IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 26, 2017, accepted March 22, 2017, date of publication April 3, 2017, date of current version May 17, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2690439

Channel Estimation and IQ Imbalance Compensation for Uplink Massive MIMO Systems With Low-Resolution ADCs

YOUZHI XIONG[1], NING WEI[2], ZHONGPEI ZHANG[2], BINRUI LI[2], AND YANG CHEN[2]
[1]National Key Laboratory of Science and Technology on Communication, University of Electronic Science and Technology of China, Chengdu 610051, China
[2]University of Electronic Science and Technology of China, Chengdu 610051, China

Corresponding author: Ning Wei (wn@uestc.edu.cn)

This work was supported in part by the National High Technology Development 863 Program of China under Grant 2014AA01A704 and in part by the National Science Foundation of China under Grant 61571003.

ABSTRACT In this paper, two techniques to compensate inphase/quadrature-phase imbalance (IQI) are investigated in the uplink-quantized massive multiple-input and multiple-output systems for different models of randomized IQI parameters. One is referred to as combined-signal-based channel estimation and compensation (CCEC) and the other is denoted by effective channel estimation and compensation (ECEC). First, an independent automatic gain control (AGC) scheme is proposed to calibrate the dynamic range of both the I branch and the Q branch. By doing that, different quantization steps are used for analog-to-digital converters following the AGCs in these two branches at each receive antenna. Second, considering the impacts of both quantization and IQI, we give the details of channel estimation and IQI compensation for both the CCEC and the ECEC using bilinear generalized approximate message passing (Bi-GAMP). Moreover, to exploit the Bi-GAMP for ECEC reasonably, we theoretically derive the probability density function (pdf) of the elements in the effective channel for the case where only RX IQI is considered. Furthermore, we extend the ECEC to the case where both RX IQI and TX IQI are incorporated into the systems and derive the similar pdf as well. Finally, we use the numerical results to testify the validity of our theoretical analysis and the fact that the analytic PDF can be approximated by a Gaussian distribution when the IQI parameters are relatively small. Compared with other classical methods, the proposed methods can obtain better performance based on the Monte Carlo simulation results.

INDEX TERMS Massine MIMO, IQ imbalance, low-resolution ADCs, AGC, Bi-GAMP.

I. INTRODUCTION

Being a promising candidate technology for fifth-generation (5G) communication systems, massive multiple-input and multiple-output (MIMO) is capable of providing more spatial degrees of freedom (DoF) and high transmission rate by deploying hundreds of antennas at the base station (BS) [1]–[5]. Accordingly, from the implementation perspective, the same number of analog-to-digital converter (ADC) pairs are needed for these antennas. Meanwhile, it has been demonstrated in [6] that ADCs with more output quantize bits consume more power. To reduce the power consumption surge, as mentioned in [7], it is anticipated that the use of low-resolution ADCs is economical and in accordance with the concept of green-communication for current massive MIMO systems. However, radio frequency (RF) impairments, which are inevitable in practical systems, result in unexpected in-phase/quadrature-phase imbalance (IQI) and degrade the performance of baseband signal processing [8], [9]. To be concrete, as mentioned in [10], the impacts of IQI are one of the main problems in quantized massive MIMO systems [11], [12]. Assuming that the same quantization step is used for both in-phase and quadrature-phase branches, the branch with much lower average amplitude suffers more drastic distortion. Moreover, the phase mismatch between the real and the imaginary parts may invert the sign of the signals. Therefore, it is necessary for the quantized massive MIMO to compensate the impacts of IQI, which motivates us to construct the work in this paper. In the following IQ imbalance and IQ mismatch are used interchangeably.

A. RELATED WORK

Massive MIMO with one-bit ADCs was first introduced in [13] where the maximum ratio combination (MRC) and the zero forcing (ZF) detectors were exploited. As an extension of this work, [14] discussed the performance in terms

2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

of achievable rate for the high-order modulations with low resolution ADCs. Although the MRC and the ZF detectors are concise for implementation, their performance is inferior to the recent techniques proposed in [7] where a data-aided channel estimation method is used to achieve a better performance. The performance analysis for massive MIMO systems including low-resolution ADCs or one-bit ADCs was given in [15]–[19]. Low-resolution ADCs were recently introduced to millimeter wave communication systems in [20]–[23] and to wideband MIMO orthogonal frequency-division multiplexing (MIMO-OFDM) systems in [24]–[26]. To compensate the performance loss due to the use of low-resolution ADCs, [10], [26]–[28] offered a mixed-ADCs architecture where a few low-resolution ADCs are replaced by the high-resolution ADCs. However, these works did not investigated the impacts of IQI and assumed that the channel state information (CSI) is exactly known at the receiver. In fact, it is difficult to acquire the perfect CSI in practical systems, particularly in the systems with low resolution ADCs. To this end, [7], [15], [22], [29]–[31] discussed the channel estimation when low resolution ADCs were considered. To the best of our knowledge, [32]–[34] investigated the downlink massive MIMO with low resolution ADCs. Gokceoglu *et al.* in [32] and [33] studied the spatio-temporal waveform design for the downlink massive MIMO and MISO with 1-bit receivers. With low resolution ADCs, hybrid precoding schemes were explored in [34].

Concerning IQI, investigated early in [35] and [36] for MIMO-OFDM communication systems, recent work can be found in [11], [12], [37], and [38], where some advanced IQI compensation methods were studied whereas the quantization was not involved. Zarei *et al.* [11] developed an IQI aware widely linear minimum mean square error (LMMSE) receiver for channel state information (CSI) acquisition and data detection in a single-carrier uplink multi-cell massive MIMO systems. To theoretically analyze the impacts of IQI parameters on the sum rate of systems, the authors assumed that these IQI parameters are known at the receiver side. Kolomvakis *et al.* [12] provided a novel low complexity pilot-based estimator and IQI compensation scheme. Considering both IQI parameters and CSI, a joint estimation method was established in [39] and [40]. Supposing that both the IQI and the low-resoution ADCs are taken into account at the transmitter (i.e.,BS), Wang and Zhang in [41] incorporated a perfect specific transceiver near the transmitter to estimate the IQI parameters, which demands additional time overhead for transmitting training sequences between the specific transceiver and the BS. Additionally, IQI impacts on the precoding and beamforming technologies were examined in [42]–[44].

B. CONTRIBUTIONS

Considering the aforementioned works, in this paper, the channel estimation and the IQI compensation are studied for quantized massive MIMO systems. Toward this objective, two methods for channel estimation and IQI compensation are investigated. They are referred to as the combined-signal-based channel estimation and compensation (CCEC) and the effective channel estimation and compensation (ECEC), respectively. The CCEC estimates the CSI and performs data detection using the bilinear generalized approximate message passing (Bi-GAMP) algorithm, described in [45], by combining the received signal with its conjugated version. On the other hand, the ECEC estimates the effective channel and performs data detection by applying the Bi-GAMP algorithm with the derived probability density function (PDF). To make the Bi-GAMP algorithm work reasonably for the ECEC, we also theoretically derive the PDF of the components in the effective channel. In contrast to [7], where perfect CSI and IQI-free case were presumed, we remove these assumptions and consider a more general case in this paper. Based on the work in [11], [12], and [37], we further take the impacts of quantization errors into account and compare the proposed methods with the schemes mentioned in there. In summary, our contributions in this paper are explained as follows.

- An independent automatic gain control (AGC) scheme for each receive antenna is proposed. For the sake of clarity, we further elaborate on the implementation of the independent AGC scheme. Compared to the common AGC used in [7], the proposed AGC scheme has advantages to ease the degradation caused by the amplitude imbalance between the I branch and the Q branch, particularly when the amplitude imbalance is severe. The main reason is that the ADCs at the I branch and the Q branch can efficiently use the limited quantization bits by adopting the different quantization steps.
- Two efficient channel estimation and IQI compensation methods are constructed for quantized massive MIMO systems. In addition, a new pilot matrix is designed to avoid the signal self-interference arising from IQI during the channel training stage. To guarantee the performance of channel estimation, a modified Bi-GAMP algorithm, which incorporates the impacts of quantization errors, is utilized. For the case where only RX IQI is considered, the CCEC method is proposed to complete channel estimation and data detection by combining the received signal with its conjugated version before performing the CCEC. Although the combined signal can remove the impacts of IQI, it brings a problem that only the real part of the received signal is exploited. To fully use the received signal, the effective channel, in which the IQI parameters and the CSI are included, can be estimated using the modified Bi-GAMP on condition that the PDF of the elements in the effective channel is known. We then theoretically derive this PDF and introduce the ECEC method when RX IQI is considered. For the case where both RX IQI and TX IQI are incorporated, we also derive the similar PDF and extend the ECEC using the modified Bi-GAMP algorithm.
- The numerical results are used to certify the validity of our analytic results. Based on these results, the theoretical PDF can be approximated by a Gaussian distribution

when the IQI parameters are relatively small. For the case where the IQI parameters are large, Laplace distribution is suitable for approximation. To demonstrate the performance improvement of the proposed methods, we compare them with other techniques used in the recent works through the Monte-Carlo simulation. By doing so, from the viewpoint of bit error ratio (BER) and normalized mean square error (NMSE), the proposed methods can obtain better performance.

1) ORGANIZATION

The rest of the paper is organized as follows. We introduce the system model of quantized massive MIMO with IQI in Section II. Our proposed independent AGC and quantization scheme are discussed in Section III. The techniques of channel estimation and IQI compensation are described in Section IV for the case where only RX IQI is considered. For the case where both RX IQI and TX IQI are considered, the methods of channel estimation and IQI compensation are described in Section V. Simulation results are shown in Section VI, and the conclusion and future work are presented in Section VII.

2) NOTATION

Throughout this paper, boldface lower and upper case letters indicate column vectors and matrices, respectively. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. $\mathbf{0}_{N\times K}$ denotes the $N \times K$ all-zero matrix. diag$\{a_1, \cdots, a_N\}$ is a diagonal matrix with $a_1, \cdots, a_N$ on the main diagonal. $\mathbf{A}^H$, $\mathbf{A}^T$, $\mathbf{A}^*$ and $\mathbf{A}^{-1}$ denote the conjugate transpose, the transpose, the conjugate and the inverse of $\mathbf{A}$, respectively. In addition, a_{ij} is the entry at the ith row and the jth column of $\mathbf{A}$. We denote by $\mathbf{a}_n$ the nth row vector of $\mathbf{A}$. $\lceil a \rceil$ denotes the minimal integer no less than a. $\Re(a)$ and $\Im(a)$ indicate the real and the imaginary parts of a, respectively. $\mathbb{E}(\xi)$ and $\mathbb{V}(\xi)$ represent the expectation and variance over a random variable ξ, respectively.

II. SYSTEMS MODEL

To model the IQI at the transmitter and/or the receiver, an asymmetrical fashion described in [46] is used. This model states that the mismatches of amplitude and phase appear at the Q branch while the I branch works in an ideal behavior. As considered in [11], it is assumed that the IQI parameters are invariant during one transmission block and vary among different blocks. Throughout this paper, we focus on the single-carrier uplink single-cell massive MIMO scenario, as illustrated in Fig. 1, in which K single-antenna users simultaneously deliver their data to the BS deploying N antennas. At the kth transmitter, $\phi_{k,\mathrm{T}}$ and $g_{k,\mathrm{T}}$ are used to denote the phase and amplitude mismatches, respectively. As shown in Fig. 2, $\phi_{n,\mathrm{R}}$ and $g_{n,\mathrm{R}}$ represent the phase and amplitude mismatches at receive antenna n in the similar way. Apparently, we have $\phi_{k,\mathrm{T}} = 0$ and $g_{k,\mathrm{T}} = 1$ for perfect IQ matching at transmitter k along with $\phi_{n,\mathrm{R}} = 0$ and $g_{n,\mathrm{R}} = 1$ for perfect IQ matching at receive antenna n.

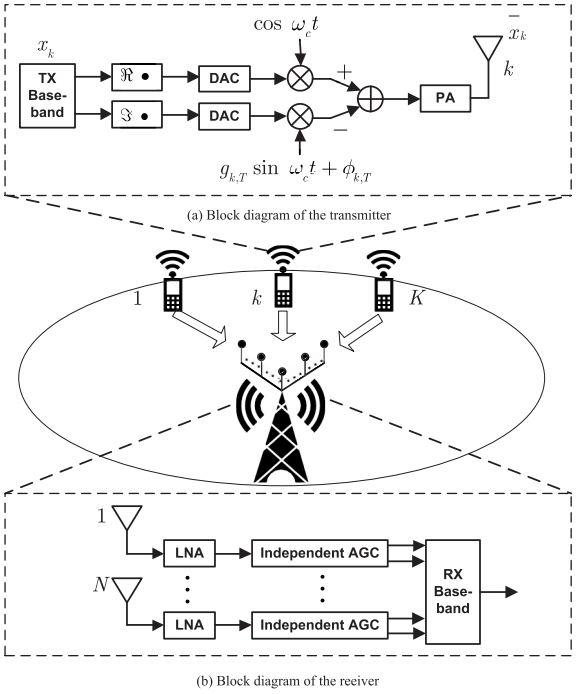


FIGURE 1. Block diagram of the transmitter and receiver affected by IQ imbalance. The independent AGCs and low-resolution ADCs are involved.

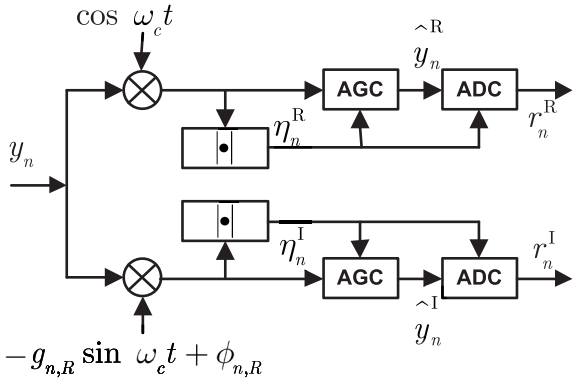


FIGURE 2. Block diagram of the independent AGC for the nth receive antenna. Note that $|\bullet|$ indicates the envelop of signals. The I and Q branch are working simultaneously with perfect synchronization.

A. IQ IMBALANCE AT THE TRANSMITTER

In this part, the signal expression corrupted by the IQI is given at the transmitter according to the procedures showing at the top of Fig. 1. The data expected to be transmitted by user k are denoted by $x_k[i]$ at time index i. $x_k[i]$ satisfies $\mathbb{E}[x_k[i]] = 0$ and $\mathbb{E}[|x_k[i]|^2] = 1$ and is randomly selected from a constellation denoted by $\mathcal{A}$ in which each point comprises Q bits. As proving in [46], when TX IQI is considered at user k, the transmitted signal $\bar{x}_k[i]$ can be obtained according to

$$\bar{x}_k[i] = G_{1,k}x_k[i] + G_{2,k}x_k[i]^*, \tag{1}$$

where $G_{1,k} = (1+g_{k,T}e^{\mathrm{j}\phi_{k,\mathrm{T}}})/2$ and $G_{2,k} = (1-g_{k,\mathrm{T}}e^{\mathrm{j}\phi_{k,\mathrm{T}}})/2$ with $g_{k,\mathrm{T}}$ and $\phi_{k,\mathrm{T}}$ being the amplitude and phase imbalances

at user k, respectively. In this paper, we assume that $g_{k,\mathrm{T}}$ follows the uniform $\mathcal{U}(1-g_{\max,\mathrm{T}}, 1+g_{\max,\mathrm{T}})$ and $\phi_{k,\mathrm{T}}$ follows the uniform $\mathcal{U}(-\phi_{\max,\mathrm{T}}, \phi_{\max,\mathrm{T}})$ since Zarei *et al.* [11], Wang and Zhang [41], and Hakkarainen *et al.* [44] chose the values of $\phi_{k,\mathrm{T}}$ and $g_{k,\mathrm{T}}$ in a proper range randomly and uniformly. By stacking $\bar{x}_k[i]$ together for all users, we have

$$\bar{\mathbf{x}}[i] = \begin{bmatrix}\bar{x}_1[i] & \bar{x}_2[i] & \cdots & \bar{x}_K[i]\end{bmatrix}^T = \mathbf{G}_1\mathbf{x}[i] + \mathbf{G}_2\mathbf{x}^*[i], \tag{2}$$

where

$$\mathbf{x}[i] = \begin{bmatrix}x_1[i] & x_2[i] & \cdots & x_K[i]\end{bmatrix}^T,$$
$$\mathbf{G}_1 = \mathrm{diag}\{G_{1,1}, G_{1,2}, \cdots, G_{1,K}\}$$

and

$$\mathbf{G}_2 = \mathrm{diag}\{G_{2,1}, G_{2,2}, \cdots, G_{2,K}\},$$

from which we have the property $\mathbf{G}_1 + \mathbf{G}_2 = \mathbf{I}_K$.

B. IQ IMBALANCE AT THE RECEIVER

If all users deliver the data simultaneously, $\bar{\mathbf{x}}[i]$ will be transmitted through the wireless channel denoted by $\mathbf{H} \in \mathbb{C}^{N\times K}$. When RX IQI is not involved, the received signal at the BS can be expressed as

$$\mathbf{y}[i] = \mathbf{H}\bar{\mathbf{x}}[i] + \mathbf{w}[i], \tag{3}$$

where $\mathbf{H} = \begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_K\end{bmatrix}$, assumed to be flat-fading over one block, denotes the channel matrix with $\mathbf{h}_k = \check{\mathbf{R}}_k\mathbf{v}_k \in \mathbb{C}^{N\times 1}$ being the channel vector from the kth user to the BS [11]. The components of $\mathbf{v}_k$ are mutually independent and identically distributed (i.i.d.) random variables characterized by the complex gaussian distribution $\mathcal{CN}(0, 1)$ and $\mathbf{R}_k = \mathbb{E}\{\mathbf{h}_k\mathbf{h}_k^H\} = \check{\mathbf{R}}_k\check{\mathbf{R}}_k^H$ represents the channel covariance matrix [11]. Here we let $\mathbf{R}_k = \mathbf{I}_N$ for convenience. The additive white gaussian noise (AWGN) is denoted by $\mathbf{w}[i]$, whose elements independently follow the distribution $\mathcal{CN}(0, \sigma^2)$.

If the RF front-ends at the BS are also inaccurate, in the time-domain, the received baseband signal can be modeled as [11]

$$\begin{aligned}\bar{\mathbf{y}}[i] &= \mathbf{K}_1\mathbf{y}[i] + \mathbf{K}_2\mathbf{y}^*[i] \\ &= (\mathbf{K}_1\mathbf{H}\mathbf{G}_1 + \mathbf{K}_2\mathbf{H}^*\mathbf{G}_2^*)\mathbf{x}[i] \\ &\quad + (\mathbf{K}_1\mathbf{H}\mathbf{G}_2 + \mathbf{K}_2\mathbf{H}^*\mathbf{G}_1^*)\mathbf{x}^*[i] \\ &\quad + (\mathbf{K}_1\mathbf{w}[i] + \mathbf{K}_2\mathbf{w}[i]^*)\end{aligned} \tag{4}$$

where

$$\mathbf{K}_1 = \mathrm{diag}\{K_{1,1}, K_{1,2}, \cdots, K_{1,N}\},$$
$$\mathbf{K}_2 = \mathrm{diag}\{K_{2,1}, K_{2,2}, \cdots, K_{2,N}\}$$

with $K_{1,n} = (1+g_{n,\mathrm{R}}e^{-\mathrm{j}\phi_{n,\mathrm{R}}})/2$ and $K_{2,n} = (1-g_{n,\mathrm{R}}e^{\mathrm{j}\phi_{n,\mathrm{R}}})/2$ for $n = 1, 2, \cdots, N$. Similarly, $g_{n,\mathrm{R}}$ and $\phi_{n,\mathrm{R}}$ denote the amplitude and phase imbalances at the nth antenna of the BS, respectively. In this paper, it is also assumed that $g_{k,\mathrm{R}}$ has the uniform distribution over $\mathcal{U}(1 - g_{\max,\mathrm{R}}, 1 + g_{\max,\mathrm{R}})$ and $\phi_{k,\mathrm{R}}$ has the uniform distribution over $\mathcal{U}(-\phi_{\max,\mathrm{R}}, \phi_{\max,\mathrm{R}})$ [11], [41], [44]. From the expressions of $\mathbf{K}_1$ and $\mathbf{K}_2$, we have $\mathbf{K}_1 + \mathbf{K}_2^* = \mathbf{I}_N$. Using the augmented representation given in [11], (4) can be rewritten as

$$\tilde{\mathbf{y}}[i] = \Phi\tilde{\mathbf{H}}\Xi\tilde{\mathbf{x}}[i] + \Phi\tilde{\mathbf{w}}[i], \tag{5}$$

where $\tilde{\mathbf{y}}[i] = \begin{bmatrix}\Re(\bar{\mathbf{y}}[i])^T & \Im(\bar{\mathbf{y}}[i])^T\end{bmatrix}^T$, $\tilde{\mathbf{x}}[i] = \begin{bmatrix}\Re(\mathbf{x}[i])^T & \Im(\mathbf{x}[i])^T\end{bmatrix}^T$, $\tilde{\mathbf{w}}[i] = \begin{bmatrix}\Re(\mathbf{w}[i])^T & \Im(\mathbf{w}[i])^T\end{bmatrix}^T$ and the representations of Φ, $\tilde{\mathbf{H}}$ and Ξ are, respectively, given by

$$\Phi = \begin{bmatrix}\Re(\mathbf{K}_1 + \mathbf{K}_2) & \Im(\mathbf{K}_2 - \mathbf{K}_1) \\ \Im(\mathbf{K}_1 + \mathbf{K}_2) & \Re(\mathbf{K}_1 - \mathbf{K}_2)\end{bmatrix}, \tag{6a}$$

$$\tilde{\mathbf{H}} = \begin{bmatrix}[r]\Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H})\end{bmatrix} \tag{6b}$$

and

$$\Xi = \begin{bmatrix}[r]\Re(\mathbf{G}_1 + \mathbf{G}_2) & \Im(\mathbf{G}_2 - \mathbf{G}_1) \\ \Im(\mathbf{G}_1 + \mathbf{G}_2) & \Re(\mathbf{G}_1 - \mathbf{G}_2)\end{bmatrix}. \tag{6c}$$

It can be seen from (4) that the received signal $\bar{\mathbf{y}}[i]$ is impacted by $\mathbf{x}^*[i]$ interpreted as the self-interference of $\mathbf{x}[i]$. In the frequency-domain, this self-interference causes the mirror interference expected to be suppressed. Therefore, to ease the effects of the self-interference is one of the motivations of this paper, which will be studied in the succeeding sections.

In fact, although a narrowband system is considered in this paper, the succeeding investigations can be extended to a wideband system such as the MIMO-OFDM system. Because the signal model of (5.43) in [46] has the same form as (4) except that k in (5.43) denotes the index of subcarrier, it is worth extending our work to a broadband system if the quantization is considered. Therefore, this realistic system with low-resolution ADCs and IQ imbalance is left for the future work.

III. INDEPENDENT AGC AND QUANTIZATION USING FEW BITS

IQI can cause some problems for quantization. In [7], where IQI is not incorporated, a common AGC is used to control the signal amplitude before the signal is divided into I branch and Q branch. However, due to the IQI, particularly when the amplitude imbalance is severe, using the same dynamic range for both I branch and Q branch (e.g., using the common AGC in [7]) may cause some unexpected problems. When the amplitude imbalance is severe, i.e., $g_{n,\mathrm{R}} \ll 1$ (or $g_{n,\mathrm{R}} \gg 1$), it is easy to find that the average magnitude of the quadrature-phase is much less (or larger) than that of the in-phase. It has been mentioned that using low-resolution ADCs in massive MIMO systems is an efficient way to reduce power consumption. Therefore, adopting the same quantization step as the I branch, the Q branch cannot distinguish the signals using limited quantization bits. In other words, the equivalent quantization bits are inadequate for the Q branch.

To handle this problem caused by amplitude imbalance, an independent AGC scheme is proposed. As depicted in Fig. 2, two independent and separated AGCs are used for I branch and Q branch at each receive antenna. Due to the

clear analogy in signal processing between the I branch and the Q branch, we take the I branch as an example to show the processes of AGC and ADC. In order to calibrate the dynamic range, a power tracking method, which is carried out before the AGC and ADC at the I branch, is used to calculate the signal power within a time interval and control the output of the AGC accordingly. To this end, a predefined time window with length L is incorporated to calculate the average power of the signal (i.e., $\Re(\bar{y}_n[l])$) within the time window at the I branch. Therefore, the moving-average power can be calculated as

$$P_n^{\mathrm{R}}[i] = \begin{cases} \frac{1}{i}\sum_{l=0}^{i-1} |\Re(\bar{y}_n[l])|^2, & \text{if } i \le L \\ \frac{1}{L}(LP_n^{\mathrm{R}}[i-1] + |\Re(\bar{y}_n[i])|^2 \\ \quad - |\Re(\bar{y}_n[i-L])|^2), & \text{otherwise.} \end{cases} \tag{7}$$

We then use the *three-sigma* rule described in [47] to calculate the scaling factor η_n^{R} as

$$\eta_n^{\mathrm{R}}[i] = 3\sqrt{P_n^{\mathrm{R}}[i]}, \tag{8}$$

with which the AGC can control the dynamic range of the down-converted signal. Considering the scaling factor at all antennas, the scaling factor matrix can be written as

$$\Omega^{\mathrm{R}}[i] = \mathrm{diag}\left\{\eta_1^{\mathrm{R}}[i], \eta_2^{\mathrm{R}}[i], \cdots, \eta_N^{\mathrm{R}}[i]\right\}. \tag{9}$$

Similarly, $\Omega^{\mathrm{I}}[i]$ for the Q branch can be obtained by calculating $P_n^{\mathrm{I}}[i]$, $n = 1, 2, \cdots, N$. Next, the scaled signals at the I branches for all antennas are achieved by

$$\hat{\mathbf{y}}^{\mathrm{R}}[i] = \left(\Omega^{\mathrm{R}}[i]\right)^{-1} \bar{\mathbf{y}}^{\mathrm{R}}[i]. \tag{10}$$

It should be pointed out that the values of elements in $\hat{\mathbf{y}}^{\mathrm{R}}[i]$ are distributed in the range $[-1, 1]$ with high probability before quantization.

Finally, the quantization of scaled signal is carried out by the ADCs at the I branch. The quantized signal, which is fed forward to the baseband digital signal processing (DSP) at the receiver, is expressed as

$$\mathbf{r}^{\mathrm{R}}[i] = \Omega^{\mathrm{R}}[i]\mathcal{Q}\left(\hat{\mathbf{y}}^{\mathrm{R}}[i]\right), \tag{11}$$

where $\mathcal{Q}(\cdot)$ is the element-wise quantization operator on a floating-point number. For the Q branch, we also have $\mathbf{r}^{\mathrm{I}}[i] = \Omega^{\mathrm{I}}[i]\mathcal{Q}\left(\check{\mathbf{y}}^{\mathrm{I}}[i]\right)$ which is also fed forward to the baseband DSP at the receiver. Based on $\mathbf{r}^{\mathrm{R}}[i]$ and $\mathbf{r}^{\mathrm{I}}[i]$, the complex signal in the baseband DSP yields

$$\mathbf{r}[i] = \mathbf{r}^{\mathrm{R}}[i] + \mathrm{j}\mathbf{r}^{\mathrm{I}}[i]. \tag{12}$$

In this paper, a uniform scalar quantization with step Δ is considered. Supposing that two real-valued b-bit ADCs are used for each receive antenna, the quantization step Δ is equal to 2^{-b}. For a real-valued ADC and corresponding input $\hat{y}$, the final quantized output can be modeled as

$$r = \eta\mathcal{Q}(\hat{y}) = \begin{cases} \eta\left\lceil \frac{\hat{y}}{\Delta} - \frac{1}{2} \right\rceil \Delta, & \text{if } -\frac{2^b\Delta}{2} \le \hat{y} \le \frac{2^b\Delta}{2}, \\ \frac{\eta 2^b\Delta}{2}, & \text{if } \hat{y} > \frac{2^b\Delta}{2}, \\ -\frac{\eta 2^b\Delta}{2}, & \text{if } \hat{y} < -\frac{2^b\Delta}{2}. \end{cases} \tag{13}$$

Correspondingly, the upper and lower quantization boundaries of r in (13) can be determined by a de-quantization operation

$$\mathcal{Q}^{-1}(r) = \begin{cases} \left(r - \frac{\eta\Delta}{2}, +\infty\right), & \text{if } r = r_{\max}, \\ \left(-\infty, -r + \frac{\eta\Delta}{2}\right), & \text{if } r = r_{\min}, \\ \left(r - \frac{\eta\Delta}{2}, r + \frac{\eta\Delta}{2}\right), & \text{otherwise,} \end{cases} \tag{14}$$

where $r_{\max}$ and $r_{\min}$ are the maximum and the minimum quantization levels, respectively. Eq. (14) will be used in (21) later. In this paper, we assume that all antennas adopt identical low-resolution ADCs with $b \in \{1, 2, 3, 4\}$, which is much more efficient in terms of cost and power consumption as compared to the conventional massive MIMO systems equipped with high-resolution ADCs.

In addition, to make analysis more convenient, we adopt the additive quantization noise model (AQNM) which is extensively employed in quantized MIMO systems [17], [18], [23]. Recalling the received signal model in (4) and (5), it is easy to find that they are consistent with the signal model described in [17], and then the corresponding quantized signals can be given by

$$\mathbf{r}[i] = \alpha\bar{\mathbf{y}}[i] + \mathbf{e}[i] \tag{15}$$

and

$$\tilde{\mathbf{r}}[i] = \alpha\tilde{\mathbf{y}}[i] + \tilde{\mathbf{e}}[i], \tag{16}$$

where $\tilde{\mathbf{r}}[i] = \left[\Re(\mathbf{r}[i])^T \quad \Im(\mathbf{r}[i])^T\right]^T$ belongs to $\mathbb{C}^{2N\times 1}$ and $\tilde{\mathbf{e}}[i] = \left[\Re(\mathbf{e}[i])^T \quad \Im(\mathbf{e}[i])^T\right]^T \in \mathbb{C}^{2N\times 1}$, which denotes the additive quantization errors vector uncorrelated with $\tilde{\mathbf{y}}[i]$. The covariance matrix of $\tilde{\mathbf{e}}[i]$, as in [17], is characterized by

$$\Sigma_{\tilde{\mathbf{e}}[i]\tilde{\mathbf{e}}[i]} = \alpha(1-\alpha)\mathrm{diag}\{\mathbb{E}\{(\Phi\tilde{\mathbf{H}}\Xi\Xi^H\tilde{\mathbf{H}}^H\Phi^H + \frac{\sigma^2}{2}\mathbf{I}_{2N})\}\}. \tag{17}$$

Here it is worth noticing that α is a coefficient whose value is related to the quantization bit and given by Table 1 when $b \le 5$. For $b > 5$, the value of α is determined

TABLE 1. α **for different quantization bits ($b \le 5$).**

b	1	2	3	4	5
α	0.6366	0.8825	0.96546	0.990503	0.997501

by $\alpha \approx 1 - \frac{\pi\sqrt{3}}{2}2^{-2b}$ according to [17], [18] and [23]. From here on, we will omit the time index for simplicity.

IV. CHANNEL ESTIMATION AND DATA DETECTION UNDER RX IQI AND QUANTIZATION ERRORS

Based on the discussions in [12] and [41], where the RF front-ends at the users side are assumed to be perfect, one system only experiencing RX IQI is considered in this section. By doing so, the acquisition of CSI and IQI parameters are investigated under quantization errors. First a pilot matrix scheme is designed for channel estimation. To compensate the impacts of IQI at the receiver, the CCEC method is proposed based on the classical Bi-GAMP algorithm. Finally, the proposed ECEC method is presented to obtain better performance.

A. PILOT MATRIX DESIGN

For any coherent detection, the CSI needs to be estimated at the receiver. Considering the existence of IQI, the predefined pilots should be able to estimate the unknown CSI and IQI parameters. Supposing that $\mathbf{x}_k \in \mathbb{C}^{2K\times 1}$ denotes the pilot sequence at the first $2K$ symbols of one block for user k, we let $\mathbf{X}_\mathrm{p} = \begin{bmatrix}\mathbf{x}_1^T & \cdots & \mathbf{x}_K^T\end{bmatrix}^T \in \mathbb{C}^{K\times 2K}$ be the pilot matrix which satisfies $\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^H = \mathbf{I}_K$ and $\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^T = \mathbf{0}_K$ to guarantee the orthogonality of pilots among different users. Besides, the self-interference is expected to be suppressed during the channel training stage. For this purpose, we suggest a specific pilot matrix

$$\mathbf{X}_\mathrm{p} = \begin{bmatrix}\mathbf{X}_{1,\mathrm{p}} & \mathbf{X}_{2,\mathrm{p}}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_K/\sqrt{2} & \mathrm{j}\mathbf{I}_K/\sqrt{2}\end{bmatrix}. \tag{18}$$

Next, the CSI and IQI parameters estimation are discussed based on the designed pilot matrix and its received version.

B. CHANNEL ESTIMATION AND DATA DETECTION USING COMBINED SIGNALS

Since TX IQI is not considered, we have $\mathbf{G}_1 = \mathbf{I}_K$ and $\mathbf{G}_2 = \mathbf{0}_K$. As stated above, $\mathbf{X}_\mathrm{p}$ is used for training at the beginning of one block. Meanwhile, with L being the data length, $\mathbf{X}_\mathrm{d} \in \mathbb{C}^{K\times L}$ is used to denote the data matrix following the pilot matrix $\mathbf{X}_\mathrm{p}$. Therefore, the signal to be transmitted within one block can be denoted by $\mathbf{X} = \begin{bmatrix}\mathbf{X}_\mathrm{p} & \mathbf{X}_\mathrm{d}\end{bmatrix}$. Considering the IQ mismatch, noise and quantization errors, the corresponding received signal $\mathbf{R} = \begin{bmatrix}\mathbf{R}_\mathrm{p} & \mathbf{R}_\mathrm{d}\end{bmatrix} \in \mathbb{C}^{N\times(2K+L)}$ is given by

$$\mathbf{R} = \alpha\left(\mathbf{K}_1\mathbf{H}\mathbf{X} + \mathbf{K}_2\mathbf{H}^*\mathbf{X}^* + \mathbf{K}_1\mathbf{W} + \mathbf{K}_2\mathbf{W}^*\right) + \mathbf{E}, \tag{19}$$

where $\mathbf{W} \in \mathbb{C}^{N\times(2K+L)}$ and $\mathbf{E} \in \mathbb{C}^{N\times(2K+L)}$ denote the noise and quantization errors over one block, respectively.

Using the property that $\mathbf{K}_1 + \mathbf{K}_2^* = \mathbf{I}_N$, the combined signal $\mathbf{R}^c$ can be obtained by adding the quantized signal $\mathbf{R}$ to its conjugated version $\mathbf{R}^*$ and is given by

$$\begin{aligned}\mathbf{R}^c &= \mathbf{R} + \mathbf{R}^* = \alpha(\mathbf{H}\mathbf{X} + \mathbf{H}^*\mathbf{X}^*) + 2\Re(\alpha\mathbf{W} + \mathbf{E}) \\ &= \alpha\tilde{\mathbf{H}}\tilde{\mathbf{X}} + 2\Re(\alpha\mathbf{W} + \mathbf{E}),\end{aligned} \tag{20}$$

where $\tilde{\mathbf{H}} = \begin{bmatrix}2\Re(\mathbf{H}) & 2\Im(-\mathbf{H})\end{bmatrix} \in \mathbb{C}^{N\times 2K}$ and $\tilde{\mathbf{X}} = \begin{bmatrix}\Re(\mathbf{X})^T & \Im(\mathbf{X})^T\end{bmatrix}^T \in \mathbb{C}^{2K\times(2K+L)}$. When combined signal $\mathbf{R}^c$ is used, it is obvious from (20) that the effects of IQI vanish but at the cost of the imaginary part loss in quantized signal. Additionally, if combined signal $\mathbf{R}^c$ is considered, only the real parts of noise and quantization errors remain. It is worth noting that the imaginary part $(\mathbf{R} + \mathbf{R}^*)$ cannot be directly used because the IQI is coupled in it. According to the fact that all components in $\mathbf{H}$ are selected following a complex Gaussian distribution $\mathcal{CN}(0, 1)$, it can be concluded that all elements in $\tilde{\mathbf{H}}$ have a real Gaussian distribution $\mathcal{N}(0, 1)$. In addition, supposing that a square QAM constellation with M points is used for $\mathbf{X}$, the elements in $\tilde{\mathbf{X}}$ can be regarded as points in the $\sqrt{M}$-PAM constellation denoted by $\mathcal{A}^p$.

It is easy to find that (20) is consistent with the generalized bilinear problem described in [45] except for quantization errors which can be modeled as additive white noise [48]. Hence, we can modify the Bi-GAMP algorithm therein to perform the channel estimation and data detection at the BS side. The details of the modified algorithm are described in Algorithm 1, which is referred to as the proposed CCEC when the combined signal is used.

In Algorithm 1, t and $T_{\max}$ denote the current iteration index and the maximum number of iterations, respectively. The value of $T_{\max}$ is 100 in [7]. In fact, the Bi-GAMP algorithm can converge after 20-30 iterations although it shows a slow convergence at low SNR regime. ϵ represents the stopping condition based on the normalized change in the residual. And $\beta \in (0, 1]$ is a damping factor used for avoiding the divergence of this algorithm. The value of β is 0.3 in [49]. Here, we provide some explanations about this algorithm and omit the iteration index for brevity.

First, the operations in lines 9-10 calculate the posterior mean and variance of $z_{n,l}$. Referring to [10] and operation (14), we have

$$\hat{z}_{n,l} = \frac{\int_{Q^{-1}(r^c_{n,l})} z_{n,l}\mathcal{CN}(z_{n,l}; \hat{p}_{n,l}, v^p_{n,l})\mathrm{d}\, z_{n,l}}{\int_{Q^{-1}(r^c_{n,l})} \mathcal{CN}(z_{n,l}; \hat{p}_{n,l}, v^p_{n,l})\mathrm{d}\, z_{n,l}}, \tag{21a}$$

$$v^z_{n,l} = \frac{\int_{Q^{-1}(r^c_{n,l})} |z_{n,l} - \hat{z}_{n,l}|^2\mathcal{CN}(z_{n,l}; \hat{p}_{n,l}, v^p_{n,l})\mathrm{d}\, z_{n,l}}{\int_{Q^{-1}(r^c_{n,l})} \mathcal{CN}(z_{n,l}; \hat{p}_{n,l}, v^p_{n,l})\mathrm{d}\, z_{n,l}}, \tag{21b}$$

by which $v^s_{n,l}$ and $\hat{s}_{n,l}$ in lines 11-14 can be obtained. Considering that the integrations in (21) make the computation more complex, some simplifications are needed for practical implementation. If regarding the quantization errors as AWGN and referring to [45], we can get

$$\hat{s}_{n,l} = \frac{r^c_{n,l} - \hat{p}_{n,l}}{v^p_{n,l} + \gamma}, \tag{22a}$$

$$v^s_{n,l} = \frac{1}{v^p_{n,l} + \gamma}, \tag{22b}$$

where $\gamma = 4\left(0.5\alpha^2\sigma^2 + \sigma_q^2\right)$ and σ_q^2 denotes the variance of quantization error. Without integration, the computational complexity of (22) is reduced significantly.

Algorithm 1 Bi-GAMP-Based Channel Estimation and Data Detection

Input: Compensated observations $\mathbf{R}^{\mathrm{c}}$, pilots $\tilde{\mathbf{X}}_{\mathrm{p}}$

Output: $\hat{\tilde{\mathbf{H}}}$ and $\hat{\tilde{\mathbf{X}}}_{\mathrm{d}}$

1: **Initialize:** β, ϵ; $\forall n, l$: $\hat{s}_{n,l}(0) = 0$, $v^s_{n,l}(0) = 0$, $\bar{v}^p_{n,l}(0) = 0$, $v^p_{n,l}(0) = 0$; $\sum_k \triangleq \sum_{k=1}^{2K}$, $\sum_n \triangleq \sum_{n=1}^{N}$, $\sum_l \triangleq \sum_{l=1}^{2K+L}$;
$\forall n, k, l$: $\bar{x}_{k,l}(0) = 0$, $\hat{\tilde{x}}_{k,l}(0) = 0$, $v^{\tilde{x}}_{k,l}(0) = 1$, $\tilde{h}_{k,l}(0) = 0$, $\hat{\tilde{h}}_{k,l}(0) = 10^{-5}$, $v^{\tilde{h}}_{k,l}(0) = 1$
2: **for** $t = 1$ **to** $T_{\max}$ **do**
3: $\forall n, l: \bar{v}^p_{n,l}(t) = \sum_k |\hat{\tilde{h}}_{n,k}(t)|^2 v^x_{k,l}(t) + v^{\tilde{h}}_{n,k}(t)|\hat{x}_{k,l}(t)|^2$
4: $\forall n, l: \bar{v}^p_{n,l}(t) = \beta\alpha^2 \bar{v}^p_{n,l}(t) + (1-\beta)\bar{v}^p_{n,l}(t-1)$
5: $\forall n, l: \bar{p}_{n,l}(t) = \alpha \sum_k \hat{\tilde{h}}_{n,k}(t)\hat{\tilde{x}}_{k,l}(t)$
6: $\forall n, l: v^p_{n,l}(t) = \bar{v}^p_{n,l}(t) + \alpha^2 \sum_k v^{\tilde{h}}_{n,k}(t) v^{\tilde{x}}_{k,l}(t)$
7: $\forall n, l: v^p_{n,l}(t) = \beta v^p_{n,l}(t) + (1-\beta) v^p_{n,l}(t-1)$
8: $\forall n, l: \hat{p}_{n,l}(t) = \bar{p}_{n,l}(t) - \hat{s}_{n,l}(t-1)\bar{v}^p_{n,l}(t)$
9: $\forall n, l: \hat{z}_{n,l}(t) = \mathbb{E}\{z_{n,l}(t)|\hat{p}_{n,l}(t), v^p_{n,l}(t) + \sigma^2\}$
10: $\forall n, l: v^z_{n,l}(t) = \mathbb{V}\{z_{n,l}(t)|\hat{p}_{n,l}(t), v^p_{n,l}(t) + \sigma^2\}$
11: $\forall n, l: v^s_{n,l}(t) = (1 - v^z_{n,l}(t)/(v^p_{n,l}(t) + \sigma^2))/(v^p_{n,l}(t) + \sigma^2)$
12: $\forall n, l: v^s_{n,l}(t) = \beta v^s_{n,l}(t) + (1-\beta) v^s_{n,l}(t-1)$
13: $\forall n, l: \hat{s}_{n,l}(t) = \left(\hat{z}_{n,l}(t) - \hat{p}_{n,l}(t)\right)/(v^p_{n,l}(t) + \sigma^2)$
14: $\forall n, l: \hat{s}_{n,l}(t) = \beta \hat{s}_{n,l}(t) + (1-\beta)\hat{s}_{n,l}(t-1)$
15: $\forall n, k: \bar{h}_{n,k}(t) = \beta \hat{\tilde{h}}_{n,k}(t) + (1-\beta)\bar{h}_{n,k}(t-1)$
16: $\forall k, l: \bar{x}_{k,l}(t) = \beta \hat{\tilde{x}}_{k,l}(t) + (1-\beta)\bar{x}_{k,l}(t-1)$
17: $\forall k, l: v^r_{k,l}(t) = \left(\alpha^2 \sum_n |\bar{h}_{n,k}(t)|^2 v^s_{n,l}(t)\right)^{-1}$
18: $\forall k, l: \zeta = \bar{x}_{k,l}(t)\left(1 - \alpha^2 v^r_{k,l}(t) \sum_n v^{\tilde{h}}_{n,k}(t) v^s_{n,l}(t)\right)$
19: $\forall k, l: \hat{r}_{k,l}(t) = \zeta + \alpha v^r_{k,l}(t) \sum_n \bar{h}^*_{n,k}(t)\hat{s}_{n,l}(t)$
20: **if** $l > 2K$ **then**
21: $\forall k, l: \hat{\tilde{x}}_{k,l}(t+1) = \mathbb{E}\{\tilde{x}_{k,l}(t)|\hat{r}_{k,l}(t), v^r_{k,l}(t)\}$
22: $\forall k, l: v^{\tilde{x}}_{k,l}(t+1) = \mathbb{V}\{\tilde{x}_{k,l}(t)|\hat{r}_{k,l}(t), v^r_{k,l}(t)\}$
23: **else**
24: $\forall k, l: v^{\tilde{x}}_{k,l}(t+1) = 0; \quad \hat{\tilde{x}}_{k,l}(t) = \tilde{x}_{\mathrm{p},k,l}$
25: **end if**
26: $\forall n, k: v^q_{n,k}(t) = \left(\sum_l |\hat{\tilde{x}}_{k,l}(t+1)|^2 v^s_{n,l}(t)\right)^{-1}$
27: $\forall n, k: \zeta = \alpha \bar{h}_{n,k}(t)\left(1 - v^q_{n,k}(t) \sum_l v^{\tilde{x}}_{k,l}(t+1) v^s_{n,l}(t)\right)$
28: $\forall n, k: \hat{q}_{n,k}(t) = \zeta + v^q_{n,k}(t) \sum_l \hat{\tilde{x}}^*_{k,l}(t+1)\hat{s}_{n,l}(t)$
29: $\forall n, k: \hat{\tilde{h}}_{n,k}(t+1) = \mathbb{E}\{\tilde{h}_{n,k}(t)|\hat{q}_{n,k}(t), v^q_{n,k}(t)\}$
30: $\forall n, k: v^{\tilde{h}}_{n,k}(t+1) = \mathbb{V}\{\tilde{h}_{n,k}(t)|\hat{q}_{n,k}(t), v^q_{n,k}(t)\}$
31: **if** $(\sum_n \sum_l |\bar{p}_{n,l}(t) - \bar{p}_{n,l}(t-1)|^2)/\sum_n \sum_l |\bar{p}_{n,l}(t)|^2 < \epsilon$ **then**
32: stop
33: **end if**
34: **end for**

Next, the operations in lines 21-22 are taken with regard to the posterior probability

$$\mathcal{P}(\tilde{x}_{k,l}) = \frac{\mathcal{CN}(\tilde{x}_{k,l}; \hat{r}_{k,l}, v^r_{k,l})}{\sum_{\tilde{x}'_{k,l} \in \mathcal{A}^p} \mathcal{CN}(\tilde{x}'_{k,l}; \hat{r}_{k,l}, v^r_{k,l})}. \tag{23}$$

Assuming that the square QAM constellation is used, the posterior mean and variance of $\tilde{x}_{k,l}$ can be given by

$$\hat{\tilde{x}}_{k,l} = \sum_{\tilde{x}_{k,l} \in \mathcal{A}^p} \tilde{x}_{k,l} \mathcal{P}(\tilde{x}_{k,l}), \tag{24a}$$

$$v^{\tilde{x}}_{k,l} = \sum_{\tilde{x}_{k,l} \in \mathcal{A}^p} |\tilde{x}_{k,l} - \hat{\tilde{x}}_{k,l}|^2 \mathcal{P}(\tilde{x}_{k,l}). \tag{24b}$$

Finally, assuming that the elements in $\tilde{\mathbf{H}}$ follow $\mathcal{N}(0, \sigma_h^2)$ and referring to Section III-C in [45], the operations in lines 29-30 can be simplified as

$$\hat{\tilde{h}}_{k,l} = \frac{\hat{q}_{n,k}\sigma_h^2}{v^q_{n,k} + \sigma_h^2}, \tag{25a}$$

$$v^{\tilde{h}}_{k,l} = \frac{v^q_{n,k}\sigma_h^2}{v^q_{n,k} + \sigma_h^2}. \tag{25b}$$

We notice that the pilot matrix $\tilde{\mathbf{X}}_{\mathrm{p}}$ is known during the training stage, which differs from [45] in which pilots do not exists in $\mathbf{X}$. Therefore, the corresponding variances for $\tilde{\mathbf{X}}_{\mathrm{p}}$ are zero, which means $v^{\tilde{x}}_{k,l} = 0$ for $l \leq 2K$. Additionally, based on [45], the effects of quantization are incorporated in (21) or (22). This modified algorithm can be regarded as a data-aided channel estimation which further used to recover the unknown data.

As a comparison, we introduce another way to estimate the channel information and IQI parameters, which is mentioned in our previous work in [50]. Based on the received training sequences in (20) and LMMSE technique, the BS first acquire the CSI using the combined signal received in training stage. Next, the BS estimates the IQ parameters via Least Square (LS) technique using the estimated CSI and local pilot $\mathbf{X}_{\mathrm{p}}$. Finally, on the basis of estimated CSI and IQ parameters, the IQ compensation and data detection are carried out at the BS via LMMSE. The similar work can also be found in [12] for massive MIMO systems without quantization errors.

C. PROPOSED EFFECTIVE CHANNEL ESTIMATION AND COMPENSATION WITH IQI AT THE RECEIVER

As stated previously, the CCEC only exploits the real part of the received signal, i.e., only $\mathbf{R}^c = 2\Re(\mathbf{R})$ is used. To fully exploit the received signal, in this subsection, an ECEC method is proposed to estimate the effective channel and recover the data. Using the property that $\mathbf{K}_1 + \mathbf{K}_2^* = \mathbf{I}_N$, Φ in (6a) can be rewritten as

$$\Phi = \begin{bmatrix} \Re(\mathbf{K}_1 + \mathbf{K}_2) & \Im(\mathbf{K}_2 - \mathbf{K}_1) \\ \Im(\mathbf{K}_1 + \mathbf{K}_2) & \Re(\mathbf{K}_1 - \mathbf{K}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_N \\ \mathbf{A} & \mathbf{B} \end{bmatrix}, \tag{26}$$

where $\mathbf{A} = \mathrm{diag}\{a_1, a_2, \cdots, a_N\}$ and $\mathbf{B} = \mathrm{diag}\{b_1, b_2, \cdots, b_N\}$ with $a_i = g_{i,\mathrm{R}} \sin(\phi_{i,\mathrm{R}})$ and $b_i = g_{i,\mathrm{R}} \cos(\phi_{i,\mathrm{R}})$ for $i = 1, 2, \cdots, N$. Then the effective channel is defined as

$$\mathbf{H}^{\mathrm{eff}} = \Phi\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \mathbf{H}^{\mathrm{eff}}_{11} & \mathbf{H}^{\mathrm{eff}}_{12} \\ \mathbf{H}^{\mathrm{eff}}_{21} & \mathbf{H}^{\mathrm{eff}}_{22} \end{bmatrix}, \tag{27}$$

where $\mathbf{H}_{11}^{\text{eff}} = \Re(\mathbf{H})$, $\mathbf{H}_{12}^{\text{eff}} = -\Im(\mathbf{H})$, $\mathbf{H}_{21}^{\text{eff}} = \mathbf{A}\Re(\mathbf{H}) + \mathbf{B}\Im(\mathbf{H})$, and $\mathbf{H}_{22}^{\text{eff}} = \mathbf{B}\Re(\mathbf{H}) - \mathbf{A}\Im(\mathbf{H})$. From $\mathbf{H}^{\text{eff}}$, we find that the elements in $\mathbf{H}_{11}^{\text{eff}}$ and $\mathbf{H}_{12}^{\text{eff}}$ independently follow the real Gaussian distribution $\mathcal{N}(0, 0.5)$ according to the fact that the elements in $\mathbf{H}$ independently have the complex Gaussian distribution $\mathcal{CN}(0, 1)$.

We subsequently determine the PDF of the elements in $\mathbf{H}_{21}^{\text{eff}}$ and $\mathbf{H}_{22}^{\text{eff}}$. Without loss of generality, the arbitrary entry $y_{i,j}$ in $\mathbf{H}_{21}^{\text{eff}}$ and $z_{i,j}$ in $\mathbf{H}_{22}^{\text{eff}}$ are considered because all elements in $\mathbf{H}_{11}^{\text{eff}}$ or $\mathbf{H}_{12}^{\text{eff}}$ are random variables. Due to the similarity between $y_{i,j}$ and $z_{i,j}$, only the PDF of $y_{i,j}$ is discussed. If $h_{i,j}^{\text{R}}$ and $h_{i,j}^{\text{I}}$ denote the real and imaginary parts of $h_{i,j}$ located at the ith row and the jth column of $\mathbf{H}$, $y_{i,j}$ is expressed as

$$y_{i,j} = a_i h_{i,j}^{\text{R}} + b_i h_{i,j}^{\text{I}}. \tag{28}$$

Lemma 1: The PDF of $y_{i,j}$ is independent of the phase mismatch $\phi_{i,\text{R}}$ and is determined by the distribution of $g_{i,\text{R}}$, $h_{i,j}^{\text{R}}$ and $h_{i,j}^{\text{I}}$. Furthermore, the analytic PDF of $y_{i,j}$ is given by

$$f_{y_{i,j}}(y_{i,j}) = \begin{cases} \dfrac{1}{4\sqrt{\pi}\, g_{\max,\text{R}}} \left[\Gamma\left(0, \dfrac{y_{i,j}^2}{g_{\text{U,R}}^2}\right) - \Gamma\left(0, \dfrac{y_{i,j}^2}{g_{\text{L,R}}^2}\right) \right], & \text{if } y_{i,j} \neq 0 \\ \dfrac{1}{2\sqrt{\pi}\, g_{\max,\text{R}}} \ln\left(\dfrac{g_{\text{U,R}}}{g_{\text{L,R}}}\right), & \text{otherwise,} \end{cases} \tag{29}$$

where $g_{\text{L,R}} \triangleq (1 - g_{\max,\text{R}})$ and $g_{\text{U,R}} \triangleq (1 + g_{\max,\text{R}})$.

Proof: Please refer to Appendix VII. ■

In the similar way, we can also conclude that $f_{z_{i,j}}(z_{i,j})$ is independent of the phase mismatch $\phi_{i,\text{R}}$ and is given by

$$f_{z_{i,j}}(z_{i,j}) = \frac{1}{2g_{\max,\text{R}}} \int_{g_{\text{L,R}}}^{g_{\text{U,R}}} \frac{1}{\sqrt{\pi g_{i,\text{R}}^2}} \exp(-\frac{z_{i,j}^2}{g_{i,\text{R}}^2}) \mathrm{d}\, g_{i,\text{R}}, \tag{30}$$

which can be further simplified to get the same result as (29).

Lemma 2: The elements in $\mathbf{H}_{21}^{\text{eff}}$ are mutually independent and follow the same distribution. The same conclusion can be obtained for the elements in $\mathbf{H}_{22}^{\text{eff}}$. Moreover, the arbitrary element in $\mathbf{H}_{21}^{\text{eff}}$ is independent of the arbitrary element in $\mathbf{H}_{22}^{\text{eff}}$.

Proof: Please refer to Appendix VII. ■

The complicated expression in (29) is not tractable and requires simplifying. For this purpose, given that $g_{\max,\text{R}}$ is relatively small (such as $g_{\max,\text{R}} < 0.2$), (44) can approximately be simplified as

$$\begin{aligned} f_{y_{i,j}}(y_{i,j}) &= \frac{1}{2\sqrt{\pi}\, g_{\max,\text{R}}} \int_{y_{i,j}/g_{\text{L,R}}}^{y_{i,j}/g_{\text{U,R}}} \frac{1}{t} \exp(-t^2) \mathrm{d}\, t \\ &\overset{(a)}{\approx} \frac{1}{2\sqrt{\pi}\, g_{\max,\text{R}}} \left(\frac{y_{i,j}}{g_{\text{L,R}}} - \frac{y_{i,j}}{g_{\text{U,R}}}\right) \left(\frac{y_{i,j}}{1 - g_{\max,\text{R}}^2}\right)^{-1} \\ &\quad \times \exp\left(-\frac{y_{i,j}^2}{(1 - g_{\max,\text{R}}^2)^2}\right) \\ &= \frac{1}{\sqrt{\pi}} \exp\left(-\frac{y_{i,j}^2}{(1 - g_{\max,\text{R}}^2)^2}\right) \\ &\overset{(b)}{\approx} \frac{1}{\sqrt{\pi}(1 - g_{\max,\text{R}}^2)} \exp\left(-\frac{y_{i,j}^2}{(1 - g_{\max,\text{R}}^2)^2}\right), \end{aligned} \tag{31}$$

from which we can see that $f_{y_{i,j}}(y_{i,j})$ approximately approaches a Gaussian distribution with zero mean and variance $(1-g_{\max,\text{R}}^2)^2/2$. Notice that (a) is obtained using a Taylor series expansion at $t_0 = \frac{y_{i,j}/g_{\text{L,R}} - y_{i,j}/g_{\text{U,R}}}{2}$ and (b) is based on $1 - g_{\max,\text{R}}^2 \approx 1$ when $g_{\max,\text{R}}$ is relatively small.

For the case where $g_{\max,R}$ is large, the Laplace distribution is suitable to approximate the PDF of $y_{i,j}$ (or $z_{i,j}$). Due to the fact that $\mathbb{E}[h_{i,j}^{\text{R}}] = 0$ along with the independence between a_i and $h_{i,j}^{\text{R}}$, we have $\mathbb{E}[y_{i,j}] = 0$ and

$$\mathbb{V}[y_{i,j}] = \mathbb{E}[y_{i,j}^2] = \mathbb{E}[a_i^2 (h_{i,j}^{\text{R}})^2] = \frac{1}{2}\left(1 + \frac{g_{\max,\text{R}}^2}{3}\right). \tag{32}$$

Then $f_{y_{i,j}}(y_{i,j})$ can be approximated by

$$f_{y_{i,j}}(y_{i,j}) \approx \frac{1}{\sqrt{2\mathbb{V}[y_{i,j}]}} \exp\left(-\frac{\sqrt{2}|y_{i,j}|}{\sqrt{\mathbb{V}[y_{i,j}]}}\right), \tag{33}$$

which will be verified with numerical results in Section VI.

Considering the effective channel $\mathbf{H}^{\text{eff}}$ and the input-output relationship over one block, (16) yields

$$\tilde{\mathbf{R}} = \alpha \mathbf{H}^{\text{eff}} \tilde{\mathbf{X}} + \alpha \Phi \tilde{\mathbf{W}} + \tilde{\mathbf{E}}, \tag{34}$$

where $\tilde{\mathbf{R}} \in \mathbb{C}^{2N \times (2K+L)}$ and $\tilde{\mathbf{E}} \in \mathbb{C}^{2N \times (2K+L)}$ are constructed by $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{e}}$ over one block, respectively. The effective noise defined as $\tilde{\mathbf{w}}^{\text{eff}} = \alpha \Phi \tilde{\mathbf{w}}$, where $\tilde{\mathbf{w}}$ is any column of $\tilde{\mathbf{W}}$, has the variance

$$\begin{aligned} \Sigma_{\tilde{\mathbf{w}}^{\text{eff}} \tilde{\mathbf{w}}^{\text{eff}}} &= \alpha^2 \mathbb{E}\{\Phi \tilde{\mathbf{w}} \tilde{\mathbf{w}}^H \Phi^H\} \\ &= \frac{\alpha^2 \sigma^2}{2} \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_N \\ \mathbf{0}_N & (1 + \dfrac{g_{\text{R,max}}}{3}) \mathbf{I}_N \end{bmatrix}. \end{aligned} \tag{35}$$

Also, (17) can be rewritten as

$$\Sigma_{\tilde{\mathbf{e}}\tilde{\mathbf{e}}} = \frac{\alpha(1-\alpha)}{2} \left(\begin{bmatrix} \mathbf{I}_N & \mathbf{0}_N \\ \mathbf{0}_N & (1 + \dfrac{g_{\text{R,max}}}{3}) \mathbf{I}_N \end{bmatrix} + \sigma^2 \mathbf{I}_{2N} \right). \tag{36}$$

We notice that, when treating the quantization errors as additive white noise [48], (34) is consistent with the generalized bilinear problem described in [45].

Based on the PDF $f_{y_{i,j}}(y_{i,j})$ of each block in $\mathbf{H}^{\text{eff}}$, some procedures in Algorithm 1 should be modified to make it work as we expect. Hence, the input is replaced by $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{X}}_{\text{p}}$. The output is denoted by $\hat{\mathbf{H}}^{\text{eff}}$ and $\hat{\tilde{\mathbf{X}}}_{\text{d}}$. Meanwhile, we redefine $\sum_n \triangleq \sum_{n=1}^{2N}$. In (22), we calculate $\gamma = \sigma_{\text{eff}}^2 + \sigma_q^2$, in which σ_{eff}^2 and σ_q^2 are obtained from (35) and (36), respectively. For the manipulations at lines 29-30 of Algorithm 1, the estimated values and variances of the elements in each block should be calculated using their associated PDF separately. In other words, when calculating the mean and variance according

to (25), different σ_h^2 should be used for each block in $\mathbf{H}^{\text{eff}}$. For instance, when the estimated values and variances of the elements in $\mathbf{H}_{21}^{\text{eff}}$ are calculated, the PDF in (29) is expected to be used. Suppose that IQI value is practical small, we can use (31) for further simplicity and then adopt $\sigma_h^2 = (1-g_{\max,\text{R}}^2)^2/2$ in (25). Other steps in Algorithm 1 can be kept invariant. In summary, this modified algorithm is denoted by the proposed ECEC which only need to estimate the effective channel.

V. CHANNEL ESTIMATION AND DATA DETECTION UNDER BOTH RX AND TX IQI INVOLVING QUANTIZATION ERRORS

In this section, we consider a general case that signal is corrupted by both RX IQI and TX IQI. In this case, the combined signal cannot remove the effects of IQI. Alternatively, the ECEC method is expected to used.

A. PROPOSED EFFECTIVE CHANNEL ESTIMATION AND COMPENSATION WITH IQI AT BOTH THE RECEIVER AND THE TRANSMITTER

The proposed ECEC can be extended to this general case. First, the PDF of the elements in the effective channel needs to be derived. Based on this PDF, the modified Bi-GAMP algorithm can be utilized to complete effective channel estimation and data detection. Using the property that $\mathbf{G}_1 + \mathbf{G}_2 = \mathbf{I}_K$, Ξ in (6c) is rewritten as

$$\Xi = \begin{bmatrix} \Re(\mathbf{G}_1+\mathbf{G}_2) & \Im(\mathbf{G}_2-\mathbf{G}_1) \\ \Im(\mathbf{G}_1+\mathbf{G}_2) & \Re(\mathbf{G}_1-\mathbf{G}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_K & \mathbf{C} \\ \mathbf{0}_K & \mathbf{D} \end{bmatrix}, \quad (37)$$

where $\mathbf{C} = \text{diag}\{c_1, c_2, \cdots, c_K\}$, $\mathbf{D} = \text{diag}\{d_1, d_2, \cdots, d_K\}$ with $c_j = -g_{j,\text{T}}\sin(\phi_{j,\text{T}})$ and $d_j = g_{j,\text{T}}\cos(\phi_{j,\text{T}})$ for $j = 1, 2, \cdots, K$. By using the similar way in Section IV-C, the effective channel is defined as

$$\mathbf{H}^{\text{eff}} = \Phi\tilde{\mathbf{H}}\Xi \triangleq \begin{bmatrix} \mathbf{H}_{11}^{\text{eff}} & \mathbf{H}_{12}^{\text{eff}} \\ \mathbf{H}_{21}^{\text{eff}} & \mathbf{H}_{22}^{\text{eff}} \end{bmatrix}, \quad (38)$$

where $\mathbf{H}_{11}^{\text{eff}} = \Re(\mathbf{H})$, $\mathbf{H}_{12}^{\text{eff}} = \Re(\mathbf{H})\mathbf{C} - \Im(\mathbf{H})\mathbf{D}$, $\mathbf{H}_{21}^{\text{eff}} = \mathbf{A}\Re(\mathbf{H}) + \mathbf{B}\Im(\mathbf{H})$ and $\mathbf{H}_{22}^{\text{eff}} = (\mathbf{A}\Re(\mathbf{H}) + \mathbf{B}\Im(\mathbf{H}))\mathbf{C} + (\mathbf{B}\Re(\mathbf{H}) - \mathbf{A}\Im(\mathbf{H}))\mathbf{D}$. From $\mathbf{H}^{\text{eff}}$, we find that the components in $\mathbf{H}_{11}^{\text{eff}}$ independently follow the real Gaussian distribution $\mathcal{N}(0, 0.5)$ according to the fact that the elements in $\mathbf{H}$ are independently distributed in terms of the complex Gaussian distribution $\mathcal{CN}(0, 1)$. The PDF of elements in $\mathbf{H}_{21}^{\text{eff}}$ and $\mathbf{H}_{12}^{\text{eff}}$ can be obtained by the processes given in Section IV-C.

We subsequently need to determine the PDF of the elements in $\mathbf{H}_{22}^{\text{eff}}$. Without loss of generality, an arbitrary entry $y_{i,j}$ in $\mathbf{H}_{22}^{\text{eff}}$ is discussed because the same derivation can be applied to other elements. From (38), $y_{i,j}$ can be expressed as

$$y_{i,j} = (a_i c_j + b_i d_j)h_{i,j}^{\text{R}} + (b_i c_j - a_i d_j)h_{i,j}^{\text{I}}. \quad (39)$$

Lemma 3: The PDF of $y_{i,j}$ is independent of the phase mismatches $\phi_{i,\text{R}}$ and $\phi_{i,\text{T}}$ and is determined by the distribution of $g_{i,\text{R}}$, $g_{i,\text{T}}$, $h_{i,j}^{\text{R}}$ and $h_{i,j}^{\text{I}}$. Furthermore, at the bottom of this page, the analytic PDF of $y_{i,j}$ is given by (40) where $B_1 \triangleq g_{\text{L,R}}g_{\text{L,T}}$, $B_2 \triangleq g_{\text{L,R}}g_{\text{U,T}}$, $B_3 \triangleq g_{\text{U,R}}g_{\text{L,T}}$ and $B_4 \triangleq g_{\text{U,R}}g_{\text{U,T}}$ with $g_{\text{L,R}} \triangleq (1-g_{\max,\text{R}})$, $g_{\text{U,R}} \triangleq (1+g_{\max,\text{R}})$, $g_{\text{L,T}} \triangleq (1-g_{\max,\text{T}})$ and $g_{\text{U,T}} \triangleq (1+g_{\max,\text{T}})$.

Proof: Please refer to Appendix VII. ■

Due to the independence between $g_{i,R}$, $g_{j,T}$ and $h_{i,j}^R$, we have the mean of $y_{i,j}$ as $\mathbb{E}[y_{i,j}] = 0$ and the variance of $y_{i,j}$ according to

$$\begin{aligned}\mathbb{V}[y_{i,j}] &= \mathbb{E}[y_{i,j}^2] = \mathbb{E}[g_{i,\text{R}}^2 g_{j,\text{T}}^2 (h_{i,j}^{\text{R}})^2] \\ &= \frac{1}{2}\left(1 + \frac{g_{\max,\text{R}}^2}{3}\right)\left(1 + \frac{g_{\max,\text{T}}^2}{3}\right). \end{aligned} \quad (41)$$

Considering the effective channel $\mathbf{H}^{\text{eff}}$, (5) yields

$$\tilde{\mathbf{R}} = \alpha\mathbf{H}^{\text{eff}}\tilde{\mathbf{X}} + \alpha\Phi\tilde{\mathbf{W}} + \tilde{\mathbf{E}}, \quad (42)$$

which has the same form as (34). Therefore, the variances of the effective noise and the quantization noise are also given by (35) and (36), respectively. Again, we notice that (42) is consistent with the generalized bilinear problem described in [45] when treating the quantization errors as additive white noise [48].

Based on the PDF $f_{y_{i,j}}(y_{i,j})$ of each block in $\mathbf{H}^{\text{eff}}$, some procedures in Algorithm 1 should be modified to make it work as we anticipate. Hence, the input is replaced by $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{X}}_{\text{p}}$. The output is denoted by $\hat{\mathbf{H}}^{\text{eff}}$ and $\hat{\tilde{\mathbf{X}}}_{\text{d}}$. Meanwhile, we redefine $\sum_n \triangleq \sum_{n=1}^{2N}$. In (22), we calculate $\gamma = \sigma_{\text{eff}}^2 + \sigma_q^2$, in which

$$f_{y_{i,j}}(y_{i,j}) = \begin{cases} \dfrac{1}{4\sqrt{\pi}g_{\max,\text{R}}} \times \Bigg[\int_{y_{i,j}/B_2}^{y_{i,j}/B_1} \dfrac{1}{t}\ln\dfrac{y_{i,j}}{t}\exp(-t^2)\text{d}t - \dfrac{1}{2}\ln(g_{\text{L,R}}g_{\text{L,T}})\left(\Gamma\left(0, \dfrac{y_{i,j}^2}{B_2^2}\right) - \Gamma\left(0, \dfrac{y_{i,j}^2}{B_1^2}\right)\right) \\ \quad + \dfrac{1}{2}\ln\left(\dfrac{g_{\text{U,T}}}{g_{\text{L,T}}}\right)\left(\Gamma\left(0, \dfrac{y_{i,j}^2}{B_3^2}\right) - \Gamma\left(0, \dfrac{y_{i,j}^2}{B_2^2}\right)\right) \\ \quad + \dfrac{1}{2}\ln(g_{\text{U,R}}g_{\text{U,T}})\left(\Gamma\left(0, \dfrac{y_{i,j}^2}{B_4^2}\right) - \Gamma\left(0, \dfrac{y_{i,j}^2}{B_3^2}\right)\right) - \int_{y_{i,j}/B_3}^{y_{i,j}/B_4} \dfrac{1}{t}\ln\dfrac{y_{i,j}}{t}\exp(-t^2)\text{d}t\Bigg], & \text{if } y_{i,j} \neq 0 \\ \dfrac{1}{4\sqrt{\pi}g_{\max,\text{R}}} \times \Bigg[\dfrac{1}{2}\left((\ln B_2)^2 - (\ln B_1)^2\right) - \ln\left(\dfrac{B_2}{B_1}\right)\ln(g_{\text{L,R}}g_{\text{L,T}}) + \ln\left(\dfrac{B_3}{B_2}\right)\ln\left(\dfrac{g_{\text{U,T}}}{g_{\text{L,T}}}\right) \\ \quad - \dfrac{1}{2}\left((\ln B_4)^2 - (\ln B_3)^2\right) + \ln\left(\dfrac{B_4}{B_3}\right)\ln(g_{\text{U,R}}g_{\text{U,T}})\Bigg], & \text{otherwise.} \end{cases} \quad (40)$$

σ^2_{eff} and σ^2_q are obtained from (35) and (36), respectively. For the manipulations at lines 29-30 of Algorithm 1, the estimated values and variances of the elements in each block should be calculated using their associated PDF separately. In other words, when calculating the mean and variance according to (25), different σ^2_h should be used for each block in $\mathbf{H}^{\text{eff}}$. For instance, when the estimated values and variances of the elements in $\mathbf{H}^{\text{eff}}_{22}$ are calculated, the PDF in (40) is expected to be used. In summary, this modified algorithm is denoted by the proposed ECEC which only need to estimate the effective channel and perform data detection.

B. COMPUTATIONAL COMPLEXITY FOR MODIFIED BI-GAMP ALGORITHM

In order to make the complexity analysis convenient for each iteration, we assess the complexity in terms of floating-point operations (FLOPs) used in [51]. It is assumed that the operation of function exp(·) in (23) can be implemented by a look-up table. In Algorithm 1, we see that all steps are based on real-valued operations. In addition, it is easy to find that the multiplication and the addition of two real numbers need 1 FLOPs, respectively. We denote by $L' = 2K + L$ the length of one block and let $K' = 2K$, $N' = 2N$. Based on these considerations, the complexity of operations in Algorithm 1 can be calculated as follows.

The operations in lines 3-8 require $\left(7K' + 12\right) N'L'$ FLOPs. When (22) is adopted, $13N'L'$ FLOPs are needed for lines 9-14. The operations in lines 15-16 demand $\left(3N'K' + 3K'L'\right)$ FLOPs. For operations in lines 17-19, $\left(4N' + 9\right) K'L'$ FLOPs are required. For obtaining the expectation and variance, the operations in lines 21-22 need $4\sqrt{Q}K'L'$ FLOPs. And the operations in lines 26-28 need $\left(4L' + 7\right) N'K'$ FLOPs. When (25) is used to get the expectation and variance, the complexity is $6N'K'$ FLOPs for lines 29-30. Calculating the terminal condition in line 31 requires $\left(3N'L' + 1\right)$ FLOPs.

VI. SIMULATION RESULTS AND DISCUSSION

Monte-Carlo simulations are used to evaluate the performance of the proposed methods. We also validate our previous theoretical analysis by comparing it with the numerical results.

A. SIMULATION PARAMETERS

For the setup, a single-cell system is considered. At the center of this cell, one BS employing $N = 64$ antennas serves $K = 16$ users, which uniformly distribute around the BS on a circle with radius of one. Moreover, the BS uses an uniform linear array (ULA). All antennas at the BS are assumed to use identical ADCs with equal number of resolution bits. In all simulations, We assume that all users employ the same transmission power $p_k = 1$ for $\forall k \in \{k\}_{k=1}^K$. Given an SNR, the noise power can be calculated as $\sigma^2 = (\sum_{k=1}^K p_k)/(10^{0.1\text{SNR}})$. We also make an assumption that the data sequences transmitted by all users are randomly selected from the constellation of QPSK when evaluating BER performance. In the simulations, the length of one transmission block is set to be $L = 132$, at the head of which 32 samples are used for training. In addition, all the simulation results associated to BER are averaged over 5000 realizations of the channel and IQI.

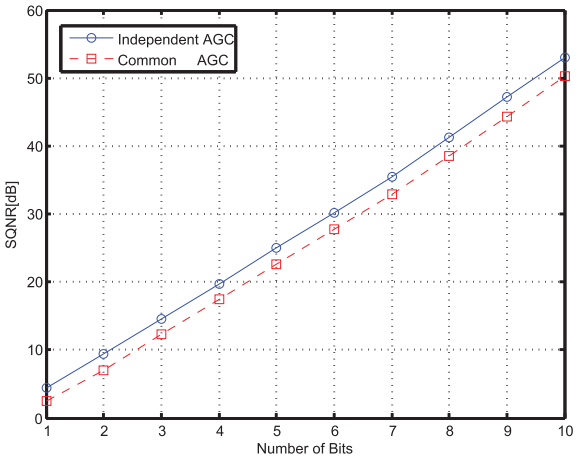


FIGURE 3. Signal-to-quantization-noise ratio (SQNR) vs. the number of bits for proposed independent AGC and common AGC. In this figure, we set that the amplitude ratio between the Q branch and I branch is 0.2.

B. VERIFICATION OF DERIVATIONS

In Fig. 3, the signal-to-quantization-noise ratio (SQNR) for proposed independent AGC and common AGC are depicted. the SQNR in this figure is defined as [6]

$$\text{SQNR} = 10\log_{10}\frac{||\bar{\mathbf{y}}||_{\text{F}}}{||\mathbf{r} - \bar{\mathbf{y}}||_{\text{F}}},$$

where $\mathbf{r}$ is the quantized version of $\bar{\mathbf{y}}$ and $||\cdot||_{\text{F}}$ denotes Frobenius norm. To manifest the superiority of proposed independent AGC, we assume that the averaging amplitude ratio between the Q branch and I branch is 0.2 and we do not consider the phase mismatch. From the results in this figure, it can be seen that our proposed independent AGC scheme outperforms the common AGC scheme, which controls the signals amplitude of Q branch and I branch using the same scaling factor. Regarding various widths of quantization bits, it can be also seen that our proposed independent AGC is preferable. Moreover, it is also obvious that one more quantization bit can approximately bring 5.8dB-6dB gain, which is consistent with the fact about 6.02dB can be obtained when one more bit is added for the quantizer using uniform quantization step [6].

In Fig. 4, we show the analytic PDF of elements in the effective channel $\mathbf{H}^{\text{eff}}$ and the corresponding simulation results. In this figure, we only consider RX IQI and set $g_{\text{max,R}} = 0.8$, $\phi_{\text{max,R}} = 10°$. Moreover, "Sim" denotes the simulation results and "Ana." denotes the analytic results. As noted in Section IV-C, the elements in $\mathbf{H}^{\text{eff}}_{11}$ and $\mathbf{H}^{\text{eff}}_{12}$ have the Gaussian distribution $\mathcal{N}(0, 0.5)$, which are omitted in this figure. For the elements in $\mathbf{H}^{\text{eff}}_{21}$ and $\mathbf{H}^{\text{eff}}_{22}$, their analytic PDFs

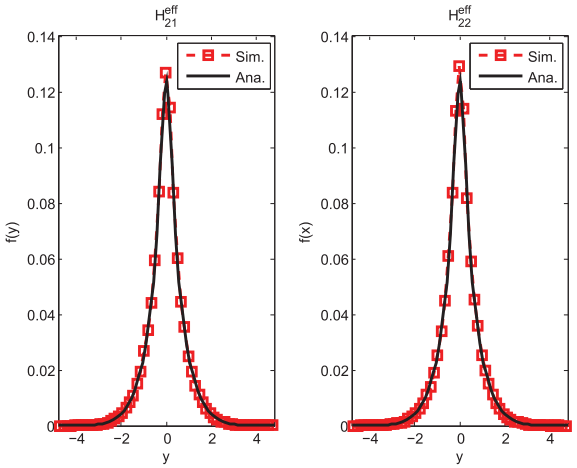


FIGURE 4. PDF of $\mathbf{H}_{21}^{\text{eff}}$ and $\mathbf{H}_{22}^{\text{eff}}$. Only RX IQI is considered. The related parameters are set by $g_{\max,\mathrm{R}} = 0.8$ and $\phi_{\max,\mathrm{R}} = 10°$. "Sim." denotes the simulation results and "Ana." denotes the analytic results.

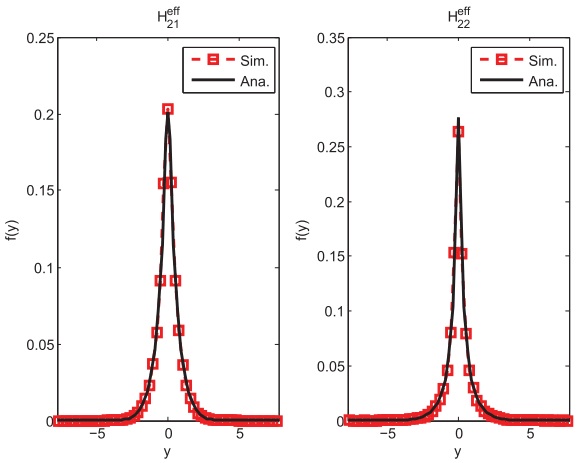


FIGURE 5. PDF of $\mathbf{H}_{21}^{\text{eff}}$ and $\mathbf{H}_{22}^{\text{eff}}$. Both RX IQI and TX IQI are considered. The related parameters are set by $g_{\max,\mathrm{R}} = 0.8$, $\phi_{\max,\mathrm{R}} = 10°$, $g_{\max,\mathrm{T}} = 0.8$ and $\phi_{\max,\mathrm{T}} = 10°$.

are determined by (29) which is consistent with the simulation results. In other words, this consistency demonstrates the validity of our theoretical analysis. It should be pointed out that the elements in $\mathbf{H}_{21}^{\text{eff}}$ and $\mathbf{H}_{22}^{\text{eff}}$ follow the same distribution, which can also be verified by the formulations of $f_{y_{i,j}}(y_{i,j})$ and $f_{z_{i,j}}(z_{i,j})$ in Section IV-C.

When both RX IQI and TR IQI are considered, in Fig. 5, we exhibit the analytic PDF of elements in the effective channel $\mathbf{H}^{\text{eff}}$ and the corresponding simulation results. Here, we set $g_{\max,\mathrm{R}} = 0.8$, $\phi_{\max,\mathrm{R}} = 10°$, $g_{\max,\mathrm{T}} = 0.8$ and $\phi_{\max,\mathrm{T}} = 10°$. As noted in Section V, the elements in $\mathbf{H}_{11}^{\text{eff}}$ have the Gaussian distribution $\mathcal{N}(0, 0.5)$. For the elements in $\mathbf{H}_{12}^{\text{eff}}$, their analytic PDF are determined by replacing $g_{\max,\mathrm{R}}$ with $g_{\max,\mathrm{T}} = 0.8$ in (29). Similarly, for the elements in $\mathbf{H}_{21}^{\text{eff}}$, their PDFs are determined by setting $g_{\max,\mathrm{R}} = 0.8$ in (29). We see that all analytic results are consistent with the simulation results. For the elements in $\mathbf{H}_{22}^{\text{eff}}$, their analytic PDF are determined by setting $g_{\max,\mathrm{R}} = 0.8$ and $g_{\max,\mathrm{T}} = 0.8$ in (40). The consistency between analytic results and simulation results demonstrates the correctness of our theoretical analysis.

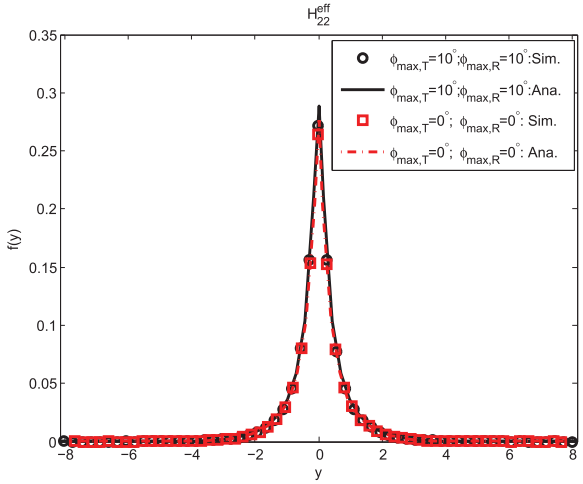


FIGURE 6. The PDF of the elements in $\mathbf{H}_{22}^{\text{eff}}$ for different phase mismatches. The parameters associated to the amplitude mismatches are set by $g_{\max,\mathrm{R}} = 0.8$ and $g_{\max,\mathrm{T}} = 0.8$.

The results in Fig. 6 certify that the PDF of the elements in the effective channel $\mathbf{H}^{\text{eff}}$ is independent of the phase mismatch. Here, we set $g_{\max,\mathrm{R}} = 0.8$ and $g_{\max,\mathrm{T}} = 0.8$. The results are shown for the case where systems are impacted by the phase mismatches modeled by $\phi_{\max,\mathrm{R}} = 10°$ and $\phi_{\max,\mathrm{T}} = 10°$. As a comparison, we also plot the results when the systems are not affected by the phase mismatches. It is obvious that the distribution of elements in the effective channel does not vary as the phase mismatches are changed at both the transmitter and the receiver, which is confirmed by both the theoretical results and simulation results.

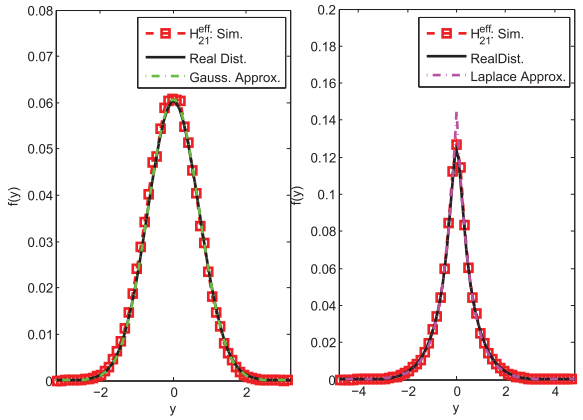


FIGURE 7. Gaussian distribution approximation when $g_{\max,\mathrm{R}} = 0.1$ and $\phi_{\max,\mathrm{R}} = 10°$ (left) and Laplace distribution approximation when $g_{\max,\mathrm{R}} = 0.8$ and $\phi_{\max,\mathrm{R}} = 10°$ (right). Only RX IQI is considered.

Here only RX IQI is considered. The results in the left subfigure of Fig. 7 certify that a Gaussian distribution can be used to approximate the practical distribution when the

IQI values are practically small. If we set $g_{\max,\mathrm{R}} = 0.1$ and $\phi_{\max,\mathrm{R}} = 10°$, it can be seen that the Gaussian distribution can fit the practical distribution well. Much smaller $g_{\max,\mathrm{R}}$ can give a more accurate approximation. The results in the right subfigure of Fig. 7 certify that a Laplace distribution can be used to approximate the practical distribution when the IQI values are relatively large. We set $g_{\max,\mathrm{R}} = 0.8$ and $\phi_{\max,\mathrm{R}} = 10°$ to make the amplitude more severe. The fitness between the practical PDF and the approximated one leads to the conclusion that we could use the Laplace distribution to approximate the practical distribution which has a complicated expression.

C. PERFORMANCE COMPARISON OF CCEC, ECEC AND OTHER STATE-OF-THE-ART METHODS

Hereafter, the performance of our proposed methods will be shown. We first give some explanations about some state-of-the-art techniques so as to compare them with the proposed methods. The *LMMSE without IQI comp.* represents that the LMMSE technique is directly utilized for channel estimation and data detection by using the combined signal. Taking quantization errors into account, we also show the IQI compensation method mentioned in [12], which first estimates the CSI using the combined training signal, and then estimates the IQI parameters before carrying out IQI compensation. The LMMSE IQU and LMMSE IQA are referred to [11], in which more details of these two methods can be found. It is worth noting that IQI parameters are known for the LMMSE IQU and the LMMSE IQA when calculating the auto-correlation matrix and the cross-correlation matrix. For our proposed CCEC and ECEC, we set $\beta = 0.2$, $T_{\max} = 90$ and $\epsilon = 10^{-5}$ (in [7] $\epsilon = 10^{-8}$ and in [49] $\epsilon = 10^{-3}$) in Algorithm 1 for the succeeding simulation results.

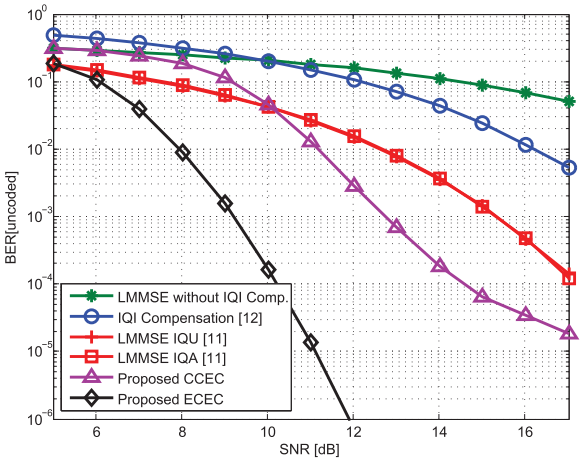


FIGURE 8. BER vs. SNR for different data detection methods using 4-bit quantization. Only RX IQI is considered. The parameters are set by $g_{\max,\mathrm{R}} = 0.2$ and $\phi_{\max,\mathrm{R}} = 10°$ for all solid lines.

In Fig. 8, the performance associated with BER is shown when 4-bit quantization is used for all ADCs. Using the combined signal, the proposed CCEC outperforms the method used in [12]. And at the high SNR regime, the proposed CCEC can obtain more gain. However, as mentioned previously, using the combined signal results in performance loss due to the fact that only the real part of received signal is exploited. In contrast, the proposed ECEC can fully utilize the received signal and is superior to the proposed CCEC, which is demonstrated in Fig. 8. Moreover, compared with the LMMSE IQU and the LMMSE IQA, the proposed ECEC provides an improvement of BER. Here, LMMSE IQU and LMMSE IQA almost have the same performance since the pilots in this study can avoid the self-interference during the training stage. From the simulation results, the proposed ECEC can obtain a better performance because it exploits that property that each block in the effective channel has different distribution when performing channel estimation and IQI compensation.

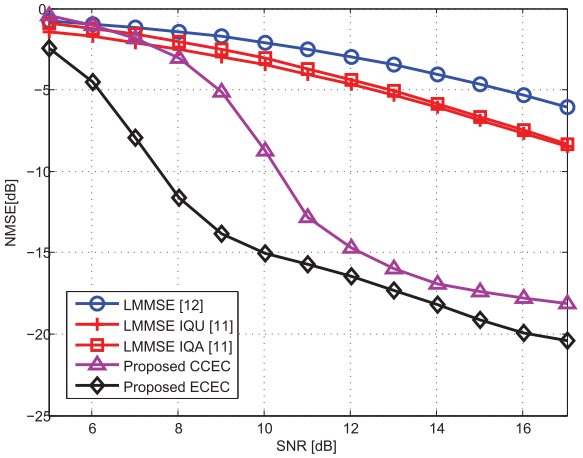


FIGURE 9. NMSE vs. SNR for different channel estimation methods using 4-bit quantization. Only RX IQI is considered. The parameters are set by $g_{\max,\mathrm{R}} = 0.2$ and $\phi_{\max,\mathrm{R}} = 10°$ for all solid lines.

From the viewpoint of the channel estimation, in Fig. 9, we show the normalized mean square error (NMSE) according to

$$\mathrm{NMSE}(\hat{\mathbf{H}}) = 10\log_{10}\frac{||\hat{\mathbf{H}} - \mathbf{H}||_{\mathrm{F}}}{||\mathbf{H}||_{\mathrm{F}}},$$

$$\mathrm{NMSE}(\hat{\mathbf{H}}_{\mathrm{eff}}) = 10\log_{10}\frac{||\hat{\mathbf{H}}_{\mathrm{eff}} - \mathbf{H}_{\mathrm{eff}}||_{\mathrm{F}}}{||\mathbf{H}_{\mathrm{eff}}||_{\mathrm{F}}}.$$

where $\hat{\mathbf{H}}$ (or $\hat{\mathbf{H}}_{\mathrm{eff}}$) denotes the estimated version of $\mathbf{H}$ (or $\mathbf{H}_{\mathrm{eff}}$). We observe from the results in Fig. 9 that the proposed CCEC can obtain a preferable NMSE than the LMMSE given in [12] when the combined signal is used during the training stage. If the effective channel needs to be estimated, the proposed ECEC also outperforms the LMMSE IQU and the LMMSE IQA proposed in [11]. Additionally, it can be seen that a better NMSE is obtained if the effective channel is estimated from the received pilots instead of estimating the CSI from the combined pilots, i.e., the ECEC outperforms the CCEC in channel estimation.

Under RX IQI and the different quantization bits, the BER results for different methods are depicted in Fig. 10.

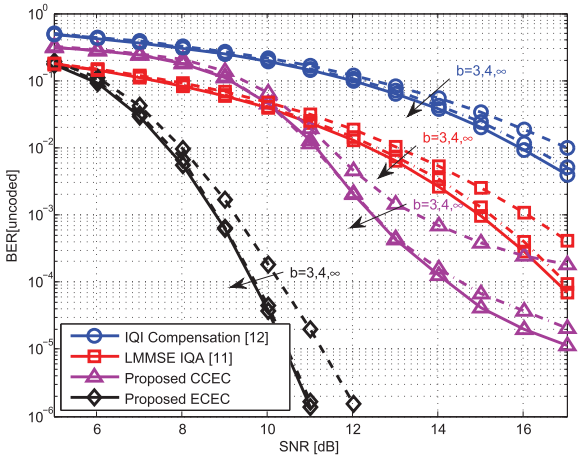


FIGURE 10. BER vs. SNR for different data detection methods. Only RX IQI is considered. The parameters are set by $g_{\max,R} = 0.1$ and $\phi_{\max,R} = 10°$. Results for different quantization bits are depicted.

From these results associated with the proposed ECEC, the performance of using 4-bit are quite close to that of using full-bit, particularly at the low SNR regime where noise is predominant. The same conclusion can be obtained for the CCEC. Supposing that the resources are limited, employing 4-bit ADCs at each antenna is an acceptable trade-off. For the case where 3-bit ADCs are employed, we see that a large performance gap appears at the high SNR regime where quantization errors are predominant over the noise. If the quantized signal is used by the receiver, the proposed ECEC can outperform the LMMSE IQA when the effective channel needs to be estimated. Likewise, If only the combined signal is used to estimate the CSI, the proposed CCEC is superior to the IQ compensation method used in [12].

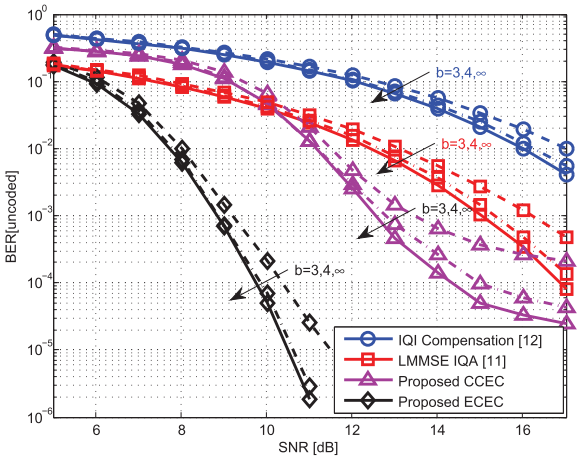


FIGURE 11. BER vs. SNR for different channel estimation methods. Both RX IQI and TX IQI are considered. The parameters are set by $g_{\max,R} = 0.1$, $\phi_{\max,R} = 10°$, $g_{\max,T} = 0.1$ and $\phi_{\max,T} = 10°$. Results for different quantization bits are depicted.

Suppose that both the RX IQI and the TX IQI are involved in the quantized MIMO systems, we show the BER results in Fig. 11. The parameters are set by $g_{\max,R} = 0.1$, $\phi_{\max,R} = 10°$, $g_{\max,T} = 0.1$ and $\phi_{\max,T} = 10°$. In this case, it can be concluded that our proposed ECEC can also outperform the LMMSE IQA when the effective channel needs to be estimated. Likewise, if only the combined signals are used to estimate the CSI, our proposed CCEC is superior to the IQ compensation method used in [12]. From the results in Fig. 11 and Fig. 10, we see that TX IQI can degrades the performance using the CCEC or the ECEC. On the other hand, the LMMSE IQA can obtain the same performance when TX IQI is considered because the IQI parameters are perfectly known at the receiver when calculating the auto-correlation matrix and the cross-correlation matrix in [11].

VII. CONCLUSION

In this paper, we have studied the CCEC and the ECEC for channel estimation and IQI compensation in the uplink quantized massive MIMO systems. In addition, an independent automatic gain control (AGC) scheme for I branch and Q branch has been proposed to control the dynamic range for both branches. Taking the impacts of both quantization and IQI into account, we have given the details of the channel estimation and the IQI compensation for both the CSEC and the ECEC using the Bi-GAMP algorithm, respectively. For the proposed ECEC, we have theoretically derived the PDF of the elements in the effective channel. We use the numerical results to testify the validity of our analytic results and the fact that the analytic PDF can be approximated by a Gaussian distribution when the IQI parameters are relatively small. Compared with other techniques by simulation results, the proposed methods can obtain better performance.

In the future, we will extend our work to the scenario with multiple cells. Considering the impacts of low-resolution ADCs and IQI, we will analyze the cell coverage under this scenario, particularly when the mmwave cellular networks are involved. In addition, the wideband systems with frequency dependent IQI are worth investigating for future work as well.

APPENDIX A
PROOF OF LEMMA 1

First, we can rewrite $y_{i,j} = g_{i,\mathrm{R}} \sin(\phi_{i,\mathrm{R}}) h^{\mathrm{R}}_{i,j} + g_{i,\mathrm{R}} \cos(\phi_{i,\mathrm{R}}) h^{\mathrm{I}}_{i,j}$. Based on the stated definitions and the predefined PDF of $g_{i,\mathrm{R}}$, $\phi_{i,\mathrm{R}}$ and $h_{i,j}$, the PDF of $y_{i,j}$ is then calculated as

$$\begin{aligned} f_{y_{i,j}}(y_{i,j}) &= \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} f(y_{i,j}|g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}) \\ &\quad \times f(g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}) \mathrm{d}\, g_{i,\mathrm{R}} \mathrm{d}\, \phi_{i,\mathrm{R}} \\ &\overset{(a)}{=} \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \frac{1}{\sqrt{\pi(a_i^2 + b_i^2)}} \exp\left(-\frac{y_{i,j}^2}{a_i^2 + b_i^2}\right) \\ &\quad \times \frac{1}{4 g_{\max,\mathrm{R}}} \frac{1}{\phi_{\max,\mathrm{R}}} \mathrm{d}\, g_{i,\mathrm{R}} \mathrm{d}\, \phi_{i,\mathrm{R}} \\ &\overset{(b)}{=} \frac{1}{2 g_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \frac{1}{\sqrt{\pi g_{i,\mathrm{R}}^2}} \exp\left(-\frac{y_{i,j}^2}{g_{i,\mathrm{R}}^2}\right) \mathrm{d}\, g_{i,\mathrm{R}}, \end{aligned} \tag{43}$$

in which (a) is based on the fact that $y_{i,j}$ is a Gaussian random variable with variance $a_i^2 + b_i^2$ [52] given $g_{i,\mathrm{R}}$ and $\phi_{i,\mathrm{R}}$, which are two independent random variables. Suppose that the amplitude mismatch is not too large, we assume $1 - g_{\max,\mathrm{R}} > 0$. If we let $t = y_{i,j}/g_{i,\mathrm{R}}$ ($y_{i,j} \neq 0$), then we can get $f_{y_{i,j}}(y_{i,j})$ according to

$$\begin{aligned} f_{y_{i,j}}(y_{i,j}) &= \frac{1}{2\sqrt{\pi}g_{\max,\mathrm{R}}} \int_{y_{i,j}/g_{\mathrm{U,R}}}^{y_{i,j}/g_{\mathrm{L,R}}} \frac{1}{t} \exp(-t^2) \mathrm{d}\, t \\ &\overset{(a)}{=} \frac{1}{2\sqrt{\pi}g_{\max,\mathrm{R}}} \left(\Gamma\left(0, \frac{y_{i,j}^2}{g_{\mathrm{U,R}}^2}\right) - \Gamma\left(0, \frac{y_{i,j}^2}{g_{\mathrm{L,R}}^2}\right) \right), \end{aligned} \tag{44}$$

in which we resort to the Gamma function defined by (8.350.2) in [53]. When $y_{i,j} = 0$, we can easily obtain the result given in (29). In the same way, we can calculate the PDF of $y_{i,j}$ conditioned by $\phi_{i,\mathrm{R}}$ according to

$$\begin{aligned} & f_{y_{i,j}|\phi_{i,\mathrm{R}}}(y_{i,j}|\phi_{i,\mathrm{R}}) \\ &= \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} f(y_{i,j}|g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}) f(g_{i,\mathrm{R}}) \mathrm{d}\, g_{i,\mathrm{R}} \\ &\overset{(a)}{=} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \frac{1}{\sqrt{\pi(a_i^2 + b_i^2)}} \exp\left(-\frac{y_{i,j}^2}{a_i^2 + b_i^2}\right) \frac{1}{2g_{\max,\mathrm{R}}} \mathrm{d}\, g_{i,\mathrm{R}} \\ &\overset{(b)}{=} \frac{1}{2g_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \frac{1}{\sqrt{\pi g_{i,\mathrm{R}}^2}} \exp\left(-\frac{y_{i,j}^2}{g_{i,\mathrm{R}}^2}\right) \mathrm{d}\, g_{i,\mathrm{R}} \\ &= f_{y_{i,j}}(y_{i,j}). \end{aligned} \tag{45}$$

It is obvious from (45) that $f_{y_{i,j}}(y_{i,j})$ is independent of the phase mismatch $\phi_{i,\mathrm{R}}$.

APPENDIX B
PROOF OF LEMMA 2

We first prove that all elements in $\mathbf{H}_{21}^{\mathrm{eff}}$ are mutually independent and follow the same distribution. For $i \neq i'$ or $j \neq j'$, we calculate the joint PDF of $y_{i,j}$ and $y_{i',j'}$ as

$$\begin{aligned} & f_{y_{i,j}y_{i',j'}}(y_{i,j}, y_{i',j'}) \\ &= \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \\ &\quad \times f(y_{i,j}, y_{i',j'}|g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}, g_{i',\mathrm{R}}, \phi_{i',\mathrm{R}}) \\ &\quad \times f(g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}, g_{i',\mathrm{R}}, \phi_{i',\mathrm{R}}) \mathrm{d}\, g_{i,\mathrm{R}} \mathrm{d}\, \phi_{i,\mathrm{R}} \mathrm{d}\, g_{i',\mathrm{R}} \mathrm{d}\, \phi_{i',\mathrm{R}} \\ &\overset{(a)}{=} \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} f(y_{i,j}|g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}) \\ &\quad \times f(y_{i',j'}|g_{i',\mathrm{R}}, \phi_{i',\mathrm{R}}) f(g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}) f(g_{i',\mathrm{R}}, \phi_{i',\mathrm{R}}) \\ &\quad \times \mathrm{d}\, g_{i,\mathrm{R}} \mathrm{d}\, \phi_{i,\mathrm{R}} \mathrm{d}\, g_{i',\mathrm{R}} \mathrm{d}\, \phi_{i',\mathrm{R}} \\ &\overset{(b)}{=} f_{y_{i,j}}(y_{i,j}) f_{y_{i',j'}}(y_{i',j'}), \end{aligned} \tag{46}$$

in which (a) is based on the fact that $y_{i,j}$ and $y_{i',j'}$ are mutually independent given $g_{i,\mathrm{R}}$, $\phi_{i,\mathrm{R}}$, $g_{i',\mathrm{R}}$ and $\phi_{i',\mathrm{R}}$. From (46), we see that $y_{i,j}$ and $y_{i',j'}$ are mutually independent. Additionally, it is easy to find from (29) that $y_{i,j}$ and $y_{i',j'}$ have the same distribution.

We then prove that the arbitrary element in $\mathbf{H}_{21}^{\mathrm{eff}}$ is independent of the arbitrary element in $\mathbf{H}_{22}^{\mathrm{eff}}$. To do so, we only need to verify the independence between $y_{i,j}$ and $z_{i,j}$ because it is obvious that $y_{i,j}$ and $z_{i',j'}$ are independent for $i \neq i'$ or $j \neq j'$. Recalling that $y_{i,j} = a_i h_{i,j}^{\mathrm{R}} + b_i h_{i,j}^{\mathrm{I}}$ and $z_{i,j} = b_i h_{i,j}^{\mathrm{R}} - a_i h_{i,j}^{\mathrm{I}}$, we can rewrite them in a vector form as

$$\begin{bmatrix} y_{i,j} \\ z_{i,j} \end{bmatrix} = \begin{bmatrix} [r]a_i & b_i \\ b_i & -a_i \end{bmatrix} \begin{bmatrix} h_{i,j}^{\mathrm{R}} \\ h_{i,j}^{\mathrm{I}} \end{bmatrix} = \mathbf{T} \begin{bmatrix} h_{i,j}^{\mathrm{R}} \\ h_{i,j}^{\mathrm{I}} \end{bmatrix}. \tag{47}$$

Due to $\mathbf{T}\mathbf{T}^H = g_{i,\mathrm{R}}\mathbf{I}_2$ and the independence between $h_{i,j}^{\mathrm{R}}$ and $h_{i,j}^{\mathrm{I}}$, we can conclude that the independence between $y_{i,j}$ and $z_{i,j}$ is also satisfied.

APPENDIX C
PROOF OF LEMMA 3

Based on the independence between $g_{i,\mathrm{R}}$, $\phi_{i,\mathrm{R}}$, $g_{j,\mathrm{T}}$ and $\phi_{j,\mathrm{T}}$ along with their PDF, we have

$$f(g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}, g_{j,\mathrm{T}}, \phi_{j,\mathrm{T}}) = \frac{1}{2g_{\max,\mathrm{T}}} \frac{1}{2g_{\max,\mathrm{R}}} \frac{1}{2\phi_{\max,\mathrm{T}}} \frac{1}{2\phi_{\max,\mathrm{R}}}.$$

Using definitions $g_{\mathrm{R,T}} \triangleq g_{\max,\mathrm{R}} g_{\max,\mathrm{T}}$ and $\phi_{\mathrm{R,T}} \triangleq \phi_{\max,\mathrm{R}}\phi_{\max,\mathrm{T}}$, we can calculate the PDF of $y_{i,j}$ according to (48) at the top of the next page, (a) is obtained because $y_{i,j}$ is a Gaussian random variable with variance $(a_i c_j + b_i d_j)^2 + (b_i c_j - a_i d_j)^2 = g_{i,\mathrm{R}}^2 g_{i,\mathrm{T}}^2$ given $g_{i,\mathrm{R}}$ and $\phi_{i,\mathrm{R}}$, which are two independent variables. Using the same derivation as (45), we can also obtain that $f(y_{i,j})$ is independent of the phase mismatches $\phi_{i,\mathrm{R}}$ and $\phi_{j,\mathrm{T}}$. Based on this fact, when calculating $f_{y_{i,j}}(y_{i,j})$, we set $\phi_{i,R} = 0$ and $\phi_{j,T} = 0$, following which we can get

$$y_{i,j} = g_{i,\mathrm{R}} g_{j,\mathrm{T}} h_{i,j}^{\mathrm{R}}. \tag{49}$$

Defining $z_{i,j} = g_{i,\mathrm{R}} g_{j,\mathrm{T}}$ for simplicity, we first obtain the PDF of $z_{i,j}$ and then calculate the PDF of $y_{i,j}$. Since we have already known the PDF of $g_{i,\mathrm{R}}$ and $g_{j,\mathrm{T}}$, which are denoted by $f_{g_{i,\mathrm{R}}}(g_{i,\mathrm{R}})$ and $f_{g_{j,\mathrm{T}}}(g_{j,\mathrm{T}})$, respectively, we can achieve [54]

$$f_{z_{i,j}}(z_{i,j}) = \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \frac{1}{|g_{i,\mathrm{R}}|} f_{g_{i,\mathrm{R}}}(g_{i,\mathrm{R}}) f_{g_{j,\mathrm{T}}}\left(\frac{z_{i,j}}{g_{i,\mathrm{R}}}\right) \mathrm{d}\, g_{i,\mathrm{R}}. \tag{50}$$

Here, we assume that $g_{\mathrm{L,R}} > 0$ and $g_{\max,\mathrm{T}} < g_{\max,\mathrm{R}}$ (when $g_{\max,\mathrm{T}} > g_{\max,\mathrm{R}}$, we can obtain the similar results using the succeeding procedures). Based on these assumptions and the fact that uniform distributions are used to model $g_{i,\mathrm{R}}$ and $g_{j,\mathrm{T}}$, we can obtain the PDF of $z_{i,j}$ according to

$$f_{z_{i,j}}(z_{i,j}) = \begin{cases} \dfrac{1}{g_{\mathrm{R,T}}} \ln\left(\dfrac{z_{i,j}}{g_{\mathrm{L,R}} g_{\mathrm{L,T}}}\right), & \text{if } B_1 \leqslant z_{i,j} < B_2 \\ \dfrac{1}{g_{\mathrm{R,T}}} \ln\left(\dfrac{g_{\mathrm{U,T}}}{g_{\mathrm{L,T}}}\right), & \text{if } B_2 \leqslant z_{i,j} \leqslant B_3 \\ \dfrac{1}{g_{\mathrm{R,T}}} \ln\left(\dfrac{g_{\mathrm{U,R}} g_{\mathrm{U,T}}}{z_{i,j}}\right), & \text{if } B_3 < z_{i,j} \leqslant B_4 \\ 0, & \text{otherwise,} \end{cases} \tag{51}$$

$$f_{y_{i,j}}(y_{i,j}) = \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \int_{-\phi_{\max,T}}^{\phi_{\max,\mathrm{T}}} \int_{g_{\mathrm{L,T}}}^{g_{\mathrm{U,T}}} f(y_{i,j}|g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}, g_{j,\mathrm{T}}, \phi_{j,\mathrm{T}}) f(g_{i,\mathrm{R}}, \phi_{i,\mathrm{R}}, g_{j,\mathrm{T}}, \phi_{j,\mathrm{T}}) \mathrm{d}\, g_{i,\mathrm{R}} \mathrm{d}\, \phi_{i,\mathrm{R}} \mathrm{d}\, g_{j,\mathrm{T}} \mathrm{d}\, \phi_{j,\mathrm{T}}$$
$$\overset{(a)}{=} \int_{-\phi_{\max,\mathrm{R}}}^{\phi_{\max,\mathrm{R}}} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \int_{-\phi_{\max,\mathrm{T}}}^{\phi_{\max,\mathrm{T}}} \int_{g_{\mathrm{L,T}}}^{g_{\mathrm{U,T}}} \frac{1}{\sqrt{\pi g_{i,\mathrm{R}}^2 g_{j,\mathrm{T}}^2}} \exp\left(-\frac{y_{i,j}^2}{g_{i,\mathrm{R}}^2 g_{j,\mathrm{T}}^2}\right) \frac{1}{16 g_{\mathrm{R,T}} \phi_{\mathrm{R,T}}} \mathrm{d}\, g_{i,\mathrm{R}} \mathrm{d}\, \phi_{i,\mathrm{R}} \mathrm{d}\, g_{j,\mathrm{T}} \mathrm{d}\, \phi_{j,\mathrm{T}}$$
$$\overset{(b)}{=} \int_{g_{\mathrm{L,R}}}^{g_{\mathrm{U,R}}} \int_{g_{\mathrm{L,T}}}^{g_{\mathrm{U,T}}} \frac{1}{\sqrt{\pi g_{i,\mathrm{R}}^2 g_{j,\mathrm{T}}^2}} \exp\left(-\frac{y_{i,j}^2}{g_{i,\mathrm{R}}^2 g_{j,\mathrm{T}}^2}\right) \frac{1}{2 g_{\max,\mathrm{R}}} \frac{1}{2 g_{\max,\mathrm{T}}} \mathrm{d}\, g_{i,\mathrm{R}} \mathrm{d}\, g_{j,\mathrm{T}}. \tag{48}$$

where $g_{\mathrm{R,T}} \triangleq 4 g_{\max,\mathrm{R}} g_{\max,\mathrm{T}}$. Subsequently, using $y_{i,j} = z_{i,j} h_{i,j}^{\mathrm{R}}$ and the PDF of $z_{i,j}$ in (51), we can calculate the PDF of $y_{i,j}$ according to

$$f_{y_{i,j}}(y_{i,j}) = \int_{z_{i,j}} \frac{1}{|z_{i,j}|} f_{z_{i,j}}(z_{i,j}) f_{h_{i,j}^{\mathrm{R}}}\left(\frac{y_{i,j}}{z_{i,j}}\right) \mathrm{d}\, z_{i,j}$$
$$= \frac{1}{\sqrt{\pi}} \int_{z_{i,j}} \frac{1}{|z_{i,j}|} f_{z_{i,j}}(z_{i,j}) \exp\left(-\frac{y_{i,j}^2}{z_{i,j}^2}\right) \mathrm{d}\, z_{i,j}. \tag{52}$$

We further assume that $1 - g_{\max,\mathrm{R}} > 0$ and $1 - g_{\max,\mathrm{T}} > 0$. Plugging (51) into (52), we can get $f_{y_{i,j}}(y_{i,j})$ which is calculated as

$$f_{y_{i,j}}(y_{i,j}) = \frac{1}{4\sqrt{\pi} g_{\max,\mathrm{R}} g_{\max,\mathrm{T}}} \times \left[\int_{B_1}^{B_2} \frac{1}{z_{i,j}} \ln\left(\frac{z_{i,j}}{g_{\mathrm{L,R}} g_{\mathrm{L,T}}}\right) \times \exp\left(-\frac{y_{i,j}^2}{z_{i,j}^2}\right) \mathrm{d}\, z_{i,j} + \ln\left(\frac{g_{\mathrm{U,T}}}{g_{\mathrm{L,T}}}\right) \int_{B_2}^{B_3} \frac{1}{z_{i,j}} \exp\left(-\frac{y_{i,j}^2}{z_{i,j}^2}\right) \mathrm{d}\, z_{i,j} + \int_{B_3}^{B_4} \frac{1}{z_{i,j}} \ln\left(\frac{g_{U,R} g_{\mathrm{U,T}}}{z_{i,j}}\right) \exp\left(-\frac{y_{i,j}^2}{z_{i,j}^2}\right) \mathrm{d}\, z_{i,j} \right]. \tag{53}$$

For the case $y_{i,j} \neq 0$, we let $t = y_{i,j}/z_{i,j}$, by which we can further calculate each integration in (53) according to

$$\lambda_1 = \int_{B_1}^{B_2} \frac{1}{z_{i,j}} \ln\left(\frac{z_{i,j}}{g_{\mathrm{L,R}} g_{\mathrm{L,T}}}\right) \exp\left(-\frac{y_{i,j}^2}{z_{i,j}^2}\right) \mathrm{d}\, z_{i,j}$$
$$= \int_{y_{i,j}/B_1}^{y_{i,j}/B_2} \frac{1}{t} \ln\left(\frac{y_{i,j}}{t}\right) \exp(-t^2) \mathrm{d}\, t - \frac{1}{2} \ln(g_{\mathrm{L,R}} g_{\mathrm{L,T}}) \left(\Gamma\left(0, \frac{y_{i,j}^2}{B_2^2}\right) - \Gamma\left(0, \frac{y_{i,j}^2}{B_1^2}\right) \right), \tag{54a}$$

$$\lambda_2 = \int_{B_3}^{B_4} \frac{1}{z_{i,j}} \ln\left(\frac{g_{U,R} g_{\mathrm{U,T}}}{z_{i,j}}\right) \exp\left(-\frac{y_{i,j}^2}{z_{i,j}^2}\right) \mathrm{d}\, z_{i,j}$$
$$= \frac{1}{2} \ln(g_{\mathrm{U,R}} g_{\mathrm{U,T}}) \left(\Gamma\left(0, \frac{y_{i,j}^2}{B_4^2}\right) - \Gamma\left(0, \frac{y_{i,j}^2}{B_3^2}\right) \right) - \int_{y_{i,j}/B_3}^{y_{i,j}/B_4} \frac{1}{t} \ln \frac{y_{i,j}}{t} \exp(-t^2) \mathrm{d}\, t. \tag{54b}$$

$$\lambda_3 = \int_{B_3}^{B_4} \frac{1}{z_{i,j}} \ln\left(\frac{g_{U,R} g_{\mathrm{U,T}}}{z_{i,j}}\right) \exp\left(-\frac{y_{i,j}^2}{z_{i,j}^2}\right) \mathrm{d}\, z_{i,j}$$
$$= \frac{1}{2} \ln(g_{\mathrm{U,R}} g_{\mathrm{U,T}}) \left(\Gamma\left(0, \frac{y_{i,j}^2}{B_4^2}\right) - \Gamma\left(0, \frac{y_{i,j}^2}{B_3^2}\right) \right) - \int_{y_{i,j}/B_3}^{y_{i,j}/B_4} \frac{1}{t} \ln \frac{y_{i,j}}{t} \exp(-t^2) \mathrm{d}\, t. \tag{54c}$$

Then, substituting λ_1, λ_2 and λ_3 into (53) yields the result in (40) for $y_{i,j} \neq 0$. When $y_{i,j} = 0$, we can easily achieve the final result by setting $y_{i,j} = 0$ in (53) and reach the result in (40) for $y_{i,j} \neq 0$. The proof is completed by now.

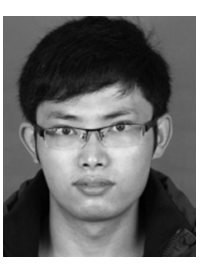
ACKNOWLEDGMENT

The authors would like to thank the Editor and Reviewers for the constructive comments on the previous manuscript. They also would like to thank Shiyu Liao for providing the significant comments to revise this manuscript.

REFERENCES

[1] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, Mar. 2015.
[2] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
[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. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 4th Quart., 2015.
[5] K. Zheng, L. Zhao, J. Mei, B. Shao, W. Xiang, and L. Hanzo, "Survey of large-scale MIMO systems," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1738–1760, 3rd Quart., 2015.
[6] L. Bin, T. W. Rondeau, J. H. Reed, and C. W. Bostian, "Analog-to-digital converters," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 69–77, Nov. 2005.
[7] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADcs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
[8] 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.
[9] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.
[10] T. C. Zhang, C. K. Wen, S. Jin, and T. Jiang, "Mixed-ADc massive MIMO detectors: Performance analysis and design optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738–7752, Nov. 2016.

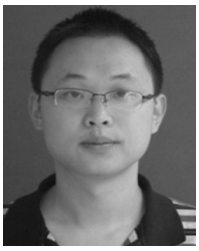
[11] S. Zarei, W. Gerstacker, J. Aulin, and R. Schober, "I/Q imbalance aware widely-linear receiver for uplink multi-cell massive MIMO systems: Design and sum rate analysis," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3393–3407, May 2016.
[12] N. Kolomvakis, M. Matthaiou, and M. Coldrey, "IQ imbalance in multiuser systems: channel estimation and compensation," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 3039–3051, Jul. 2016.
[13] C. Risi, D. Persson, and E. G. Larsson. (Apr. 2014). "Massive MIMO with 1-bit ADc." [Online]. Available: https://arxiv.org/abs/1404.7736v1
[14] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer. (Jul. 2016). "Throughput analysis of massive MIMO uplink with low-resolution ADcs." [Online]. Available: https://arxiv.org/abs/1602.01139v2
[15] Y. Li, C. Tao, L. Liu, G. Seco-Granados, and A. L. Swindlehurst, "Channel estimation and uplink achievable rates in one-bit massive MIMO systems," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.
[16] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, Jr., (May 2016). "Hybrid architectures with few-bit ADc receivers: Achievable rates and energy-rate tradeoffs." [Online]. Available: https://arxiv.org/abs/1605.00668v1
[17] L. Fan, S. Jin, C. K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADc," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.
[18] W. Tan, S. Jin, C. K. Wen, and Y. Jing, "Spectral efficiency of mixed-ADc receivers for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7841–7846, 2016.
[19] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADcs," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, May 2016.
[20] J. Mo and R. W. Heath, Jr., "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.
[21] U. Ugurlu and R. Wichman, "Enabling low-resolution ADc with high-order modulations for millimeter-wave systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
[22] J. Mo, P. Schniter, and R. W. Heath, Jr., (Oct. 2016). "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADcs." [Online]. Available: https://arxiv.org/abs/1610.02735v1
[23] W. B. Abbas, F. G. Cuba, and M. Zorzi. (Jul. 2016). "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADcs." [Online]. Available: https://arxiv.org/abs/1607.03725v1
[24] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.
[25] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, Jr., "Uplink performance of wideband massive MIMO with one-bit ADcs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.
[26] N. Liang and W. Zhang, "Mixed-ADc massive MIMO uplink in frequency-selective channels," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4652–4666, Nov. 2016.
[27] N. Liang and W. Zhang, "Mixed-ADc massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.
[28] J. Liu, J. Xu, W. Xu, S. Jin, and X. Dong, "Multiuser massive MIMO relaying with mixed-ADc receiver," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 76–80, Jan. 2017.
[29] J. Choi, J. Mo, and R. W. Heath, Jr., "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADcs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.
[30] Y. Li, C. Tao, L. Liu, A. Mezghani, and A. L. Swindlehurst, "How much training is needed in one-bit massive MIMO systems at low SNR?" in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
[31] L. Fan, D. Qiao, S. Jin, C. K. Wen, and M. Matthaiou, "Optimal pilot length for uplink massive MIMO systems with low-resolution ADc," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2016, pp. 1–5.
[32] A. Gokceoglu, E. Björnson, E. G. Larsson, and M. Valkama, "Spatio-temporal waveform design for multiuser massive MIMO downlink with 1-bit receivers," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 2, pp. 347–362, Mar. 2017.
[33] A. Gokceoglu, E. Björnson, E. G. Larsson, and M. Valkama, "Waveform design for massive MISO downlink with energy-efficient receivers adopting 1-bit ADcs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–7.
[34] L. F. Lin, W. H. Chung, H. J. Chen, and T. S. Lee, "Energy efficient hybrid precoding for multi-user massive MIMO systems using low-resolution ADcs," in *Proc. IEEE Int. Workshop Signal Process. Syst. (SiPS)*, Oct. 2016, pp. 115–120.
[35] M. Valkama, M. Renfors, and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2335–2344, Oct. 2001.
[36] A. Tarighat and A. H. Sayed, "MIMO OFDM receivers for systems with IQ imbalances," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3583–3596, Sep. 2005.
[37] A. Hakkarainen, J. Werner, K. R. Dandekar, and M. Valkama, "Analysis and augmented spatial processing for uplink OFDMA MU-MIMO receiver with transceiver I/Q imbalance and external interference," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3422–3439, May 2016.
[38] A. A. A. Boulogeorgos, V. M. Kapinas, R. Schober, and G. K. Karagiannidis, "I/Q-imbalance self-interference coordination," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4157–4170, Jun. 2016.
[39] X. Cheng, Z. Luo, and S. Li, "Joint estimation for I/Q imbalance and multipath channel in millimeter-wave SC-FDE systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 6901–6912, Sep. 2016.
[40] W. Zhang, R. C. de Lamare, C. Pan, and M. Chen, "Joint TX/RX IQ imbalance parameter estimation using a generalized system model," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2015, pp. 4704–4709.
[41] S. Wang and L. Zhang, "Signal processing in massive MIMO with IQ imbalance and low-resolution ADcs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8298–8312, Dec. 2016.
[42] W. Zhang, R. C. de Lamare, C. Pan, and M. Chen, "Widely linear block-diagonalization type precoding in massive MIMO systems with IQ imbalance," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Oct. 2015, pp. 1789–1794.
[43] R. Hamila, Ö. Özdemir, and N. Al-Dhahir, "Beamforming OFDM performance under joint phase noise and I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2978–2989, May 2016.
[44] 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.
[45] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.
[46] T. Schenk, *RF Imperfections in High-Rate Wireless Systems: Impact and Digitial Compensation*. Dordrecht, The Netherlands: Springer, 2008.
[47] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
[48] D. S. Palguna, D. J. Love, T. A. Thomas, and A. Ghosh, "Millimeter wave receiver design using low precision quantization and parallel $\Delta\Sigma$ architecture," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6556–6569, Oct. 2016.
[49] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, Apr. 2015.
[50] Y. Xiong, N. Wei, and Z. Zhang, "Compensation of IQ imbalance for quantized massive MIMO systems," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2016, pp. 1–5.
[51] S. Wu, L. Kuang, Z. Ni, J. Lu, D. D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.
[52] S. M. Ross, *Introduction to Probability Models*. Burlington, MA, USA: Academic, 2010.
[53] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Burlington, MA, USA: Academic, 2007.
[54] A. Papoulis and S. U. Pillai, *Probabilty, Random Variables and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 2002.



YOUZHI XIONG received the B.S. degree in electrical engineering from Henan University, Kaifeng, China in 2011, and the M.S. degree in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2014, where he is currently pursuing the Ph.D. degree. His research interests include massive MIMO, channel estimation, and data detection.

NING WEI received the Ph.D. degree in electrical engineering from the University of Electronic Science and Technology of China, in 2008. He was a Visitor of The University of Texas at Austin, Austin, TX, USA, in 2012. He is currently an Associate Professor with the University of Electronic Science and Technology of China. His research interests are wireless communications networks and signal processing, including ad hoc networks, multi-RAT, channel coding, MIMO system, cooperative communications, and physical-layer network coding.

ZHONGPEI ZHANG received the Ph.D. degree in traffic information engineering from Southwest Jiaotong University, Chengdu, China, in 2000. From 2001 to 2003, he held a post-doctoral position with Tsinghua University. From 2004 to 2005, he was a Research Scientist with Oulu University, Oulu, Finland. He is currently a Professor with the University of Electronic Science and Technology of China. His research interests include wireless communication networks, equalization and iterative receiver, space-time coding, and noncoherent detection algorithm.



BINRUI LI received the B.Eng. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, where he is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communication. His research interests include wireless communications and signal processing, including massive MIMO channel estimation, beamforming/precoding, compressed sensing, and pilot contamination elimination/reduction techniques.

YANG CHEN received the B.S. degree in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, and the M.S. degree in communication and information systems from Southwest Jiaotong University, Chengdu, China, in 2015. He is currently pursuing the Ph.D. degree in communication and information system with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China. His research interests include massive MIMO and physical layer security.