \frac{2M}{\mathrm{tr}\left((\hat{\mathbf{H}}_{\mathrm{U,eq}}^T\hat{\mathbf{H}}_{\mathrm{U,eq}})^{-1}\right)}.$$

Proposition 2: The BS transmits signal by applying the ZF precoder based on the estimated uplink equivalent channel. When $N \to \infty$, the normalization factor can be approximated as

$$\beta_{\mathrm{eq}}^{\circ} = (N-M)\frac{M\mathrm{tr}(\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r})}{N\mathrm{tr}\left((\mathbf{K}_{\mathrm{U},t}^T\mathbf{K}_{\mathrm{U},t} + 2\sigma_{\mathrm{eq}}^2\mathbf{I}_{2M})^{-1}\right)}, \tag{17}$$

where σ_{eq}^2 is the average estimation error of the equivalent channel.

Proof 2: Refer to Appendix B.2.

With the precoding matrix $\mathbf{W}$, the received signal at the mth branch in the downlink can be written as

$$y_{\mathrm{D},m}=\sqrt{\beta_{\mathrm{eq}}}[\mathbf{G}]_{mm}d_m+\sqrt{\beta_{\mathrm{eq}}}\sum_{i=1,i\neq m}^{2M}[\mathbf{G}]_{mi}d_i+[\mathbf{K}_{\mathrm{D},r}]_m.\mathbf{v}_{\mathrm{D}}, \tag{18}$$

where $\mathbf{G}=\mathbf{H}_{\mathrm{D,eq}}\tilde{\mathbf{I}}_{2N}\hat{\mathbf{H}}_{\mathrm{U,eq}}(\hat{\mathbf{H}}_{\mathrm{U,eq}}^T\hat{\mathbf{H}}_{\mathrm{U,eq}})^{-1}\tilde{\mathbf{I}}_{2M}$.

Notice that we assume the I and Q branches at the UEs are detected independently. In (18), the first term is the effective signal we want, the second term is the interference for the mth branch from other branches which includes two parts, the interference from branches of other UEs, i.e. $\{d_i\}_{i=1,i\neq m,m'}^{2M}$ and from the corresponding branch at the same UE, i.e. $d_{m'}$. The last term is the noise.

Proposition 3: When the ZF precoder with the estimated equivalent channel is applied at the BS, the approximation of SINR for the mth branch with a large number of BS antennas can be expressed as

$$\mathrm{SINR}_m=\frac{\beta_{\mathrm{eq}}^{\circ}G_{mm}}{\beta_{\mathrm{eq}}^{\circ}\sum\limits_{i=1,i\neq m}^{2M}G_{mi}+\frac{M\sigma^2}{P_{\mathrm{BS}}}}, \tag{19}$$

with

$$G_{ij}=[\mathbf{K}_{\mathrm{D},r}\tilde{\mathbf{I}}_{2M}\mathbf{T}(\mathbf{K}_{\mathrm{U},t}^T)^{-1}\tilde{\mathbf{I}}_{2M}]_{ij}^2, \tag{20}$$

in which the elements of $\mathbf{T}$ are expressed as

$$[\mathbf{T}]_{ij}=\sqrt{(\bar{D}_{ij}+\tilde{D}_{ij})/D_j}, \tag{21}$$

where

$$D_j=(1+2\eta_j\sigma_{eq}^2)^2(N+1)+4\eta_j\sigma_{\mathrm{eq}}^2\gamma_3^{\mathrm{U}},$$
$$\bar{D}_{ij}=\begin{cases}(N+1)\gamma_1^{\mathrm{D}}, & j=i,\\ \gamma_2^{\mathrm{D}}+\gamma_2^{\mathrm{U}}-2\gamma_2^{\mathrm{M}}, & j=i',\\ (1+\gamma_3^{\mathrm{U}}-2\gamma_3^{\mathrm{M}})/2, & \text{else},\end{cases}$$
$$\tilde{D}_{ij}=\begin{cases}\eta_i\sigma_{\mathrm{eq}}^2, & j=i,\\ \left[\eta_i\gamma_3^{\mathrm{U}}+\eta_j(1+\gamma_3^{\mathrm{U}}-2\gamma_3^{\mathrm{M}})+2\eta_i\eta_j\gamma_3^{\mathrm{U}}\sigma_{\mathrm{eq}}^2\right]\sigma_{\mathrm{eq}}^2, & \text{else}.\end{cases}$$

Here $\eta_i=[\mathbf{K}_{\mathrm{U},t}^{-1}]_{ii}^2+[\mathbf{K}_{\mathrm{U},t}^{-1}]_{i'i}^2$, and $\gamma_{\cdot}^{\cdot}$ are determined by the parameters of IQI, which are defined as

$$\begin{aligned}
&\gamma_1^{\mathrm{D}}=\zeta_{t0}^2\zeta_{r0}^2\xi_{t0}^2\xi_{r0}^2,\ \gamma_1^{\mathrm{M}}=\zeta_{t0}\zeta_{r0}\xi_{t0}\xi_{r0},\\
&\gamma_2^{\mathrm{D}}=-(2\zeta_{t1}\zeta_{r1}-\zeta_{t1}-\zeta_{r1})-\xi_{t1}\xi_{r2}/2+1/2,\\
&\gamma_2^{\mathrm{U}}=(4\zeta_{r2}-1)\xi_{r2}+1,\\
&\gamma_2^{\mathrm{M}}=2\zeta_{t0}\xi_{t0}[\zeta_{r0}(\xi_{r0}-\xi_{r3})+\zeta_{r3}(2\xi_{r3}-\xi_{r0})],\\
&\gamma_3^{\mathrm{U}}=(4\zeta_{r2}-1)\xi_{r2}+2,\\
&\gamma_3^{\mathrm{M}}=\zeta_{t0}\xi_{t0}[\zeta_{r0}(3\xi_{r0}-2\xi_{r3})+2\zeta_{r3}(2\xi_{r3}-\xi_{r0})],
\end{aligned} \tag{22}$$

where

$$\begin{aligned}
&\zeta_{t0}=\mathbb{E}\{1/(1+\epsilon_{t,n}^{\mathrm{BS}\,2})^{\frac{1}{2}}\},\ \zeta_{r0}=\mathbb{E}\{1/(1+\epsilon_{r,n}^{\mathrm{BS}\,2})^{\frac{1}{2}}\},\\
&\zeta_{t1}=\mathbb{E}\{1/(1+\epsilon_{t,n}^{\mathrm{BS}\,2})\},\ \zeta_{r1}=\mathbb{E}\{1/(1+\epsilon_{r,n}^{\mathrm{BS}\,2})\},\\
&\zeta_{r2}=\mathbb{E}\{\epsilon_{r,n}^{\mathrm{BS}\,2}/(1+\epsilon_{r,n}^{\mathrm{BS}\,2})^2\},\ \zeta_{r3}=\mathbb{E}\{\epsilon_{r,n}^{\mathrm{BS}\,2}/(1+\epsilon_{r,n}^{\mathrm{BS}\,2})^{\frac{3}{2}}\},\\
&\xi_{t0}=\mathbb{E}\{\cos(\theta_{t,n}^{\mathrm{BS}}/2)\},\ \xi_{r0}=\mathbb{E}\{\cos(\theta_{r,n}^{\mathrm{BS}}/2)\},\\
&\xi_{t1}=\mathbb{E}\{\cos\theta_{t,n}^{\mathrm{BS}}\},\ \xi_{r1}=\mathbb{E}\{\cos\theta_{r,n}^{\mathrm{BS}}\},\\
&\xi_{r2}=\mathbb{E}\{\cos^2\theta_{r,n}^{\mathrm{BS}}\},\ \xi_{r3}=\mathbb{E}\{\cos^3(\theta_{r,n}^{\mathrm{BS}}/2)\},
\end{aligned}$$

Proof 3: See Appendix B.3.

Proposition 3 describes the impact of IQI on the SINR per branch. We know that $\gamma_{\cdot}^{\cdot}$ are bounded only by parameters of IQI at the BS and UEs, irrespective of the number of antennas at the BS. Therefore, when IQI only exists at the UEs, $G_{mm}, G_{mm'}=O(1)$ and $G_{mi}=O(\frac{1}{N}), i\neq m,m'$. Since β_{eq} is linear with N, we can easily find that the SINR is bounded as the number of BS antennas or the BS transmit power goes to infinity. However, when IQI only exists at the BS, $G_{mm'}=O(\frac{1}{N})$, which means the SINR increases without bound as $N\to\infty$. Note that when the BS and the UEs have ideal IQ branches, the SINR increases without bound as $N\to\infty$ or $P_{\mathrm{BS}}\to\infty$.

Taking proposition 3 into (15), we can obtain some corollaries about the achievable rate as follows.

Corollary 1: When the BS transmits signals with ZF precoding based on the estimated equivalent channel, as $N\to\infty$, the average achievable downlink rate with IQI at the UEs has the asymptotic property as

$$\lim_{N\to\infty}R=\frac{1}{2M}\sum_{i=1}^{2M}\log_2\left(1+\mathrm{SINR}_i^{\infty}\right), \tag{23}$$

with

$$\mathrm{SINR}_i^{\infty}=\left(\frac{[\mathbf{K}_{\mathrm{D},r}]_{ii}[\mathbf{K}_{\mathrm{U},t}^{-1}]_{ii}+[\mathbf{K}_{\mathrm{D},r}]_{ii'}[\mathbf{K}_{\mathrm{U},t}^{-1}]_{i'i}}{[\mathbf{K}_{\mathrm{D},r}]_{ii}[\mathbf{K}_{\mathrm{U},t}^{-1}]_{i'i}+[\mathbf{K}_{\mathrm{D},r}]_{ii'}[\mathbf{K}_{\mathrm{U},t}^{-1}]_{i'i'}}\right)^2.$$

The gap between the average achievable downlink rate in the ideal branches case and the rate with IQI only at the BS has the asymptotic property as

$$\lim_{N\to\infty}\Delta R^{\mathrm{oBS}}=\log_2\left(\frac{\gamma_3^{\mathrm{U}}+\frac{P_{\mathrm{BS}}}{\sigma^2}\left[1+(1+2\sigma_{\mathrm{eq}}^2)\gamma_3^{\mathrm{U}}-2\gamma_3^{\mathrm{M}}\right]}{\gamma_1^{\mathrm{D}}/(1+2\sigma_{\mathrm{eq}}^2)}\right). \tag{24}$$

Proof 4: Since $[\mathbf{T}]_{ii}=O(1)$ and $[\mathbf{T}]_{ij}=O(\frac{1}{N})$, $j\neq i$, (23) follows directly from (15) and (19).

In the ideal IQ branches case, the achievable rate is given by

$$R^{\mathrm{ideal}}=\log_2\left(1+\frac{(N-M)\frac{P_{\mathrm{BS}}}{M\sigma^2}}{\frac{P_{\mathrm{BS}}}{\sigma^2}\sigma_{\mathrm{eq}}^2+1+2\sigma_{\mathrm{eq}}^2}\right). \tag{25}$$

When IQI only exists at the BS, i.e. $\epsilon_{\cdot}^{\mathrm{UE}}=0, \theta_{\cdot}^{\mathrm{UE}}=0$, we have $\mathbf{K}_{\mathrm{D},r} = \mathbf{K}_{\mathrm{D},r}^{-1} = \mathbf{I}_M$ and $\eta_i = 1$ in (20) and (21). Based on (19), the average rate with IQI only at the BS, R^{oBS}, can be expressed as

$$R^{\mathrm{oBS}} = \log_2\left(1+\frac{(N+1)\gamma_1^{\mathrm{D}}/(1+2\sigma_{\mathrm{eq}}^2)}{M[1+(1+2\sigma_{\mathrm{eq}}^2)\gamma_3^{\mathrm{U}}-2\gamma_3^{\mathrm{M}}]+\frac{M\sigma^2}{P_{\mathrm{BS}}}\gamma_3^{\mathrm{U}}}\right),$$

which applies the approximation that $\gamma_2^{\mathrm{D}}+\gamma_2^{\mathrm{U}}-2\gamma_2^{\mathrm{M}} \approx 1+\gamma_3^{\mathrm{U}}-2\gamma_3^{\mathrm{M}}$. Thus, the asymptotic property of $\Delta R^{\mathrm{oBS}} = R^{\mathrm{ideal}} - R^{\mathrm{oBS}}$ can be given as (24).

Corollary 1 shows that the existence of IQI at the UEs limits the achievable rate to a finite ceiling, which is only determined by the parameters of IQI at the UEs. Notice that when IQI of UEs only has amplitude imbalance, the achievable rate will grow without bound as N goes to infinity, which reveals that the phase imbalance at the UEs plays a more important role on the achievable rate. Moreover, when the IQI only exists at the BS, the rate increases without bound, which has a gap determined by the number of UEs and the estimation error compared with the ideal IQ branches case. It means that the impact of IQI at the UEs is much higher than that of IQI at the BS. Moreover, we know the IQI at the UEs is more serious than the IQI at the BS [11], which also indicates the IQI at the UEs has more adverse influence on the achievable downlink rate.

Corollary 2: With ZF precoding based on the estimated equivalent channel in the uplink, the average achievable downlink rate with IQI at both the BS and UEs has the following asymptotic property as $P_{\mathrm{BS}} \to \infty$:

$$\lim_{P_{\mathrm{BS}}\to\infty} R = \frac{1}{2M}\sum_{m=1}^{2M}\log_2\left(1+\frac{G_{mm}}{\sum\limits_{i=1,i\neq m}^{2M} G_{mi}}\right). \tag{26}$$

Proof 5: The proof follows directly from (15).

This corollary reveals that the achievable downlink rate converges to a finite ceiling as $P_{\mathrm{BS}} \to \infty$. Notice that when there is no IQI, we know that the ceiling of the achievable rate depends only on the estimation error of the uplink channel which destroys the orthogonality of the inter-user reception when the ZF precoding technique is used. Here, the difference between TX IQI and RX IQI at the BS also contributes to the orthogonality destruction. As well, the existence of IQI at the UEs introduces the interference from the corresponding I or Q branch at the same UE.

4.2 Precoder Based on Wireless Channel

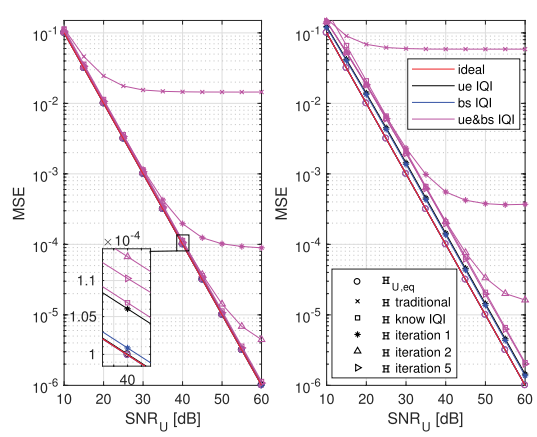
In Sect. 4.1, we consider the ZF precoder based on equivalent channel, since it's generally used in systems protocol with the TDD mode. As mentioned above, the hardware impairments of the TX and RX at the BS and TX and RX at the UEs are independent. The existence of IQI in the uplink still affects the performance of the downlink, even though the TX and the RX at the downlink have the ideal IQ branches. Thus, we investigate the achievable downlink rate with ZF precoder with the wireless channel to eliminate the impact of the TX and RX IQI in the uplink. The information of the wireless channel can be acquired by the estimator in Sect. 3.2. The performance of the receiver with ZF precoding based on the estimated wireless channel can be easily obtained from the analysis results in Sect. 4.1, by setting $\mathbf{K}_{\mathrm{U},r} = \mathbf{I}_{2N}, \mathbf{K}_{\mathrm{U},t} = \mathbf{I}_{2M}$ and substituting σ_{eq}^2 with σ_{H}^2. From the result in (23), we can easily find out that the TX IQI at the UEs has a similar impact as the RX IQI at the UEs. Thus, the ceilings caused by TX at the UEs will be improved by mitigating the effect of TX and RX IQI.

5. Numerical Results

In this section, we verify our analysis by simulations. In all the simulation scenarios, we assume that the parameters of the amplitude and phase imbalance, i.e., $\epsilon_{\cdot}$ and $\theta_{\cdot}$ mentioned above, are drawn from Gaussian distributions $\mathcal{N}(0,\sigma_\epsilon^2)$ and $\mathcal{N}(0,\sigma_\theta^2)$, respectively. When measured in dB, σ_ϵ is computed as $20\log_{10}(1+\sigma_\epsilon)$. The signal-to-noise ratio (SNR) for each stream at the BS is defined as SNR= $10\log_{10}(P_{\mathrm{BS}}/\sigma^2)$. We adopt the discrete Fourier sequences with length $\tau = 2M$ as training sequences, which satisfy the assumption in Proposition 1. For different cells, the values of the IQI at the BS and UEs are different. We investigate the average performance of cells which have the similar BS and UEs hardware implementations, i.e. the coefficients of IQI have the same distribution.

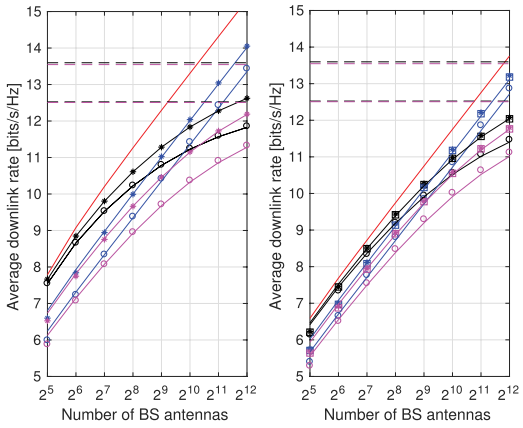
Note that in Figs. 1–7, four scenarios are considered as follows, ideal IQ branches (red lines), IQI only exists at the UEs (black lines), IQI only exists at the BS (blue lines), and IQI exists at both the BS and UEs (magenta lines), which are marked as "ideal", "ue IQI", "bs IQI" and "ue&bs IQI", respectively.

Firstly, we exemplify the impact of IQI on the ac-



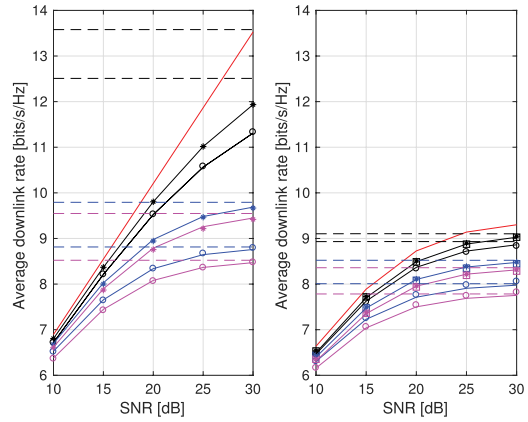
(a) $(\sigma_\epsilon, \sigma_\theta) = (1\ \mathrm{dB}, 2^\circ)$ (b) $(\sigma_\epsilon, \sigma_\theta) = (2\ \mathrm{dB}, 5^\circ)$

Fig. 1 MSE of the equivalent channel and the wireless channel estimation in Sect. 3 versus $\mathrm{SNR_U}$, for $N = 128$, $M = 10$.



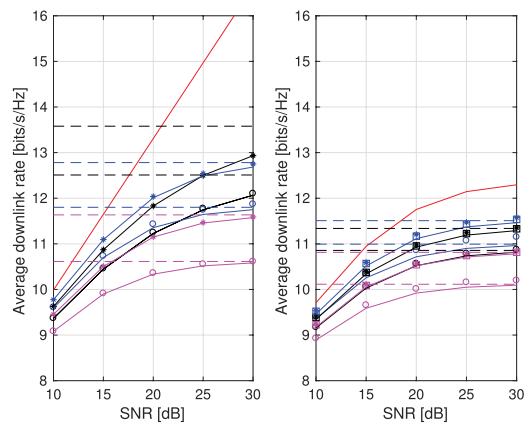
(a) perfect CSI (b) estimated CSI

Fig. 2 Achievable average downlink rate versus the number of BS antennas, M=10, SNR=20 dB, $(\sigma_\epsilon, \sigma_\theta)$ = (1 dB, 2°).



(a) perfect CSI (b) estimated CSI

Fig. 3 Achievable average downlink rate versus SNR, N=128, M=10, $(\sigma_\epsilon, \sigma_\theta)$ = (1 dB, 2°).



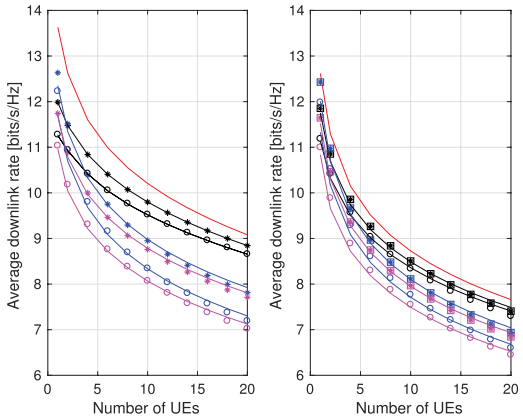
(a) perfect CSI (b) estimated CSI

Fig. 4 Achievable average downlink rate versus SNR, N=1024, M=10, $(\sigma_\epsilon, \sigma_\theta)$ = (1 dB, 2°).

curacy of channel estimators proposed in Sect. 3. Figure 1 shows the estimation error of the real-valued channel as a function of SNR in the uplink which is defined as $\mathrm{SNR_U} = 10\log_{10}(P_{\mathrm{UE}}/\sigma^2)$. For the estimator of the uplink equivalent channel (marked as "$\mathbf{H}_{\mathrm{U,eq}}$"), curves of MSE of the uplink equivalent channel overlap that of the ideal IQ branches case in Fig. 1, and the estimation error is inversely proportional to $\mathrm{SNR_U}$. It reveals that the impact of IQI can be removed for estimation of the uplink equivalent channel, which coincides with Proposition 1. For the proposed estimator of the wireless channel (marked as "**H** iteration 1/2/5"), the estimation error has a non-zero error floor at high SNRs only when both BS and UEs have IQI, and the floor can be lowered by increasing the number of iteration in Algorithm 1. Taking $(\sigma_\epsilon, \sigma_\theta)$ = (2 dB, 5°) as an example, MSE of the wireless channel is close to the floor when SNR reaches about 40 dB, and the floor is close to zero when the number of iteration times is 5. Compared with the traditional method ignoring IQI (marked as "**H** traditional"), the superiority of the proposed wireless channel estimation method gets more obvious when IQI gets more severe. Moreover, we find out that the performance of the proposed method is close to that of the channel estimation with the perfect IQI information (marked as "**H** know IQI") that is under the result of "**H** iteration 5" in Fig. 1, when the number of iteration times is 5, even with severe IQI. It should be noted that the estimator with one time iteration is acceptable in practical applications, because (i) it has lower computational cost, and (ii) the improvement brought by more iteration times is small when $\mathrm{SNR_U} \leq 30$ dB and the higher $\mathrm{SNR_U}$ is not applicable.

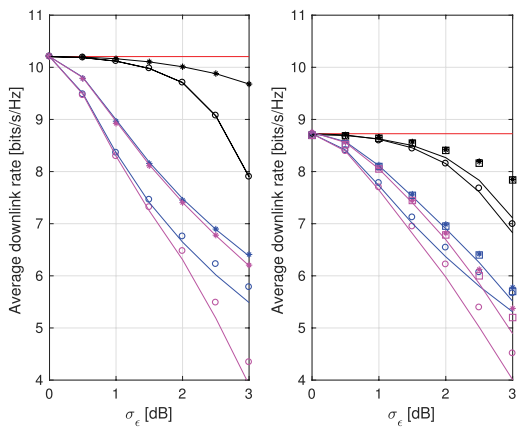
Next, we evaluate the impact of the IQI on the achievable rate of the system with the ZF precoding. Precoders with the prefect CSI and the estimated CSI obtained by the proposed methods in Sect. 3 are investigated, respectively. Note that we assume that the SNR in the uplink, $\mathrm{SNR_U}$, is 20 dB, and the number of iterations used for the wireless channel estimation is 1 in Figs. 2–7 since the improvement brought by more iteration times is small. Simulation results of the precoders based on the uplink equivalent channel $\mathbf{H}_{\mathrm{U,eq}}$ and the wireless channel **H** are marked as "∘" and "∗", respectively, and the solid lines represent the analytical results correspondingly. Meanwhile, the dashed lines denote ceilings of the achievable rates. In the estimated CSI case, we find out that the results of the proposed wireless channel estimation is almost the same as the case with the perfect IQI information (marked as "□").

Average rates versus the number of BS antennas are illustrated in Fig. 2. We can find out that when IQI only exists at the BS, the average rate increases proportionally with the number of BS antennas, and compared to the ideal IQ branches case, there is a gap determined by the number of UEs as shown in (24). However, when IQI exists at the UEs, the rate is limited to a finite ceiling derived in (23). Meanwhile, when the number of BS antennas is limited, e.g. $N < 1024$, the impact of IQI at the BS is more severe than that of IQI at the UEs. On the other hand, when the number of BS antennas goes to infinity, IQI at the UEs plays a more important role on the performance degradation. The reason is that IQI at one UE only affects its own signals of I and Q branches, but the variation of TX and RX IQI at the BS



(a) perfect CSI (b) estimated CSI

Fig. 5 Achievable average downlink channel rate versus the number of users, N=128, SNR=20 dB, $(\sigma_\epsilon, \sigma_\theta)$ = (1 dB, 2°).

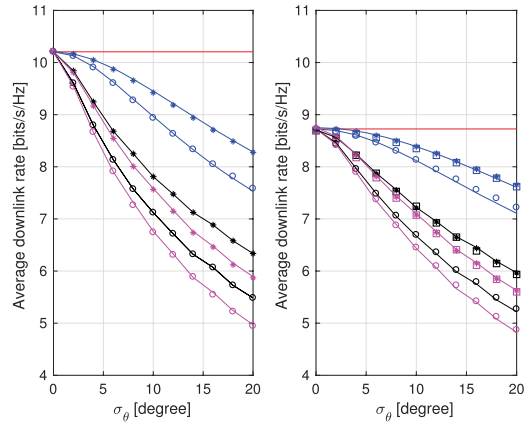


(a) perfect CSI (b) estimated CSI

Fig. 6 Achievable average downlink rate versus the standard deviation of the amplitude imbalance, N=128, M=10, SNR=20 dB, σ_θ = 0°.

destroys the orthogonality of the inter-user reception, and the destruction can be attenuated by increasing the number of BS antennas. Finally, it is worth mentioning that the precoder based on the wireless channel outperforms that based on the uplink equivalent channel.

In Fig. 3 and Fig. 4, we illustrate the average rate as a function of SNR for N = 128 and N = 1024, respectively. It shows that IQI limits the average rate to finite ceilings shown in (26), as the transmit power at the BS goes to infinity. Without doubt, the precoder based on the wireless channel performs better than the precoder based on uplink equivalent channel, since the former eliminates the impact of TX IQI at the UEs and RX IQI at the BS. When the channel estimation error exists, the effect of IQI is no longer prominent, because the imperfect channel information leads to the non-orthogonality of the inter-user reception. Comparing IQI at the UEs with IQI at the BS, it is determined by the ratio of the number of BS antennas to that of UEs which will cause more serious performance loss. Figure 3 shows IQI at the BS leads to more significant impact on performance degradation at high SNR, when the ratio is small. On the



(a) perfect CSI (b) estimated CSI

Fig. 7 Achievable average downlink rate versus the standard deviation of the phase imbalance, N=128, M=10, SNR=20 dB, σ_ϵ = 0 dB.

contrary, when the ratio is large, IQI at the UEs causes more serious impact as shown in Fig. 4. It shows the same results as in Fig. 2.

We know that the ratio of the number of BS antennas to that of UEs plays an important role on the average rate of MIMO systems. Thus, we investigate the average rate versus the number of users for a fixed number of BS antennas in Fig. 5. Obviously, it shows that when the perfect channel information is given, the impact of IQI at the UEs can be ignorable compared with the change in the number of UEs, since IQI at one UE only affects the performance of the receiver itself. In other cases, the rate decreases rapidly with the number of users. It is because the orthogonality of inter-user reception is weakened by channel estimation error and IQI at the BS.

Furthermore, we focus on the effect of the amplitude imbalance and the phase imbalance on behaviors of the system performance. In Fig. 6, we evaluate the average rate as a function of the amplitude imbalance with the matched phases between I and Q branches. It shows that the performance degradation of the system is more sensitive to the amplitude imbalance at the BS than that at the UEs. The reason is that the amplitude imbalance mainly affects diagonal elements of IQI matrices with phase balance, and in this case, I and Q branches at an RF chain have no inter-branch interference. Thus, the mismatch of TX and RX at the BS has more impact on the performance [9]. It needs to be pointed out that when the precoding matrix is design based on the estimated wireless channel, the simulation results have some deviations from our analysis results, especially for situations where the amplitude imbalance is very serious. It is because the estimation error is evident when the amplitude imbalance is too severe. In Fig. 7, we evaluate the average rate as a function of the phase imbalance with the matched amplitudes between I and Q branches. Obviously, the average rate is more sensitive to the phase imbalance at the UEs than that at the BS, since the phase imbalance leads to inter-branch interference at an RF chain. When IQI only exists at the BS, the inter-branch interference can be reduced with the

large scale antenna array. However, for IQI at the UEs, the inter-branch interference always exists, because the received signals of I and Q branches at the UEs are interfered with each other.

6. Conclusion

In this paper, we have studied massive MU-MIMO systems with IQI at both the BS and UEs, operating in the TDD mode. The uplink equivalent channel estimation with the LS criterion and the wireless channel estimation based on the estimated uplink equivalent channel were provided. We also analyzed the impact of IQI on the achievable average rate of the downlink of the system with widely-linear ZF precoding based on the estimated channel, and derived the closed-form expression of the rate. Analytical and simulation results demonstrated the effectiveness of the proposed estimators of the uplink equivalent channel and the wireless channel. We prove that the achievable average rate of the downlink is limited to a finite ceiling determined by the parameters of IQI when the transmit power goes to infinity. With the growth of the number of BS antennas, the widely-linear ZF precoding ensures that the rate increases without bound if IQI only exists at the BS. However, the rate of the system with IQI at the UEs still exhibits limitation determined by the parameters of IQI at the UEs. Moreover, the impact of the amplitude imbalance at the BS and the phase imbalance at the UEs is a more significant determiner of system performance than other imbalance. Obviously, the precoding based on the wireless channel can provide better performance than that based on uplink equivalent channel, since the former eliminates the impact of transceiver IQI in the uplink. There still exists several open topics, such as investigating the impact of IQI on other widely-linear precoding methods, adopting more practical channel models, designing optimal training sequences, and analyzing the energy efficiency of the proposed massive MIMO system.

References

[1] E.G. Larsson, O. Edfors, F. Tufvesson, and T.L. Marzetta, "Massive MIMO for next generation wireless systems," IEEE Commun. Mag., vol.52, no.2, pp.186–195, Feb. 2014.
[2] F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," IEEE Signal Process. Mag., vol.30, no.1, pp.40–60, Jan. 2013.
[3] L. Lu, G.Y. Li, A.L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," IEEE J. Sel. Topics Signal Process., vol.8, no.5, pp.742–758, Oct. 2014.
[4] S. Wagner, R. Couillet, M. Debbah, and D.T. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," IEEE Trans. Inf. Theory, vol.58, no.7, pp.4509–4537, March 2012.
[5] F. Boccardi, R.W. Heath, A. Lozano, T.L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," IEEE Commun. Mag., vol.52, no.2, pp.74–80, Feb. 2014.
[6] J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A.C. Soong, and J.C. Zhang, "What will 5G be?," IEEE J. Sel Areas Commun., vol.32, no.6, pp.1065–1082, June 2014.
[7] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," IEEE Trans. Inf. Theory, vol.60, no.11, pp.7112–7139, Nov. 2014.
[8] W. Zhang, H. Ren, C. Pan, M. Chen, R.C. de Lamare, B. Du, and J. Dai, "Large-scale antenna systems with UL/DL hardware mismatch: Achievable rates analysis and calibration," IEEE Trans. Commun., vol.63, no.4, pp.1216–1229, April 2015.
[9] H. Wei, D. Wang, J. Wang, and X. You, "Impact of RF mismatches on the performance of massive MIMO systems with ZF precoding," Science China Information Sciences, vol.59, no.2, pp.1–14, Feb. 2016.
[10] B. Razavi and R. Behzad, RF Microelectronics, Prentice Hall, New Jersey, 1998.
[11] T. Schenk, RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation, Springer Science & Business Media, 2008.
[12] A. Hakkarainen, J. Werner, K.R. Dandekar, and M. Valkama, "Widely-linear beamforming and RF impairment suppression in massive antenna arrays," J. Commun. Netw., vol.15, no.4, pp.383–397, Aug. 2013.
[13] N. Kolomvakis, M. Matthaiou, and M. Coldrey, "IQ imbalance in multiuser systems: Channel estimation and compensation," IEEE Trans. Commun., vol.64, no.7, pp.3039–3051, July 2016.
[14] S. Wang and L. Zhang, "Signal processing in massive MIMO with IQ imbalances and low-resolution ADCs," IEEE Trans. Wireless Commun., vol.15, no.12, pp.8298–8312, Dec. 2016.
[15] S. Zarei, W.H. Gerstacker, J. Aulin, and R. Schober, "I/Q imbalance aware widely-linear receiver for uplink multi-cell massive MIMO systems: Design and sum rate analysis," IEEE Trans. Wireless Commun., vol.15, no.5, pp.3393–3408, May 2016.
[16] N. Kolomvakis, M. Coldrey, T. Eriksson, and M. Viberg, "Downlink performance of regularized ZF in massive MIMO systems subject to IQ imbalance," Communications (ICC), 2017 IEEE International Conference on, pp.1–7, May 2017.
[17] N. Kolomvakis, M. Coldrey, T. Eriksson, and M. Viberg, "Massive MIMO systems with IQ imbalance: Channel estimation and sum rate limits," IEEE Trans. Commun., vol.65, no.6, pp.2382–2396, June 2017.
[18] S. Zarei, W. Gerstacker, and R. Schober, "I/Q imbalance aware widely-linear precoding for downlink massive MIMO systems," Globecom Workshops (GC Wkshps), pp.301–307, Dec. 2014.
[19] Y. Liang, H. Li, F. Li, R. Song, and L. Yang, "Channel compensation for reciprocal TDD massive MIMO-OFDM with IQ imbalance," IEEE Wireless Commun. Lett., vol.6, no.6, pp.778–781, Dec. 2017.
[20] W. Zhang, R.C. de Lamare, C. Pan, M. Chen, J. Dai, B. Wu, and X. Bao, "Widely linear precoding for large-scale MIMO with IQI: Algorithms and performance analysis," IEEE Trans. Wireless Commun., vol.16, no.5, pp.3298–3312, May 2017.
[21] Z.J. Liu, D.C. Sun, J.L. Wang, and K.C. Yi, "Impact and compensation of I/Q imbalance on channel reciprocity of time-division-duplexing multiple-input multiple-output systems," IET Communications, vol.7, no.7, pp.663–672, Feb. 2013.
[22] J. Evans and D.N.C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," IEEE Trans. Inf. Theory, vol.46, no.6, pp.2059–2078, Sept. 2000.
[23] A. Lozano, A.M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," IEEE Trans. Inf. Theory, vol.49, no.10, pp.2527–2544, Oct. 2003.

Appendix A: Old and Application-Related Results on Random Vectors

Lemma 1 ([22]): Assume that the elements of $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{N\times 1}$ are i.i.d complex Gaussian random variables with zero-mean

and unit variance, and $\mathbf{A} \in \mathbb{C}^{N\times N}$ is a Hermitian matrix with bounded spectral norm, and is independent of $\mathbf{x}$ and $\mathbf{y}$. Then, there are

(1) $\frac{1}{N}\mathbf{x}^H\mathbf{A}\mathbf{x} \xrightarrow[N\to\infty]{a.s.} \frac{1}{N}\mathrm{tr}(\mathbf{A})$,

(2) $\frac{1}{N}\mathbf{x}^H\mathbf{A}\mathbf{y} \xrightarrow[N\to\infty]{a.s.} 0$,

(3) $\mathbb{E}\left\{|\frac{1}{N}\mathbf{x}^H\mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{tr}(\mathbf{A})|^2\right\} = O(\frac{1}{N})$

where 'a.s.' denotes almost surely convergence, i.e. convergence with probability one.

Lemma 2 ([9]): For full column rank matrices $\mathbf{A} = \left[\mathbf{a}_1^H, \cdots, \mathbf{a}_M^H,\right]^T$, $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_M]$, where $\mathbf{a}_m, \mathbf{b}_m \in \mathbb{C}^{N\times 1}$, it holds that

$$[\mathbf{A}\mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}]_{ij} = \frac{\det\left(\mathbf{B}_{\mathrm{chg}}^{ij\ H}\mathbf{B}\right)}{\det\left(\mathbf{B}^H\mathbf{B}\right)}, \tag{A·1}$$

and

$$[\mathbf{A}\mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}]_{ij}^2 = \frac{\det\left(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T - \boldsymbol{\alpha}_i\mathbf{B}_{\mathrm{del}}^{j\ T}(\mathbf{B}_{\mathrm{del}}^j\mathbf{B}_{\mathrm{del}}^{j\ T})^{-1}\mathbf{B}_{\mathrm{del}}^j\boldsymbol{\alpha}_i^T\right)}{\det\left(\boldsymbol{\beta}_j\boldsymbol{\beta}_j^T - \boldsymbol{\beta}_j\mathbf{B}_{\mathrm{del}}^{j\ T}(\mathbf{B}_{\mathrm{del}}^j\mathbf{B}_{\mathrm{del}}^{j\ T})^{-1}\mathbf{B}_{\mathrm{del}}^j\mathbf{b}_j^T\right)}, \tag{A·2}$$

where $\mathbf{B}_{\mathrm{chg}}^{ij} = \left[\mathbf{b}_1, \cdots, \mathbf{b}_{j-1}, \mathbf{a}_i, \mathbf{b}_{j+1}, \cdots, \mathbf{b}_M\right]$, $\boldsymbol{\alpha}_i = \mathbf{a}_i^T\mathbf{B}$, $\boldsymbol{\gamma}_i = \mathbf{b}_i^T\mathbf{B}$, $\boldsymbol{\beta}_j = \mathbf{B}_{\mathrm{del}}^{j\ T}\mathbf{B}$, and $\mathbf{B}_{\mathrm{del}}^j = \left[\mathbf{b}_1, \cdots, \mathbf{b}_{j-1}, \mathbf{b}_{j+1}, \cdots, \mathbf{b}_M\right]$.

Lemma 3: When the number of antennas at the BS goes to infinity, the derivations of characteristic of $\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_i$ and $\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_i$ can be approximated as

$$\begin{aligned}
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_i\mathbf{h}_j^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j \xrightarrow[N\to\infty]{a.s.} (N+1)\gamma_1^{\mathrm{D}},\\
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_i\mathbf{h}_j^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j \xrightarrow[N\to\infty]{a.s.} (N+1),\\
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_i\mathbf{h}_j^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j \xrightarrow[N\to\infty]{a.s.} (N+1)\gamma_1^{\mathrm{M}}.
\end{aligned} \tag{A·3}$$

Proof 6: With the expressions of $\mathbf{K}_{\mathrm{D},t}$ and $\mathbf{K}_{\mathrm{U},r}$ defined in (3) and the definition of $\mathbf{K}_{\mathrm{U},r}^i$ and $\mathbf{K}_{\mathrm{D},t}^i, i = 1, 2, 3, 4$ in (4) and (5), we observe that

$$\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r} = \begin{bmatrix}\mathrm{diag}(\mathbf{a}_{11}+\mathbf{a}_{12}) & \mathrm{diag}(\mathbf{a}_{21}-\mathbf{a}_{22})\\ \mathrm{diag}(\mathbf{a}_{21}+\mathbf{a}_{22}) & \mathrm{diag}(\mathbf{a}_{11}-\mathbf{a}_{12})\end{bmatrix}, \tag{A·4}$$

where

$$\begin{aligned}
[\mathbf{a}_{11}]_n &= a_n\left[\cos\left(\frac{\theta_{t,n}^{\mathrm{BS}}-\theta_{r,n}^{\mathrm{BS}}}{2}\right) + \epsilon_{t,n}^{\mathrm{BS}}\epsilon_{r,n}^{\mathrm{BS}}\cos\left(\frac{\theta_{t,n}^{\mathrm{BS}}+\theta_{r,n}^{\mathrm{BS}}}{2}\right)\right],\\
[\mathbf{a}_{12}]_n &= a_n\left[\epsilon_{t,n}^{\mathrm{BS}}\cos\left(\frac{\theta_{t,n}^{\mathrm{BS}}-\theta_{r,n}^{\mathrm{BS}}}{2}\right) + \epsilon_{r,n}^{\mathrm{BS}}\cos\left(\frac{\theta_{t,n}^{\mathrm{BS}}+\theta_{r,n}^{\mathrm{BS}}}{2}\right)\right],\\
[\mathbf{a}_{21}]_n &= a_n\left[\sin\left(\frac{\theta_{t,n}^{\mathrm{BS}}+\theta_{r,n}^{\mathrm{BS}}}{2}\right) - \epsilon_{t,n}^{\mathrm{BS}}\epsilon_{r,n}^{\mathrm{BS}}\sin\left(\frac{\theta_{t,n}^{\mathrm{BS}}-\theta_{r,n}^{\mathrm{BS}}}{2}\right)\right],\\
[\mathbf{a}_{22}]_n &= a_n\left[\epsilon_{t,n}^{\mathrm{BS}}\sin\left(\frac{\theta_{t,n}^{\mathrm{BS}}-\theta_{r,n}^{\mathrm{BS}}}{2}\right) - \epsilon_{r,n}^{\mathrm{BS}}\sin\left(\frac{\theta_{t,n}^{\mathrm{BS}}+\theta_{r,n}^{\mathrm{BS}}}{2}\right)\right],\\
&\qquad n = 1, \cdots, N,
\end{aligned}$$

in which $a_n = 1/\sqrt{(1+\epsilon_{t,n}^{\mathrm{BS}^2})(1+\epsilon_{t,n}^{\mathrm{BS}^2})}$. Since the parameters of IQI are small, $\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}$ can be approximated as a Hermitian matrix.

Similarly, the expression of $\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}$ is given by

$$\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r} = \begin{bmatrix}\mathrm{diag}(1+\mathbf{a}_1) & \mathrm{diag}(\mathbf{a}_2)\\ \mathrm{diag}(\mathbf{a}_2) & \mathrm{diag}(1-\mathbf{a}_1)\end{bmatrix}, \tag{A·5}$$

where $[\mathbf{a}_1]_n = 2\epsilon_{r,n}^{\mathrm{BS}}\cos\left(\theta_{r,n}^{\mathrm{BS}}\right)/(1+\epsilon_{r,n}^{\mathrm{BS}\ 2})$, and $[\mathbf{a}_2]_n = \sin\left(\theta_{r,n}^{\mathrm{BS}}\right)$. With the help of Lemma 1, there are

$$\begin{aligned}
&\frac{1}{N}\mathbf{h}_i^T\mathbf{A}\mathbf{h}_i \xrightarrow[N\to\infty]{a.s.} \frac{1}{2N}\mathrm{tr}(\mathbf{A}),\\
&\mathrm{var}\left(\frac{1}{N}\mathbf{h}_i^T\mathbf{A}\mathbf{h}_i\right) \xrightarrow[N\to\infty]{a.s.} \frac{1}{4N^2}\left[3\mathrm{tr}(\mathbf{A}^2) - \frac{1}{2N}\mathrm{tr}^2(\mathbf{A})\right],
\end{aligned}$$

where $\mathbf{A}$ represents $\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}$ or $\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}$.

Since $\mathbf{A}$ can be approximated as an unit matrix, $\frac{1}{2N}\mathrm{tr}(\mathbf{A}^2) \approx \left[\frac{1}{2N}\mathrm{tr}(\mathbf{A})\right]^2$ for the large number of BS antennas. Based on expressions in (A·4) and (A·5), when $N \to \infty$, there are $\frac{1}{2N}\mathrm{tr}(\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}) \approx \mathbb{E}\{[\mathbf{a}_{11}]_n\}$ and $\frac{1}{2N}\mathrm{tr}(\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}) = 1$. Thus, (A·3) can be easily obtained.

Lemma 4: When $N \to \infty$, the derivations of characteristic of $\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}$ and $\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}$ can be expressed as

$$\begin{aligned}
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}\mathbf{h}_j^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{j'} \xrightarrow[N\to\infty]{a.s.} \begin{cases}\gamma_2^{\mathrm{D}}, & j=i,i',\\ 0, & \text{else},\end{cases}\\
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}\mathbf{h}_j^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{j'} \xrightarrow[N\to\infty]{a.s.} \begin{cases}\gamma_2^{\mathrm{U}}, & j=i,i',\\ 0, & \text{else},\end{cases}\\
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}\mathbf{h}_j^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{j'} \xrightarrow[N\to\infty]{a.s.} \begin{cases}\gamma_2^{\mathrm{M}}, & j=i,i',\\ 0, & \text{else}.\end{cases}
\end{aligned} \tag{A·6}$$

Proof 7: Let $\mathbf{A}$ represent $\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}$ or $\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}$, and redefine $\mathbf{h}_i = [\mathbf{h}_{Ri}^T, \mathbf{h}_{Ii}^T]^T$, $\mathbf{A} = \begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\\ \mathbf{A}_3 & \mathbf{A}_4\end{bmatrix}$, where $\mathbf{A}_i$ is the diagonal matrix. Thus, there is

$$\mathbf{h}_i^T\mathbf{A}\mathbf{h}_{i'} = -\mathbf{h}_{Ri}^T\mathbf{A}_1\mathbf{h}_{Ii} - \mathbf{h}_{Ii}^T\mathbf{A}_3\mathbf{h}_{Ii} + \mathbf{h}_{Ri}^T\mathbf{A}_2\mathbf{h}_{Ri} + \mathbf{h}_{Ii}^T\mathbf{A}_4\mathbf{h}_{Ri}.$$

For $j = i$, since $\mathbf{h}_{Ri}$, and $\mathbf{h}_{Ii}$ are i.i.d., with the help of the law of large numbers, we can obtain that

$$|\mathbf{h}_i^T\mathbf{A}\mathbf{h}_{i'}|^2 \approx \frac{1}{4}\left[\mathrm{tr}^2(\mathbf{A}_2-\mathbf{A}_3) + \mathrm{tr}(2(\mathbf{A}_2^2+\mathbf{A}_3^2) + (\mathbf{A}_1-\mathbf{A}_4)^2)\right].$$

Based on (A·4) and (A·5), there are $\mathrm{tr}(2(\mathbf{A}_2^2+\mathbf{A}_3^2) \xrightarrow{N\to\infty} 0$, and

$$\begin{aligned}
&|\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}|^2 \xrightarrow{N\to\infty} N\mathbb{E}\{[\mathbf{a}_{12}]_n^2 + [\mathbf{a}_{21}]_n^2\},\\
&|\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}|^2 \xrightarrow{N\to\infty} N\mathbb{E}\{[\mathbf{a}_1]_n^2 + [\mathbf{a}_2]_n^2\}.
\end{aligned}$$

Similarly, we can obtain that

$$\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'}\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_{i'} \xrightarrow{N\to\infty} N\mathbb{E}\{[\mathbf{a}_{12}]_n[\mathbf{a}_1]_n+[\mathbf{a}_{21}]_n[\mathbf{a}_2]_n\}.$$

With the expressions of $\mathbf{a}_\cdot$, $\gamma_2^{\cdot}$ given in (22) can be acquired.

For $j = i'$, since $\mathbf{A}$ can be approximated as a Hermitian matrix, the result is very close to the case that $j = i$.

For $j \neq i, i'$, since $\mathbf{h}_{Ri}$, $\mathbf{h}_{Ii}$, $\mathbf{h}_{Rj}$, and $\mathbf{h}_{Ij}$ are i.i.d., $\mathbf{h}_i^T\mathbf{A}\mathbf{h}_{i'}\mathbf{h}_j^T\mathbf{A}\mathbf{h}_{j'} \approx \frac{1}{4}\mathrm{tr}^2(\mathbf{A}_2-\mathbf{A}_3)\xrightarrow{N\to\infty}0$.

Thus, (A·6) can be proved.

Lemma 5: When $N \to \infty$, the derivations of characteristic of $\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j$ and $\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j$, $j \neq i, i'$, can be expressed as

$$\begin{aligned}
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j\mathbf{h}_s^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_l \xrightarrow[N\to\infty]{a.s.} \begin{cases}\frac{1}{2}, & (s,l)\in\mathcal{L},\\ 0, & \text{else},\end{cases}\\
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j\mathbf{h}_s^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_l \xrightarrow[N\to\infty]{a.s.} \begin{cases}\frac{\gamma_3^{\mathrm{U}}}{2}, & (s,l)\in\mathcal{L},\\ 0, & \text{else},\end{cases}\\
&\frac{1}{N}\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j\mathbf{h}_s^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_l \xrightarrow[N\to\infty]{a.s.} \begin{cases}\frac{\gamma_3^{\mathrm{M}}}{2}, & (s,l)\in\mathcal{L},\\ 0, & \text{else},\end{cases}
\end{aligned} \tag{A·7}$$

where the set $\mathcal{L} = \{(i,j),(i',j'),(j,i),(j',i')\}$.

Proof 8: Let $\mathbf{A}, \mathbf{B}$ represent $\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}$ or $\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}$, and redefine $\mathbf{h}_i = [\mathbf{h}_{Ri}^T, \mathbf{h}_{Ii}^T]^T$ and

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\\ \mathbf{A}_3 & \mathbf{A}_4\end{bmatrix}, \mathbf{B} = \begin{bmatrix}\mathbf{B}_1 & \mathbf{B}_2\\ \mathbf{B}_3 & \mathbf{B}_4\end{bmatrix},$$

where $\mathbf{A}_i, \mathbf{B}_i$ are the diagonal matrices. There are

$$\begin{aligned}
\mathbf{h}_i^T\mathbf{A}\mathbf{h}_j &= \mathbf{h}_{Ri}^T\mathbf{A}_1\mathbf{h}_{Rj}+\mathbf{h}_{Ii}^T\mathbf{A}_3\mathbf{h}_{Rj}+\mathbf{h}_{Ri}^T\mathbf{A}_2\mathbf{h}_{Ij}+\mathbf{h}_{Ii}^T\mathbf{A}_4\mathbf{h}_{Ij},\\
\mathbf{h}_s^T\mathbf{B}\mathbf{h}_l &= \mathbf{h}_{Rs}^T\mathbf{B}_1\mathbf{h}_{Rl}+\mathbf{h}_{Is}^T\mathbf{B}_3\mathbf{h}_{Rl}+\mathbf{h}_{Rs}^T\mathbf{B}_2\mathbf{h}_{Il}+\mathbf{h}_{Is}^T\mathbf{B}_4\mathbf{h}_{Il}.
\end{aligned}$$

For $(s,l) = (i,j)$, since $\mathbf{h}_{Ri}$, $\mathbf{h}_{Ii}$, $\mathbf{h}_{Rj}$, and $\mathbf{h}_{Ij}$ are i.i.d., with the help of the law of large numbers, there is

$$\mathbf{h}_i^T\mathbf{A}\mathbf{h}_j\mathbf{h}_i^T\mathbf{B}\mathbf{h}_j \approx \frac{N}{4}\mathrm{tr}(\mathbf{A}_1\mathbf{B}_1+\mathbf{A}_2\mathbf{B}_2+\mathbf{A}_3\mathbf{B}_3+\mathbf{A}_4\mathbf{B}_4).$$

Based on (A·4) and (A·5), we can obtain that

$$\begin{aligned}
&|\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j|^2 \xrightarrow{N\to\infty} = \frac{N}{2},\\
&|\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j|^2 \xrightarrow{N\to\infty} \frac{N}{2}\mathbb{E}\{1+[\mathbf{a}_1]_n^2+[\mathbf{a}_2]_n^2\},\\
&\mathbf{h}_i^T\mathbf{K}_{\mathrm{D},t}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j\\
&\qquad \xrightarrow{N\to\infty} \frac{N}{2}\mathbb{E}\{[\mathbf{a}_{11}]_n+[\mathbf{a}_{12}]_n[\mathbf{a}_1]_n+[\mathbf{a}_{21}]_n[\mathbf{a}_2]_n\}.
\end{aligned}$$

Taking the expressions of $\mathbf{a}_\cdot$ into the results above, the approximation in (A·7) for the case $(s,l) = (i,j)$ can be acquired.

For $(s,l) = (i',j')$, we can similarly obtain that

$$\mathbf{h}_i^T\mathbf{A}\mathbf{h}_j\mathbf{h}_{i'}^T\mathbf{B}\mathbf{h}_{j'} \approx \frac{N}{4}\mathrm{tr}(\mathbf{A}_1\mathbf{B}_4-\mathbf{A}_2\mathbf{B}_3-\mathbf{A}_3\mathbf{B}_2+\mathbf{A}_4\mathbf{B}_1),$$

which results in the approximation in (A·7) for the case $(s,l) = (i',j')$ by substituting the trace of matrices with the expectations of entries of the corresponding matrices.

For $(s,l) = (j,i)$ and $(s,l) = (j',i')$, they are very close to the cases $(s,l) = (i,j)$ and $(s,l) = (i',j')$, respectively, since $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices or very close to Hermitian matrices.

For $(s,l) \notin \mathcal{L}$, they can be easily approximated to 0 with the law of large numbers. Finally, we can prove the approximation given in (A·7).

Appendix B: Collection of Proofs

B.1 proof of Proposition 1

Since $\mathbf{S}_{\mathrm{tr,c}}\mathbf{S}_{\mathrm{tr,c}}^H = P_{\mathrm{UE}}\mathbf{I}_{2M}, \mathbf{S}_{\mathrm{tr,c}}\mathbf{S}_{\mathrm{tr,c}}^T = \mathbf{0}$, there is $\mathbf{S}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^T = \frac{P_{\mathrm{UE}}}{2}\mathbf{I}_{2M}$, i.e.$(\mathbf{S}_{\mathrm{tr}}\mathbf{S}_{\mathrm{tr}}^T)^{-1} = \frac{2}{P_{\mathrm{UE}}}\mathbf{I}_{2M}$. The element of estimation error matrix can be rewritten as

$$\begin{aligned}
[\mathbf{\Delta}_{\mathrm{eq}}]_{nm} &= \frac{2}{P_{\mathrm{UE}}}[\mathbf{K}_{\mathrm{U},r}]_{n\cdot}\mathbf{V}_{\mathrm{U,tr}}[\mathbf{S}_{\mathrm{tr}}]_{m\cdot}^T\\
&= \frac{2}{P_{\mathrm{UE}}}\left([\mathbf{K}_{\mathrm{U},r}]_{nn}[\mathbf{V}_{\mathrm{U,tr}}]_{n\cdot}+[\mathbf{K}_{\mathrm{U},r}]_{nn'}[\mathbf{V}_{\mathrm{U,tr}}]_{n'\cdot}\right)[\mathbf{S}_{\mathrm{tr}}]_{m\cdot}^T,
\end{aligned} \tag{A·8}$$

where the second equation follows the form of $\mathbf{K}_{\mathrm{U},r}$. The variance of $[\mathbf{\Delta}_{\mathrm{eq}}]_{nm}$, $n = 1,\cdots,2N, m = 1,\cdots,2M$, is acquired as

$$\sigma_{\mathrm{eq},nm}^2 = \frac{\sigma^2}{P_{\mathrm{UE}}}\|[\mathbf{K}_{\mathrm{U},r}]_{n\cdot}\|_F^2, \tag{A·9}$$

since $\mathbb{E}\{[\mathbf{V}_{\mathrm{U,tr}}]_{n\cdot}^T[\mathbf{V}_{\mathrm{U,tr}}]_{n\cdot}\} = \mathbb{E}\{[\mathbf{V}_{\mathrm{U,tr}}]_{n'\cdot}^T[\mathbf{V}_{\mathrm{U,tr}}]_{n'\cdot}\} \approx \frac{\sigma^2}{2}\mathbf{I}_\tau$, $\mathbb{E}\{[\mathbf{V}_{\mathrm{U,tr}}]_{n\cdot}^T[\mathbf{V}_{\mathrm{U,tr}}]_{n'\cdot}\} \approx \mathbf{0}$, and $[\mathbf{S}_{\mathrm{tr}}]_{m\cdot}[\mathbf{S}_{\mathrm{tr}}]_{m\cdot}^T = \frac{P_{\mathrm{UE}}}{2}$.

Thus, we can easily obtain the average estimation error as

$$\sigma_{\mathrm{eq}}^2 = \frac{1}{2N\cdot 2M}\sum_{m=1}^{2M}\sum_{n=1}^{2N}\sigma_{\mathrm{eq},nm}^2.$$

which can be expressed as (12) based on the property that $\|[\mathbf{K}_{\mathrm{U},r}]_{n\cdot}\|_F^2 + \|[\mathbf{K}_{\mathrm{U},r}]_{n'\cdot}\|_F^2 = 2$.

B.2 proof of Proposition 2

Define $\mathbf{H} = [\mathbf{h}_1,\cdots,\mathbf{h}_{2M}]$, $\mathbf{\Delta}_{\mathrm{eq}} = [\boldsymbol{\delta}_{\mathrm{eq},1},\cdots,\boldsymbol{\delta}_{\mathrm{eq},2M}]$. We can obtain that

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{\delta}_{\mathrm{eq},i}\boldsymbol{\delta}_{\mathrm{eq},i}^T\} &= \sigma_{\mathrm{eq}}^2\mathbf{K}_{\mathrm{U},r}\mathbf{K}_{\mathrm{U},r}^T,\\
\mathbb{E}\{\mathbf{K}_{\mathrm{U},r}\mathbf{h}_i\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\} &= \frac{1}{2}\mathbf{K}_{\mathrm{U},r}\mathbf{K}_{\mathrm{U},r}^T,
\end{aligned}$$

which mean that $\mathbf{H}^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{H}$ and $\mathbf{\Delta}_{\mathrm{eq}}^T\mathbf{\Delta}_{\mathrm{eq}}$ can be treated as Wishart matrices.

Using Lemma 1, we collect properties given by

$$\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_i \xrightarrow{N\to\infty} \frac{1}{2}\mathrm{tr}((\mathbf{K}_{\mathrm{U,r}}^T\mathbf{K}_{\mathrm{U,r}})),$$
$$\mathbf{h}_i^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{h}_j \xrightarrow{N\to\infty} 0,$$
$$\boldsymbol{\delta}_{\mathrm{eq},i}^T\boldsymbol{\delta}_{\mathrm{eq},i} \xrightarrow{N\to\infty} \mathrm{tr}((\mathbf{K}_{\mathrm{U,r}}^T\mathbf{K}_{\mathrm{U,r}}))\sigma_{\mathrm{eq}}^2,$$
$$\boldsymbol{\delta}_{\mathrm{eq},i}^T\boldsymbol{\delta}_{\mathrm{eq},j} \xrightarrow{N\to\infty} 0.$$

Based on the estimator of the equivalent channel, $\hat{\mathbf{H}}_{\mathrm{U,eq}} = \mathbf{H}_{\mathrm{U,eq}} + \boldsymbol{\Delta}_{\mathrm{eq}}$. With the help of the properties above and the properties of the Wishart matrix [23], there is

$$\begin{aligned}&\mathrm{tr}\left((\hat{\mathbf{H}}_{\mathrm{U,eq}}^T\hat{\mathbf{H}}_{\mathrm{U,eq}})^{-1}\right)\\ &= \mathrm{tr}\left((\mathbf{K}_{\mathrm{U},t}^T\mathbf{H}^T\mathbf{K}_{\mathrm{U},r}^T\mathbf{K}_{\mathrm{U},r}\mathbf{H}\mathbf{K}_{\mathrm{U},t} + \boldsymbol{\Delta}_{\mathrm{eq}}^T\boldsymbol{\Delta}_{\mathrm{eq}})^{-1}\right)\\ &\approx \frac{2N}{N-M}\frac{\mathrm{tr}\left((\mathbf{K}_{\mathrm{U,t}}^T\mathbf{K}_{\mathrm{U,t}} + 2\sigma_{\mathrm{eq}}^2\mathbf{I}_{2M})^{-1}\right)}{\mathrm{tr}(\mathbf{K}_{\mathrm{U,r}}^T\mathbf{K}_{\mathrm{U,r}})}.\end{aligned}$$

By combining the formula with the expression of $\mathbf{K}_{\mathrm{U,t}}$ and $\mathbf{K}_{\mathrm{U,r}}$, (17) can be achieved.

B.3 Proof of Proposition 3

From (18), the expression of the matrix $\mathbf{G}$ can be rewritten as

$$\mathbf{G} = \mathbf{K}_{\mathrm{D},r}\tilde{\mathbf{I}}_{2M}\check{\mathbf{G}}(\mathbf{K}_{\mathrm{U},t}^T)^{-1}\tilde{\mathbf{I}}_{2M}, \tag{A·10}$$

where

$$\mathbf{H}_{\mathrm{D}} = \mathbf{H}^T\mathbf{K}_{\mathrm{D},t}^T, \mathbf{H}_{\mathrm{U}} = \mathbf{K}_{\mathrm{U},r}\mathbf{H},$$
$$\check{\mathbf{G}} = \mathbf{H}_{\mathrm{D}}(\mathbf{H}_{\mathrm{U}} + \boldsymbol{\Delta}_{\mathrm{eq}}\mathbf{K}_{\mathrm{U},t}^{-1})(\mathbf{H}_{\mathrm{U}}^T\mathbf{H}_{\mathrm{U}} + (\mathbf{K}_{\mathrm{U},t}^T)^{-1}\boldsymbol{\Delta}_{\mathrm{eq}}^T\boldsymbol{\Delta}_{\mathrm{eq}}\mathbf{K}_{\mathrm{U},t}^{-1})^{-1}.$$

Redefine matrices

$$\mathbf{H}_{\mathrm{D}} = [\mathbf{h}_{\mathrm{D},1}^T, \cdots, \mathbf{h}_{\mathrm{D},2M}^T]^T, \mathbf{H}_{\mathrm{U}} = [\mathbf{h}_{\mathrm{U},1}, \cdots, \mathbf{h}_{\mathrm{U},2M}],$$
$$\hat{\mathbf{H}}_{\mathrm{U}} = \mathbf{H}_{\mathrm{U}} + \boldsymbol{\Delta}_{\mathrm{eq}}\mathbf{K}_{\mathrm{U},t}^{-1} = [\hat{\mathbf{h}}_{\mathrm{U},1}, \cdots, \hat{\mathbf{h}}_{\mathrm{U},2M}],$$
$$\hat{\mathbf{H}}_{\mathrm{chg}}^{ij} = [\hat{\mathbf{h}}_{\mathrm{U},1}, \cdots, \hat{\mathbf{h}}_{\mathrm{U},j-1}, \mathbf{h}_{\mathrm{D},i}, \hat{\mathbf{h}}_{\mathrm{U},j+1}, \cdots, \hat{\mathbf{h}}_{\mathrm{U},2M}],$$
$$\hat{\mathbf{H}}_{\mathrm{del}}^{j} = [\hat{\mathbf{h}}_{\mathrm{U},1}, \cdots, \hat{\mathbf{h}}_{\mathrm{U},j-1}, \hat{\mathbf{h}}_{\mathrm{U},j+1}, \cdots, \hat{\mathbf{h}}_{\mathrm{U},2M}].$$

With the help of Lemma 2, the entries of the matrix $\check{\mathbf{G}}$ in (A·10) can be expressed as

$$[\check{\mathbf{G}}]_{ij} = \frac{\det(\hat{\mathbf{H}}_{\mathrm{chg}}^{ij\,T}\hat{\mathbf{H}}_{\mathrm{U}})}{\det(\hat{\mathbf{H}}_{\mathrm{U}}^T\hat{\mathbf{H}}_{\mathrm{U}})} = \frac{\det\left(\left[\boldsymbol{\alpha}_i, \boldsymbol{\Phi}_{\mathrm{del}}^j\right]\right)}{\det\left(\left[\boldsymbol{\phi}_j, \boldsymbol{\Phi}_{\mathrm{del}}^j\right]\right)}, \tag{A·11}$$

where $\boldsymbol{\alpha}_i^T = \mathbf{h}_{\mathrm{D},i}^T\hat{\mathbf{H}}_{\mathrm{U}} - \delta(i,j)\hat{\mathbf{h}}_{\mathrm{U},i}^T\hat{\mathbf{H}}_{\mathrm{U}}, \boldsymbol{\phi}_j^T = \hat{\mathbf{h}}_{\mathrm{U},j}^T\hat{\mathbf{H}}_{\mathrm{U}}, \boldsymbol{\Phi}_{\mathrm{del}}^{j\,T} = \hat{\mathbf{H}}_{\mathrm{del}}^{j\,T}\hat{\mathbf{H}}_{\mathrm{U}}$ and

$$\delta(i,j) = \begin{cases} 0, & i = j \\ 1, & \text{else,}\end{cases}$$

which will not change the determinant of $\hat{\mathbf{H}}_{\mathrm{chg}}^{ij\,T}\hat{\mathbf{H}}_{\mathrm{U}}$ since $\boldsymbol{\phi}_i$ is a column of $\boldsymbol{\Phi}_{\mathrm{del}}^j$ if $i \neq j$.

According to the properties of block matrices, there are

$$\det^2\left(\left[\boldsymbol{\alpha}_i, \boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\right]\right) = \det\left(\begin{bmatrix}\boldsymbol{\alpha}_i^T\boldsymbol{\alpha}_i & \boldsymbol{\alpha}_i^T\boldsymbol{\Phi}_{\mathrm{del}}^j\\ \boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\alpha}_i & \boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\end{bmatrix}\right)$$
$$= \beta_{ij}\det\left(\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\right),$$

and

$$\det^2\left(\left[\hat{\boldsymbol{\phi}}_j^T, \hat{\boldsymbol{\Phi}}_{\mathrm{del}}^{j\,T}\right]\right) = \det\left(\begin{bmatrix}\boldsymbol{\phi}_j^T\boldsymbol{\phi}_j & \boldsymbol{\phi}_j^T\boldsymbol{\Phi}_{\mathrm{del}}^j\\ \boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\phi}_j & \boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\end{bmatrix}\right)$$
$$= \beta_j\det\left(\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\right),$$

where

$$\beta_{ij} = \boldsymbol{\alpha}_i^T\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^T\boldsymbol{\Phi}_{\mathrm{del}}^j\left(\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\right)^{-1}\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\alpha}_i,$$
$$\beta_j = \boldsymbol{\phi}_j^T\boldsymbol{\phi}_j - \boldsymbol{\phi}_j^T\boldsymbol{\Phi}_{\mathrm{del}}^j\left(\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\right)^{-1}\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\phi}_j.$$

When $N \to \infty$, we can get approximations as

$$\lim_{N\to\infty}\frac{1}{N}\hat{\mathbf{h}}_{\mathrm{U},i}^T\hat{\mathbf{h}}_{\mathrm{U},i} = 1 + 2\eta_i\sigma_{\mathrm{eq}}^2, i = 1, \cdots, 2M,$$
$$\lim_{N\to\infty}\frac{1}{N}\hat{\mathbf{h}}_{\mathrm{U},i}^T\hat{\mathbf{h}}_{\mathrm{U},j} = 0, i \neq j, i, j = 1, \cdots, 2M.$$

The approximation of $\boldsymbol{\Phi}_{\mathrm{del}}^j\left(\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\right)^{-1}\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}$ is given as

$$\lim_{N\to\infty}\boldsymbol{\Phi}_{\mathrm{del}}^j\left(\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T}\boldsymbol{\Phi}_{\mathrm{del}}^j\right)^{-1}\boldsymbol{\Phi}_{\mathrm{del}}^{j\,T} = \mathbf{I}_{2M}^{0,j},$$

where $\mathbf{I}_{2M}^{0,j}$ stands for the identity matrix whose jth diagonal element equals to 0, and there are

$$\beta_{ij} \approx \boldsymbol{\alpha}_i^T\mathbf{I}_{2M}^{1,j}\boldsymbol{\alpha}_i = |\mathbf{h}_{\mathrm{D},i}^T\hat{\mathbf{h}}_{\mathrm{U},j} - \delta(i,j)\hat{\mathbf{h}}_{\mathrm{U},i}^T\hat{\mathbf{h}}_{\mathrm{U},j}|^2,$$
$$\beta_j \approx \boldsymbol{\phi}_j^T\mathbf{I}_{2M}^{1,j}\boldsymbol{\phi}_j = |\hat{\mathbf{h}}_{\mathrm{U},j}^T\hat{\mathbf{h}}_{\mathrm{U},j}|^2.$$

where $\mathbf{I}_{2M}^{1,j} = \mathbf{I}_{2M} - \mathbf{I}_{2M}^{0,j}$. Thus, the square of $[\check{\mathbf{G}}]_{ij}$ can be approximated as

$$[\check{\mathbf{G}}]_{ij}^2 = \frac{\det^2\left(\left[\boldsymbol{\alpha}_i, \boldsymbol{\Phi}_{\mathrm{del}}^j\right]\right)}{\det^2\left(\left[\boldsymbol{\phi}_j, \boldsymbol{\Phi}_{\mathrm{del}}^j\right]\right)} \approx \frac{|\mathbf{h}_{\mathrm{D},i}^T\hat{\mathbf{h}}_{\mathrm{U},j} - \delta(i,j)\hat{\mathbf{h}}_{\mathrm{U},i}^T\hat{\mathbf{h}}_{\mathrm{U},j}|^2}{|\hat{\mathbf{h}}_{\mathrm{U},j}^T\hat{\mathbf{h}}_{\mathrm{U},j}|^2}. \tag{A·12}$$

Since entries of $\mathbf{H}$ and $\boldsymbol{\Delta}_{\mathrm{eq}}$ are independent, when $N \to \infty$, based on Lemma 3, Lemma 4 and Lemma 5, we obtain approximations as

$$|\hat{\mathbf{h}}_{\mathrm{U},i}^T\hat{\mathbf{h}}_{\mathrm{U},j}|^2 \xrightarrow[N\to\infty]{a.s.} \begin{cases} N[(N+1)(1+2\eta_j\sigma_{eq}^2)^2 + 4\eta_j\sigma_{\mathrm{eq}}^2\gamma_3^{\mathrm{U}}], & j = i,\\ N[\gamma_2^{\mathrm{U}} + (\eta_i + \eta_j + 2\eta_i\eta_j\sigma_{\mathrm{eq}}^2)\gamma_3^{\mathrm{U}}\sigma_{\mathrm{eq}}^2], & j = i',\\ N[\gamma_3^{\mathrm{U}}/2 + (\eta_i + \eta_j + 2\eta_i\eta_j\sigma_{\mathrm{eq}}^2)\gamma_3^{\mathrm{U}}\sigma_{\mathrm{eq}}^2], & \text{else,}\end{cases}$$

$$|\mathbf{h}_{\mathrm{D},i}^T\mathbf{h}_{\mathrm{U},j}|^2 \xrightarrow[N\to\infty]{a.s.} \begin{cases} N(N+1)\gamma_1^{\mathrm{D}}, & j = i,\\ N\gamma_2^{\mathrm{D}}, & j = i',\\ N/2, & \text{else,}\end{cases}$$

Applying Lemma 4 and Lemma 5, there are

$$|\mathbf{h}_{\mathrm{D},i}^T\hat{\mathbf{h}}_{\mathrm{U},j}|^2 \approx |\mathbf{h}_{\mathrm{D},i}^T\mathbf{h}_{\mathrm{U},j}|^2+\eta_j\sigma_{\mathrm{eq}}^2,$$

$$\mathbf{h}_{\mathrm{D},i}^T\hat{\mathbf{h}}_{\mathrm{U},j}\hat{\mathbf{h}}_{\mathrm{U},i}^T\hat{\mathbf{h}}_{\mathrm{U},j} \approx \begin{cases} N\gamma_2^{\mathrm{M}}, & j=i', \\ N\gamma_3^{\mathrm{M}}(1+4\eta_j\sigma_{\mathrm{eq}}^2)/2, & \text{else}, \end{cases}$$

Based on the approximations above, $[\check{\mathbf{G}}]_{ij}^2$ in (A·12) can be further approximated as

$$\left[\check{\mathbf{G}}\right]_{ij}^2 \xrightarrow[N\to\infty]{a.s.} \frac{\bar{D}_{ij}+\tilde{D}_{ij}}{\hat{D}_j}.$$

Here $\mathbf{G}=\mathbf{K}_{\mathrm{D},r}\tilde{\mathbf{I}}_{2M}\check{\mathbf{G}}(\mathbf{K}_{\mathrm{U},t}^T)^{-1}\tilde{\mathbf{I}}_{2M}$ is used to calculate the gain of signal, and $[\mathbf{G}]_{ii}$ is determined by the approximations of $[\check{\mathbf{G}}]_{ii}$, $[\check{\mathbf{G}}]_{ii'}$, $[\check{\mathbf{G}}]_{i'i}$, and $[\check{\mathbf{G}}]_{i'i'}$. Taking terms $[\check{\mathbf{G}}]_{ii}$, $[\check{\mathbf{G}}]_{ii'}$ for example, even though the cross term $[\check{\mathbf{G}}]_{ii}[\check{\mathbf{G}}]_{ii'}$ is not equal to $\sqrt{[\check{\mathbf{G}}]_{ii}[\check{\mathbf{G}}]_{ii'}}$, the difference between them can be neglected, since the law of large numbers can achieve an accurate approximation of $[\check{\mathbf{G}}]_{ii}$ when the number of BS antennas goes to infinity. Thus, the approximation of matrix of $\sqrt{[\check{\mathbf{G}}]_{ij}^2}$ is denoted as $\mathbf{T}$, and then the approximation of the power of the interference from the jth branch at the ith branch can be approximated as $G_{ij}=[\mathbf{K}_{\mathrm{D},r}\tilde{\mathbf{I}}_{2M}\mathbf{T}(\mathbf{K}_{\mathrm{U},t}^T)^{-1}\tilde{\mathbf{I}}_{2M}]_{ij}^2$ given in (20).

Nana Zhang received the B.S. degree in electronic and communication engineering from the University of Science and Technology of China, Hefei, China, in 2013. She is currently pursuing the Ph.D. degree in wireless communications with the University of Science and Technology of China, Hefei, China. Her research interests include massive MIMO and millimeter wave systems.

Huarui Yin received the B.S. degree and the Ph.D. degree in electronic engineering and information science from the University of Science and Technology of China (USTC), Hefei, Anhui, in 1996 and 2006, respectively. Since 2010, he has been with the Department of Electronic Engineering and Information Science, USTC, as an Associate Professor. His research interests include digital signal processing, low complexity receiver design, multiple access for massive connections, and throughput analysis of wireless networks.

Weidong Wang received the B.S. degree from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1989, and the M.S. degree from the University of Science and Technology of China, Hefei, China, in 1993. He is currently a Full Professor with the Department of Electronic Engineering and Information Systems, University of Science and Technology of China, Hefei, China. His research interests include wireless communication, microwave and millimeter wave, and radar technology. Prof. Wang is a member of the Committee of Optoelectronic Technology, Chinese Society of Astronautics.

Suhua Tang received the B.S. degree in Electronic Engineering in 1998 and the Ph.D. degree in Information and Communication Engineering in 2003, both from University of Science and Technology of China. From Oct. 2003 to Mar. 2014, he was with Adaptive Communications Research Laboratories, ATR, Japan. Since Apr. 2014, he is with Graduate School of Informatics and Engineering, The University of Electro-Communications, Japan, where he is an associate professor now. His research interests include green communications, ad hoc and sensor networks, inter-vehicle communications and high precision positioning. He is a member of IEEE, IEICE and IPSJ.