

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/151433>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

# Impact and Calibration of Nonlinear Reciprocity Mismatch in Massive MIMO Systems

Rongjiang Nie, Li Chen, Nan Zhao, *Senior Member, IEEE*,  
Yunfei Chen, *Senior Member, IEEE*, Weidong Wang, and Xianbin Wang, *Fellow, IEEE*

**Abstract**—Time-division-duplexing massive multiple-input multiple-output (MIMO) systems estimate the channel state information (CSI) by leveraging the uplink-downlink channel reciprocity, which is no longer valid when the mismatch arises from the asymmetric uplink and downlink radio frequency (RF) chains. Existing works treat the reciprocity mismatch as constant for simplicity. However, the practical RF chain consists of nonlinear components, which leads to nonlinear reciprocity mismatch. In this work, we examine the impact and the calibration approach of the nonlinear reciprocity mismatch in massive MIMO systems. To evaluate the impact of the nonlinear mismatch, we first derive the closed-form expression of the ergodic achievable rate. Then, we analyze the performance loss caused by the nonlinear mismatch to show that the impact of the mismatch at the base station (BS) side is much larger than that at the user equipment side. Therefore, we propose a calibration method for the BS. During the calibration, polynomial function is applied to approximate the nonlinear mismatch factor, and over-the-air training is employed to estimate the polynomial coefficients. After that, the calibration coefficients are computed by maximizing the downlink achievable rate. Simulation results are presented to verify the analytical results and to show the performance of the proposed calibration approach.

**Index Terms**—Calibration, massive MIMO, nonlinear RF chain, reciprocity mismatch.

## I. INTRODUCTION

Massive multiple-input multiple-out (MIMO) has been identified as one of the key enabling technologies for the 5th generation (5G) mobile communication networks [1]. With a great number of antennas deployed at the base station (BS), massive MIMO can significantly improve the capacity, throughput, and spectral efficiency of wireless communication systems [2].

Since the antenna number is large, the acquisition of the channel state information (CSI) becomes a great challenge [3].

This research was supported by National Key Research and Development Project (Grant NO. 2018YFB1801105), National Natural Science Foundation of China (Grant NO. 62071445), and Fundamental Research Funds for the Central Universities (Grant NO. WK3500000007). This paper was presented in part at the IEEE Wireless Communications and Networking Conference (WCNC), Nanjing, China, March 2021. (*Corresponding author: Li Chen*)

Rongjiang Nie, Li Chen, and Weidong Wang are with the CAS Key Laboratory of Wireless Optical Communication, University of Science and Technology of China, Hefei 230052, China (e-mail: johnnrj@mail.ustc.edu.cn; {chenli87, wdwang}@ustc.edu.cn).

Nan Zhao is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: zhaonan@dut.edu.cn).

Yunfei Chen is with the School of Engineering, University of Warwick, Coventry CV4 7AL, U.K. (e-mail: Yunfei.Chen@warwick.ac.uk).

Xianbin Wang is with the Department of Electrical and Computer Engineering, Western University, London, ON N6A 5B9, Canada (e-mail: xianbin.wang@uwo.ca).

To reduce the heavy overhead in obtaining the downlink CSI, massive MIMO systems are typically assumed to operate in time division duplexing (TDD) mode, and BS estimates the downlink CSI from the uplink pilots transmitted by the user equipment (UE) via exploiting the channel reciprocity [4]. In practice, the channel observed by the baseband processor consists of not only the reciprocal wireless propagation channel, but also the radio frequency (RF) gains resulting from the frequency responses of hardware devices, i.e., high power amplifier (HPA), filters, analog-to-digital converter (ADC), and digital-to-analog converter (DAC) [5]. Due to the involvement of different devices, the overall gains of the RF chains at BS and UEs are typically asymmetric, which leads to the reciprocity mismatch of the uplink and downlink channels [6].

To study how the reciprocity mismatch affects the system performance, previous works have investigated the impact of the mismatch on the MIMO system with linear precoding techniques, e.g., zero-forcing (ZF) and matched filter (MF). Generally, the reciprocity mismatch can degrade the performances of both ZF and MF precoding techniques [7]. By comparing the performances of ZF and MF with the mismatch, it can be found that ZF is more sensitive to the reciprocity mismatch and its performance loss is much larger than that of MF, especially in the high signal-to-noise ratio (SNR) regime [8]. As the antenna number of the BS grows, the performances of ZF and MF are asymptotically identical [9]. Wei *et al.* in [10] analyzed the impact of the reciprocity mismatch at the BS side and the UE side, and showed that both the reciprocity mismatch at the BS side and the UE side degraded the system performance. It is noteworthy that the performance degradation caused by the mismatch at the BS side is much more crucial than that at the UE side. The result reveals that it is important to calibrate the reciprocity mismatch at the BS side. Further, the experimental results in [11] verified the performance loss due to the reciprocity mismatch in the practical precoded system.

As the reciprocity mismatch severely degrades the system performance, reciprocity calibration has attracted great attentions in the past decade. Reciprocity calibration can be divided into two types: hardware-circuit calibration and over-the-air (OTA) calibration [12]. The hardware-circuit calibration requires auxiliary components, such as switches and couplers, to connect transmit antennas and receive antennas. Nishimori *et al.* in [13] first proposed an automatic hardware-circuit calibration for the conventional MIMO system. The calibration approach can use the transmit signal to realize a real-time calibration during the BS transmits data streams.

Then, the hardware-circuit calibration was applied to the wideband wireless system in [14], where different subcarriers were independently calibrated. In [15], an auxiliary calibration transceiver structure was proposed for the MIMO system. With the auxiliary transceiver, the calibration loop can be easily extended to the BS with multiple access points, but it inevitably increased the cost of the hardware components. To simplify both the hardware structure and the signal processing, Liu *et al.* proposed a calibration board consisting of switches and attenuators in [16]. While hardware-circuit calibration works efficiently in conventional MIMO systems with low cost of hardware components, it becomes costly in massive MIMO systems due to the large number of channels to be calibrated. To reduce the cost, a daisy chain interconnection topology of the circuits was applied to hardware-circuit calibration [17], which could reduce the transceiver interconnection effort. To further relieve the connection effort and improve the calibration performance, an optimal interconnection of the hardware-circuit calibration was presented in [18] by minimizing the Cramer-Rao lower bound in the calibration coefficients.

Unlike the hardware-circuit calibration, the OTA calibration only needs to gather the training signals among uncalibrated antennas. An OTA calibration called relative calibration was first proposed to calibrate the TDD single-input single-output system in the frequency domain [19]. Then, Kaltenberger *et al.* proposed an OTA calibration for the multi-user MIMO system in [20], which involved both BS and UEs, and they applied the total least square (LS) approach to compute the calibration coefficients. For the wideband system, the frequency-domain calibration has to be applied in each subcarrier, which means that the overhead and complexity of the frequency-domain calibration increases with the number of subcarriers. To reduce the calibration overhead and complexity for the wideband system, a time-domain reciprocity calibration was presented in [21], since the number of parameters in the time domain is much less than that in the frequency domain.

In massive MIMO systems, the reciprocity calibration method designed for conventional MIMO systems encounter new challenges due to the heavy overhead of feeding back the CSI from UEs. Based on theoretical and experimental results that the reciprocity mismatch at the single-antenna UE causes negligible impact compared with the reciprocity mismatch at the BS side equipped with large-scale antenna array [22], several calibration approaches called “single-side” or “one-side” calibration were presented for the massive MIMO system, which only calibrate the antennas at the BS. In [23], a single-side calibration method was presented for the massive MIMO Argos prototype. The performance of the Argos calibration is sensitive to the fading channel and relies on the location of the reference antenna. To overcome the shortages of Argos, H. Wei *et al.* in [24] presented a mutual coupling calibration, which utilized the strong mutual coupling effect among adjacent antennas rather than the fading channel. To compute the calibration coefficients efficiently, the LS method was applied to compute the coefficients in [25]. In [26], an OTA calibration framework was proposed based on some existing calibration schemes. As for distributed massive MIMO systems, a two-stage calibration approach was

proposed in [27] to reduce the overhead of the CSI feedback among remote access points.

All the above mentioned works have simply treated the reciprocity mismatch as a constant. However, as the reciprocity mismatch arises from the asymmetry of the transmit and receive RF chains, and practical RF chains are generally composed of nonlinear components, e.g., the nonlinear HPA [28], the reciprocity mismatch is also nonlinear, which is called nonlinear reciprocity mismatch. With the expected use of millimeter waveband and low-cost RF devices in massive MIMO, nonlinearity will become more severe for 5G and beyond networks. Hence, the nonlinear mismatch needs to be studied in massive MIMO systems. Compared with the linear reciprocity calibration, the nonlinear calibration has three great challenges. As the reciprocity mismatch factor varies with the transmit power, a simple training scheme with the single-power pilot is no longer applicable. Due to the complex expression of the transform characteristic of the nonlinear components, it is difficult to determine the relationship function between the mismatch factor and the transmit power. Further, the relationships between the calibration coefficients are also nonlinear functions so that it is difficult to solve the calibration coefficients.

The nonlinearity compensation of transmitters, e.g., the nonlinear predistortion, has been extensively studied in wireless systems [29]–[31]. The works on nonlinearity only focused on offsetting the defect of the single RF chain. Since the reciprocity mismatch is jointly caused by the defects of the transmit chain and the receive chain, existing solutions to the nonlinear transmitter compensation techniques are not sufficient to address the nonlinear reciprocity mismatch in massive MIMO systems.

Motivated by the above observations, we investigate the nonlinear reciprocity mismatch in TDD multi-user massive MIMO systems. To study the impact of the nonlinear reciprocity mismatch, we first derive the closed-form expression of the ergodic achievable rate. Then, the performance loss due to the mismatch at the BS side and UE side is analyzed, respectively. Based on these analytical results, we propose a novel nonlinear reciprocity calibration approach. To sample the nonlinear response of the RF chain varying with the transmit power, an OTA-polynomial training approach is employed in the calibration. Specifically, the polynomial functions are applied to characterize the relationship between the mismatch factor and the transmit power. The polynomial coefficients are estimated by the OTA training with multi-power training pilots. Finally, to compute the nonlinear calibration coefficients efficiently, we formulate an optimization problem which seeks to maximize the downlink achievable rate. The main contributions of the work can be summarized as follows.

- **Impact analysis of the nonlinear reciprocity mismatch:** Under the nonlinear reciprocity mismatch, we first derive the closed-form expression of the downlink ergodic achievable rate for the multi-user massive MIMO system with ZF precoding. Based on this, the impact of the nonlinear reciprocity mismatch on the system performance is examined.

- **OTA-polynomial training of nonlinear mismatch factors:** To estimate the nonlinear response of the transmit RF chain varying with the transmit power, we propose the OTA-polynomial training approach. The polynomial fitting approach is applied to approximate the nonlinear function between the nonlinear reciprocity mismatch factor and the transmit power. After that, the OTA training with multi-power training pilots is applied to estimate the polynomial coefficients.
- **Toward optimal nonlinear calibration coefficients:** To compute the nonlinear calibration coefficients efficiently, we formulate an optimization problem seeking to maximize the downlink achievable rate. A fast algorithm based on sequential linear programming is presented to efficiently solve the nonlinear calibration coefficients.

The rest of the paper is organized as follows. Section II describes the system model. The impact of the nonlinear reciprocity mismatch is analyzed in section III. In section IV, nonlinear reciprocity calibration is proposed, including the multi-power training pilots, the nonlinear reciprocity mismatch polynomial fitting, and the optimal nonlinear calibration coefficients. Simulation and numerical results are given in Section V, and the conclusion is given in VI.

Throughout the paper, vectors and matrices are denoted in bold lowercase and uppercase respectively:  $\mathbf{a}$  and  $\mathbf{A}$ . Let  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ , and  $\mathbf{A}^{-1}$  denote the transpose, conjugate transpose, and inverse of a matrix  $\mathbf{A}$  respectively.  $\text{tr}(\cdot)$  stands for the trace operator and  $\mathbb{E}(\cdot)$  represents the expectation operation. Let  $|a|$  denote the amplitude of the complex number  $a$ .  $\text{diag}(a_1, \dots, a_N)$  denotes a  $N$  by  $N$  diagonal matrix with diagonal entries given by  $a_1, \dots, a_N$ .  $\mathcal{N}(\mu, \sigma^2)$  and  $\mathcal{CN}(\mu, \sigma^2)$  represent for normal distribution and complex normal distribution with mean  $\mu$  and variance  $\sigma^2$ , respectively.  $\mathcal{U}(a, b)$  denotes uniform distribution on the interval  $[a, b]$ .  $\mathbb{C}$  and  $\mathbb{R}$  stand for the complex numbers and real numbers, respectively. Let  $[1 : N]$  denote the set  $\{1, 2, \dots, N\}$ .

## II. SYSTEM MODEL

We consider a multi-user massive MIMO system consisting of  $K$  single-antenna UEs and a BS equipped with  $M$  antennas, as illustrated in Fig. 1, where  $M$  and  $K$  are large. The transmit RF chains are subject to a typical nonlinear device, i.e., nonlinear HPA<sup>1</sup>. Without loss of generality, a memoryless HPA model called Solid State Power Amplifier (SSPA) is considered in the system, which has been extensively used to characterize the HPA [29].

Let  $x_{u,k}$  be the transmit signal of the  $k$ -th UE and  $x_{b,m}$  be the transmit signal at the  $m$ -th antenna of the BS. The transform functions of the HPAs can be characterized as

$$\hat{x}_{u,k} = \frac{\sqrt{b_0} v_k x_{u,k}}{\sqrt[2v]{1 + \left(\frac{|x_{u,k}|}{B_{\text{sat},k}}\right)^{2v}}} = \sqrt{b_0} \mathcal{B}_k(|x_{u,k}|) x_{u,k}, \quad (1)$$

<sup>1</sup>Although, in addition to HPAs, the nonlinearity may also be caused by other devices, the HPA is the major source of nonlinearity, and most presented mathematical models have been originally introduced for HPAs [32]. Therefore, we consider the nonlinearity mainly caused by HPAs.

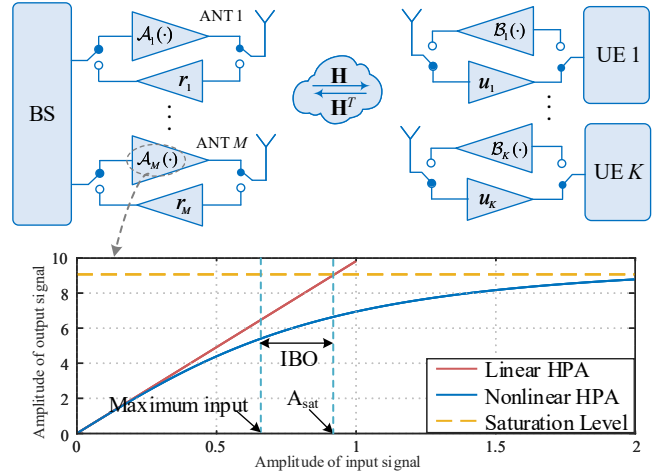


Fig. 1. A multi-user massive MIMO system with reciprocity mismatch.

$$\hat{x}_{b,m} = \frac{\sqrt{a_0} t_m x_{b,m}}{\sqrt[2v]{1 + \left(\frac{|x_{b,m}|}{A_{\text{sat},m}}\right)^{2v}}} = \sqrt{a_0} \mathcal{A}_m(|x_{b,m}|) x_{b,m}, \quad (2)$$

where  $\mathcal{B}_k(x) = v_k / \sqrt[2v]{1 + (|x|/B_{\text{sat},k})^{2v}}$  denotes the transform function of the HPA at the  $k$ -th UE,  $\mathcal{A}_m(x) = t_m / \sqrt[2v]{1 + (|x|/A_{\text{sat},m})^{2v}}$  represents the transform function of the  $m$ -th HPA at the BS,  $\nu$  refers to the smoothness factor,  $a_0 \in \mathbb{R}$  denotes the small-signal amplification gain of the HPAs at the BS,  $A_{\text{sat},m} = A_{\text{sat}} a_m \in \mathbb{R}$  is the saturation level of the  $m$ -th HPA at the BS,  $a_m \in \mathbb{R}$  represents the various saturation of different HPAs,  $t_m \in \mathbb{C}$  is the vibration of small-signal gain  $a_0$ ,  $b_0 \in \mathbb{R}$  denotes the small-signal amplification gain of the HPAs at UEs,  $B_{\text{sat},k} \in \mathbb{R}$  is the saturation level of HPAs of UEs,  $v_k \in \mathbb{C}$  represents the vibration of the small-signal gain  $b_0$ . Both  $a_m$ ,  $v_k$ , and  $|t_m|$  follow the log-normal distribution, i.e.,  $\ln a_m \sim \mathcal{N}(0, \delta_a^2)$ ,  $\ln v_k \sim \mathcal{N}(0, \delta_v^2)$ , and  $\ln |t_m| \sim \mathcal{N}(0, \delta_t^2)$ , and  $\angle t_m$  obeys the uniform distribution as  $\angle t_m \sim \mathcal{U}(-\theta_t, \theta_t)$ . Moreover, we use input back-off (IBO) to measure the relationship between the input power and the saturation level of the HPA. From [29], IBO is defined as

$$\text{IBO} = 10 \log_{10} \left( \frac{A_{\text{sat}}}{\sigma_x} \right), \quad (3)$$

where  $\sigma_x^2$  is the average power of the input signal.

The overall channel observed by the baseband processor is composed of the reciprocal wireless propagation channel and the non-reciprocal RF gain. Let  $\mathbf{H} = \mathbf{\Phi} \mathbf{H}_r \in \mathbb{C}^{k \times M}$  denote the wireless propagation channel, where  $\mathbf{H}_r \in \mathbb{C}^{k \times M}$  is the Rayleigh fading channel with each entry following  $\mathcal{CN}(0, 1)$ ,  $\mathbf{\Phi} = \text{diag}(\phi_1, \dots, \phi_K) \in \mathbb{R}^{K \times K}$ , and  $\phi_k$  denotes the large-scale path loss between the BS and the  $k$ -th UE. By considering the defects of the transmit and receive RF chain, the overall uplink and downlink channels can be modeled as

$$\mathbf{H}_{\text{UL}} = \mathbf{R} \mathbf{H}^T \mathbf{B}, \quad (4a)$$

$$\mathbf{H}_{\text{DL}} = \mathbf{U} \mathbf{H} \mathbf{A}, \quad (4b)$$

where  $\mathbf{R} = \text{diag}(r_1, \dots, r_M)$ ,  $r_m \in \mathbb{C}$  denotes the receive RF gain of the  $m$ -th antenna of the BS,  $\mathbf{A} =$



$\text{diag}(\mathcal{A}_1(|x_{b,1}|), \dots, \mathcal{A}_M(|x_{b,M}|))$ ,  $\mathbf{U} = \text{diag}(u_1, \dots, u_K)$ ,  $u_k \in \mathbb{C}$  is the receive RF gain of the  $k$ -th UE, and  $\mathbf{B} = \text{diag}(b_1, \dots, b_K)$ ,  $b_k = \mathcal{B}_k(|x_{u,k}|) \in \mathbb{C}$ . Both the amplitudes of  $r_m$  and  $u_k$  follow log-normal distribution, i.e.,  $\ln |r_m| \sim \mathcal{N}(0, \delta_r^2)$ ,  $\ln |u_k| \sim \mathcal{N}(0, \delta_u^2)$ , and their phases obey uniform distribution, i.e.,  $\angle r_m \sim \mathcal{U}(-\theta_r, \theta_r)$ ,  $\angle u_k \sim \mathcal{U}(-\theta_u, \theta_u)$ .

During the downlink phase, the BS transmits the precoded signal to UEs. Let  $\mathbf{y} = [y_1, \dots, y_K]^T$  denote the signal vector received by all UEs, where  $y_k$  is the received signal at the  $k$ -th UE. The downlink signal can be represented as

$$\mathbf{y} = \sqrt{a_0} \mathbf{H}_{\text{DL}} \underbrace{\mathbf{W} \mathbf{s}}_{\mathbf{x}_b} + \mathbf{n}, \quad (5)$$

where  $\mathbf{s} \in \mathbb{C}^{K \times 1}$  is the symbol vector with zero mean and variance  $\mathbb{E}\{\mathbf{s} \mathbf{s}^H\} = \rho_t \mathbf{I}_K$ ,  $\rho_t \in \mathbb{R}$  represents the average transmit power of the BS,  $\mathbf{W} \in \mathbb{C}^{M \times K}$  denotes the precoding matrix, and  $\mathbf{n}$  denotes the additive white Gaussian noise (AWGN) vector with  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_K)$ . In this paper, we consider the ZF precoding technique, and the precoding matrix can be given by

$$\mathbf{W} = \frac{1}{\sqrt{\beta_{\text{ZF}}}} \mathbf{H}_{\text{UL}}^* (\mathbf{H}_{\text{UL}}^T \mathbf{H}_{\text{UL}}^*)^{-1}, \quad (6)$$

where  $\beta_{\text{ZF}}$  denotes the normalization scalar defined as

$$\beta_{\text{ZF}} = \mathbb{E}\{\text{tr}[(\mathbf{H}_{\text{UL}}^T \mathbf{H}_{\text{UL}}^*)^{-1}]\}. \quad (7)$$

As the overall downlink and uplink channels are not reciprocal, the multi-user massive MIMO system suffers from the reciprocity mismatch. Since the multiplicative RF matrices  $\mathbf{A}$  and  $\mathbf{B}$  are related to the transmit power, we call it nonlinear reciprocity mismatch.

### III. PERFORMANCE ANALYSIS OF NONLINEAR RECIPROCITY MISMATCH

In this section, we first introduce the performance analysis of the existing linear reciprocity mismatch. Then, the closed-form expression of the ergodic achievable rate is derived for the massive MIMO system in the presence of the nonlinear reciprocity mismatch. Based on the closed-form expression, we further analyze the impact of the nonlinear reciprocity mismatch on the performance of the downlink transmission.

#### A. Performance of Existing Linear Reciprocity Mismatch

To analyze the system performance, we refer to the technique in [33] which is widely used in the field of massive MIMO [8], [9], [34]. This technique only utilize the statistical channel to detect the received signals, when UEs do not have any channel estimate. Therefore, by regarding the reciprocity mismatch as the constant [8]–[10], i.e.,  $\mathcal{A}_m(|x_{b,m}|) = t_m$  and  $\mathcal{B}_k(|x_{u,k}|) = v_k$ , the received signal (5) can be further denoted as

$$\begin{aligned} y_k &= \sqrt{a_0} \underbrace{u_k \mathbf{h}_k \mathbf{T} \mathbf{w}_k}_{h_{\text{eq},k,k}^{\text{lr}}} s_k + \sqrt{a_0} \sum_{i \neq k}^K \underbrace{u_k \mathbf{h}_k \mathbf{T} \mathbf{w}_i}_{h_{\text{eq},k,i}^{\text{lr}}} s_i + n_k \\ &= \underbrace{\sqrt{a_0} \mathbb{E}\{h_{\text{eq},k,k}^{\text{lr}}\}}_{s_{\text{lr},k}^{\text{ES}}} s_k + n_{\text{lr},k}^{\text{FCI}} + n_{\text{lr},k}^{\text{MUI}} + n_k, \end{aligned} \quad (8)$$

where  $\mathbf{h}_k$  is the  $k$ -th row of  $\mathbf{H}$ ,  $\mathbf{w}_k$  denotes the  $k$ -th column of  $\mathbf{W}$ ,  $\mathbf{T} = \text{diag}(t_1, \dots, t_M)$ ,  $s_{\text{lr},k}^{\text{ES}}$  denotes the effective useful signal received by the UE  $k$ ,  $n_{\text{lr},k}^{\text{FCI}}$  and  $n_{\text{lr},k}^{\text{MUI}}$  can be given by

$$n_{\text{lr},k}^{\text{FCI}} = \sqrt{a_0} (h_{\text{eq},k,k}^{\text{lr}} - \mathbb{E}\{h_{\text{eq},k,k}^{\text{lr}}\}) s_k, \quad (9)$$

$$n_{\text{lr},k}^{\text{MUI}} = \sqrt{a_0} \sum_{i \neq k}^K h_{\text{eq},k,i}^{\text{lr}} s_i. \quad (10)$$

Since the residual  $(h_{\text{eq},k,k}^{\text{lr}} - \mathbb{E}\{h_{\text{eq},k,k}^{\text{lr}}\})$  captures the fluctuation of the precoded channel,  $n_{\text{lr},k}^{\text{FCI}}$  denotes the interference resulting from the fluctuation of the precoded channel and can be called fluctuant-channel interference (FCI). Further,  $n_{\text{lr},k}^{\text{MUI}}$  represents the multiple-user interference (MUI) due to the signal for other UEs.

To study the impact of the reciprocity mismatch, the ergodic achievable rate is adopt to characterize the performance of the massive MIMO system as follows.

**Definition 1** (Downlink ergodic achievable rate). According to [6], [8], [9], the ergodic achievable rate can be defined as

$$R_k^{\text{lr}} = \log(1 + \gamma_k^{\text{lr}}), \quad \forall k \in [1 : K], \quad (11)$$

where  $\text{SINR}_k$  denotes the signal-to-interference-and-noise ratio (SINR) at the  $k$ -th UE as

$$\gamma_k^{\text{lr}} = \frac{|\mathbb{E}\{h_{\text{eq},k,k}^{\text{lr}}\}|^2}{\text{Var}\{h_{\text{eq},k,k}^{\text{lr}}\} + \sum_{i \neq k}^K \mathbb{E}\{|h_{\text{eq},k,i}^{\text{lr}}|^2\} + \frac{\sigma_n^2}{a_0 \rho_t}}. \quad (12)$$

Then, the closed-form expression of  $\text{SINR}_k$  for the massive MIMO system can be denoted as follows.

**Lemma 1** (Closed-form expression of SINR). According to [6], [9], the downlink SINR for ZF with the linear reciprocity mismatch can be given by

$$\gamma_{\text{ZF},k}^{\text{lr}} = \frac{a_0 \rho_t \frac{M-K}{\text{tr}\{\mathbf{\Phi}^{-2}\}} \text{sinc}^2(\theta_t) \text{sinc}^2(\theta_r)}{e^{\delta_r^2 + \delta_v^2 - \delta_t^2} (a_0 \rho_t \phi_k^2 \frac{M-K}{M} \varepsilon_1 + \sigma_n^2)}, \quad (13)$$

where  $\varepsilon_1 = e^{2\delta_t^2} + e^{2\delta_r^2} - 2\text{sinc}(\theta_t)\text{sinc}(\theta_r)e^{(\delta_t^2 - \delta_r^2)/2}$ .

Based on (13), we can further derive the ideal achievable rate of the massive MIMO with the ZF precoding. By setting  $\delta_t = \delta_r = \delta_v = 0$  and  $\theta_t = \theta_r = 0$ , the  $\gamma_{\text{ZF},k}^{\text{lr}}$  can degenerate into the ideal SNR denoted as

$$\gamma_{\text{ideal}} = \frac{M-K}{\text{tr}\{\mathbf{\Phi}^{-2}\}} \cdot \frac{a_0 \rho_t}{\sigma_n^2}. \quad (14)$$

By substituting (14) into (11), the ideal downlink achievable rate can be given by  $R_{\text{ideal}} = \log(1 + \gamma_{\text{ideal}})$ .

By comparing (13) and (14), it can be seen that the reciprocity mismatch leads to the interference which degrades the SINR. Although Lemma 1 shows the impact of the linear reciprocity mismatch, the nonlinear reciprocity mismatch has never been studied. Therefore, in the remainder of this section, we focus on the impact of the nonlinear reciprocity mismatch.

### B. Downlink SINDR and Ergodic Achievable Rate

To derive the closed-form expression of the achievable rate, we apply the Bussgang's theory [29], [35] to characterize the nonlinearities of the HPAs at the BS side. The theory states that the output of the nonlinear HPA can be expressed in terms of a linear scale parameter of the input signal and a nonlinear distortion which obeys the complex circular Gaussian distribution and is independent of the input. At the BS side, the output signal of HPAs can be further denoted as follows.

**Lemma 2** ([29, Eq. (18)]). By employing the Bussgang's linearization theory, the transform characteristics of the SSPA can be further characterized as

$$\hat{x}_{b,m} = g_m x_{b,m} + d_m, \quad \forall m \in [1 : M], \quad (15)$$

where  $g_m$  is the linear scale parameter denoted as

$$g_m = \frac{\mathbb{E} \left\{ x_{b,m}^* \hat{x}_{b,m} \right\}}{\mathbb{E} \left\{ |x_{b,m}|^2 \right\}} = t_m \mu \left( \frac{A_{\text{sat},m}}{\sigma_{x,m}} \right), \quad (16)$$

$\mu(x) = \frac{x}{2} [2x - \sqrt{\pi} \text{erfc}(x) \exp(x^2)(2x^2 - 1)]$ ,  $d_m \in \mathbb{C}$  denotes the nonlinear Gaussian distortion with zero mean and variance  $\sigma_{d,m}^2$  as

$$\sigma_{d,m}^2 = \mathbb{E} \left\{ |\hat{x}_{b,m}|^2 \right\} - |g_m|^2 \mathbb{E} \left\{ |x_{b,m}|^2 \right\} = |t_m|^2 \lambda_m(\sigma_{x,m}) \quad (17)$$

with

$$\lambda_m(x) = A_{\text{sat},m}^2 + \frac{A_{\text{sat},m}^4}{x^2} \exp\left(\frac{A_{\text{sat},m}^2}{x^2}\right) \bullet \text{Ei}\left(-\frac{A_{\text{sat},m}^2}{x^2}\right) - x^2 \mu\left(\frac{A_{\text{sat},m}}{x}\right)^2, \quad (18)$$

and  $\text{Ei}(x) = \int_{-\infty}^x t^{-1} e^t dt$  ( $x < 0$ ) is an exponential integral function.

According to Lemma 2, the downlink signal received by UEs can be rewritten as

$$\begin{aligned} \mathbf{y} &= \sqrt{a_0} \mathbf{U} \mathbf{H} \hat{\mathbf{x}}_b + \mathbf{n} \\ &= \sqrt{a_0} \mathbf{U} \mathbf{H} \mathbf{G} \mathbf{x}_b + \sqrt{a_0} \mathbf{U} \mathbf{H} \mathbf{d} + \mathbf{n}, \end{aligned} \quad (19)$$

where  $\mathbf{G} = \text{diag}(g_1, \dots, g_M)$  and  $\mathbf{d} = [d_1, \dots, d_M]^T$ .

Similar to (8), the UEs detect the signals by exploiting the statistical effective channel, and the signal received by the  $k$ -th UE can be further denoted as

$$y_k = \underbrace{\sqrt{a_0} \mathbb{E} \{ h_{\text{eq},k,k} \}}_{s_k^{\text{ES}}} s_k + n_k^{\text{FCI}} + n_k^{\text{MUI}} + n_k^{\text{NLD}} + n_k, \quad (20)$$

where  $h_{\text{eq},k,k} = u_k \mathbf{h}_k \mathbf{G} \mathbf{w}_k$  denotes the equivalent channel gain,  $s_k^{\text{ES}}$  denotes the effective useful signal received by the UE  $k$ ,  $n_k^{\text{FCI}}$  denotes the interference resulting from the fluctuation of the precoded channel,  $n_k^{\text{MUI}}$  represents the multiple-user interference,  $n_k^{\text{FCI}}$  and  $n_k^{\text{MUI}}$  can be expressed as

$$n_k^{\text{FCI}} = \sqrt{a_0} (h_{\text{eq},k,k} - \mathbb{E} \{ h_{\text{eq},k,k} \}) s_k, \quad (21)$$

$$n_k^{\text{MUI}} = \sqrt{a_0} \sum_{i \neq k}^K h_{\text{eq},k,i} s_i, \quad (22)$$

and  $n_k^{\text{NLD}}$  denotes the received nonlinear distortion for the  $k$ -th UE due to the nonlinear HPAs of the BS

$$n_k^{\text{NLD}} = \sqrt{a_0} u_k \mathbf{h}_k \mathbf{d}. \quad (23)$$

Compared with the linear reciprocity mismatch, the signal in the presence of the nonlinear reciprocity mismatch has an extra nonlinear distortion term  $\text{NLD}_k$ . Therefore, we use the signal-to-interference-plus-noise-and-distortion ratio (SINDR) to measure the ratio between the average power of the useful signal and the sum power of the interference, noise, and distortion.

**Definition 2** (Downlink SINDR and achievable rate). Based on (20), the SINDR at the  $k$ -th UE can be defined as (24) shown at the top of the next page, where  $\Sigma_d = \text{diag}(\sigma_{d,1}^2, \dots, \sigma_{d,M}^2)$  is the variance matrix of the distortion. The downlink ergodic achievable rate in the presence of the nonlinear reciprocity mismatch can be defined as

$$R_k = \log(1 + \gamma_k), \quad \forall k \in [1 : K]. \quad (25)$$

As UEs treat the sum of the noise and the interference as equivalent AWGN uncorrelated with the useful signal, the ergodic rate in Definition 2 is upper bounded by the capacity and legitimately achievable [6], [8], [9]. Then, we can derive the closed-form expression of the SINDR for the multi-user massive MIMO system with ZF in the presence of the nonlinear reciprocity mismatch.

**Proposition 1** (SINDR with nonlinear reciprocity mismatch). For a multi-user massive MIMO system with the ZF precoding, the closed-form expression of the SINDR in the presence of the nonlinear HPA and the reciprocity mismatch can be denoted as

$$\gamma_{\text{ZF},k}^{\text{ES}} = \frac{\mathcal{Y}_{\text{ZF},k}^{\text{ES}}}{\mathcal{Y}_{\text{ZF},k}^{\text{FCI}} + \mathcal{Y}_{\text{ZF},k}^{\text{MUI}} + \mathcal{Y}_{\text{ZF},k}^{\text{NLD}} + \sigma_n^2}, \quad (26)$$

where  $\mathcal{Y}_{\text{ZF},k}^{\text{ES}}$  denotes the power of the effective signal given by

$$\mathcal{Y}_{\text{ZF},k}^{\text{ES}} = \frac{a_0 \rho_t (M - K) |u_k \text{tr} \{ \mathbf{G}_{\text{ZF}} \mathbf{R}^* \}|^2}{M |b_k|^2 \text{tr} \{ (\mathbf{B} \Phi^2 \mathbf{B}^*)^{-1} \} \text{tr} \{ \mathbf{R} \mathbf{R}^* \}} \quad (27)$$

with  $\mathbf{G}_{\text{ZF}} = \text{diag}(g_{\text{ZF},1}, \dots, g_{\text{ZF},M})$  and  $g_{\text{ZF},m} = t_m \mu \left[ A_{\text{sat},m} \sqrt{\text{tr} \{ \mathbf{R} \mathbf{R}^* \}} / (\sqrt{|r_m|^2 \rho_t}) \right]$ .  $\mathcal{Y}_{\text{ZF},k}^{\text{FCI}}$  denotes the power of the fluctuant-channel interference and can be given by

$$\mathcal{Y}_{\text{ZF},k}^{\text{FCI}} = \frac{a_0 \rho_t |u_k|^2 \text{tr} \{ (\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}}^* - \alpha^* \mathbf{R}^*) \}}{(M - K)^{-1} M^2 |b_k|^2 \text{tr} \{ (\mathbf{B} \Phi^2 \mathbf{B}^*)^{-1} \}} \quad (28)$$

with  $\alpha = \frac{1}{M} \text{tr} \{ \mathbf{R}^{-1} \mathbf{G}_{\text{ZF}} \}$ . Further,  $\mathcal{Y}_{\text{ZF},k}^{\text{MUI}}$  is the power of the multi-user interference, which can be denoted as

$$\mathcal{Y}_{\text{ZF},k}^{\text{MUI}} = \sum_{i \neq k}^K \frac{a_0 \rho_t \phi_k^2 \text{tr} \{ (\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}}^* - \alpha^* \mathbf{R}^*) \}}{(M - K)^{-1} M^2 |u_k^{-1} b_i|^2 \phi_i^2 \text{tr} \{ (\mathbf{B} \Phi^2 \mathbf{B}^*)^{-1} \}}. \quad (29)$$

Finally,  $\mathcal{Y}_{\text{ZF},k}^{\text{NLD}}$  is the power of the nonlinear distortion given by

$$\mathcal{Y}_{\text{ZF},k}^{\text{NLD}} = a_0 |u_k|^2 \phi_k^2 \text{tr} \{ \Sigma_{\text{ZF}} \} \quad (30)$$

$$\gamma^k = \frac{a_0 \rho_t |\mathbb{E}\{h_{\text{eq},k,k}\}|^2}{a_0 \rho_t \text{Var}\{h_{\text{eq},k,k}\} + a_0 \rho_t \sum_{i \neq k}^K \mathbb{E}\{h_{\text{eq},k,i}\}^2 + a_0 |u_k|^2 \mathbb{E}\left\{\|\mathbf{h}_k \boldsymbol{\Sigma}_d^{\frac{1}{2}}\|^2\right\} + \sigma_n^2}, \quad (24)$$

with  $\boldsymbol{\Sigma}_{\text{ZF}} = \text{diag}(\sigma_{\text{ZF},1}^2, \dots, \sigma_{\text{ZF},M}^2)$  and  $\sigma_{\text{ZF},m}^2 = |t_m|^2 \lambda_m \left(\sqrt{|r_m|^2 \rho_t / \text{tr}\{\mathbf{R}\mathbf{R}^*\}}\right)$ .

*Proof:* See Appendix A. ■

From Proposition 1, it can be seen that either the RF gains  $\mathbf{G}_{\text{ZF}}$  and  $\mathbf{R}$  at the BS or the RF gains  $\mathbf{U}$  and  $\mathbf{B}$  at the UE side have impact on the downlink system performance. Especially, the difference of  $g_{\text{ZF},m}$  and  $\alpha r_m$  is a key factor for the existence of the fluctuant-channel interference and the multi-user interference. Further, the impact of mismatch at UE side is dependent on  $|u_k|^2 / (|b_k|^2 \text{tr}\{(\mathbf{B}\boldsymbol{\Phi}^2\mathbf{B}^*)^{-1}\})$ , which indicates that only the amplitude mismatch at the UE has the impact on the downlink performance.

**Remark 1** (Impact of the reciprocity mismatch). During the downlink transmission, the mismatch at the BS side is a key factor for the existence of the interference. Besides, only the amplitude of the mismatch at the UE side has the impact on the downlink performance.

Since the closed-form expression of the SINDR denoted in Proposition 1 is complicated, few insights can be directly obtained from it. In the following section, we will consider two special cases to gain deeper understanding of the impact of the nonlinear reciprocity mismatch.

### C. Impact Analysis of Nonlinear Reciprocity Mismatch

To further analyze the impact of the reciprocity mismatch at the BS side and the UE side, we derive the achievable rate at the high SINDR regime.

**Proposition 2** (Achievable rate with large SINDR). By assuming that the SINDR is large, based on the closed-form expression of SINDR denoted as (26), the achievable rate of the  $k$ -th UE of the multi-user massive MIMO system can be denoted as

$$R_k^{\text{LSNDR}} = R_{\text{Ideal}}^{\text{LSNR}} - \Delta R_{\text{BS},k} - \Delta R_{\text{UE},k}, \quad (31)$$

where  $R_{\text{Ideal}}^{\text{LSNR}}$  is given by

$$R_{\text{Ideal}}^{\text{LSNR}} = \log\left(\frac{M-K}{\text{tr}\{\boldsymbol{\Phi}^{-2}\}} \cdot \frac{\rho_t a_0}{\sigma_n^2}\right), \quad (32)$$

which represents the ideal achievable rate of each UE for the multi-user MIMO with ZF precoding at the high SNR regime.  $\Delta R_{\text{BS},k}$  denotes the average performance loss determined by the nonlinear reciprocity mismatch at the BS side and is given by

$$\Delta R_{\text{BS},k} = \log\left(\frac{\frac{M-K}{M} \rho_t a_0 \phi_k^2 \varepsilon_2 + \sigma_{\text{eq},k}^2}{\sigma_n^2 \varepsilon_3}\right) \quad (33)$$

with  $\varepsilon_2 = \text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})^*\} / M$ ,  $\varepsilon_3 = |\text{tr}\{\mathbf{G}_{\text{ZF}} \mathbf{R}^*\}|^2 / (M \text{tr}\{\mathbf{R}\mathbf{R}^*\})$ ,  $\sigma_{\text{eq},k}^2 = a_0 \phi_k^2 \text{tr}\{\boldsymbol{\Sigma}_{\text{ZF}}\} + \sigma_n^2 / |u_k|^2$ . Finally,  $\Delta R_{\text{UE},k}$  denotes the average performance

loss determined by the nonlinear reciprocity mismatch at the UE side denoted as

$$\Delta R_{\text{UE},k} = \log\left(\frac{|b_k|^2}{K} \text{tr}\{(\mathbf{B}\mathbf{B}^*)^{-1}\}\right). \quad (34)$$

*Proof:* According to (28) and (29), the sum power of the FCI and MUI can be given by

$$\begin{aligned} & \gamma_{\text{ZF},k}^{\text{FCI}} + \gamma_{\text{ZF},k}^{\text{MUI}} \\ &= \sum_{i=1}^K \frac{a_0 \rho_t \phi_k^2 \text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}}^* - \alpha^* \mathbf{R}^*)\}}{(M-K)^{-1} M^2 |u_k^{-1} b_i|^2 \phi_i^2 \text{tr}\{(\mathbf{B}\boldsymbol{\Phi}^2\mathbf{B}^*)^{-1}\}} \\ &= \frac{(M-K) a_0 \rho_t |u_k|^2 \phi_k^2 \varepsilon_2}{M \text{tr}\{(\mathbf{B}\boldsymbol{\Phi}^2\mathbf{B}^*)^{-1}\}} \sum_{i=1}^K |b_i \phi_i|^{-2} \\ &\stackrel{(a)}{=} \frac{M-K}{M} a_0 \rho_t |u_k|^2 \phi_k^2 \varepsilon_2, \end{aligned} \quad (35)$$

where  $\varepsilon_2 = \text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}}^* - \alpha^* \mathbf{R}^*)\} / M$ , and the condition (a) holds due to  $\text{tr}\{(\mathbf{B}\boldsymbol{\Phi}^2\mathbf{B}^*)^{-1}\} = \sum_{i=1}^K |b_i \phi_i|^{-2}$ .

By substituting (27), (30), and (35) into (26), the SINDR can be given by

$$\begin{aligned} \gamma_{\text{ZF},k} &= \frac{\frac{a_0 \rho_t (M-K) |u_k|^2 \varepsilon_3}{|b_k|^2 \text{tr}\{(\mathbf{B}\boldsymbol{\Phi}^2\mathbf{B}^*)^{-1}\}}}{\frac{M-K}{M} a_0 \rho_t \phi_k^2 |u_k|^2 \varepsilon_2 + a_0 |u_k|^2 \phi_k^2 \text{tr}\{\boldsymbol{\Sigma}_{\text{ZF}}\} + \sigma_n^2} \\ &\stackrel{(b)}{=} \frac{\frac{M-K}{\text{tr}\{\boldsymbol{\Phi}^{-2}\}} a_0 \rho_t \varepsilon_3 \frac{K}{|b_k|^2 \text{tr}\{(\mathbf{B}\mathbf{B}^*)^{-1}\}}}{\frac{M-K}{M} a_0 \rho_t \phi_k^2 \varepsilon_2 + a_0 \phi_k^2 \text{tr}\{\boldsymbol{\Sigma}_{\text{ZF}}\} + \frac{\sigma_n^2}{|u_k|^2}} \\ &= \frac{(M-K) a_0 \rho_t}{\text{tr}\{\boldsymbol{\Phi}^{-2}\} \sigma_n^2} \cdot \frac{\varepsilon_3 \sigma_n^2}{\frac{M-K}{M} a_0 \rho_t \phi_k^2 \varepsilon_2 + \sigma_{\text{eq},k}^2} \\ &\quad \cdot \frac{K}{|b_k|^2 \text{tr}\{(\mathbf{B}\mathbf{B}^*)^{-1}\}}, \end{aligned} \quad (36)$$

where  $\varepsilon_3 = |\text{tr}\{\mathbf{G}_{\text{ZF}} \mathbf{R}^*\}|^2 / (M \text{tr}\{\mathbf{R}\mathbf{R}^*\})$ ,  $\sigma_{\text{eq},k}^2 = a_0 \phi_k^2 \text{tr}\{\boldsymbol{\Sigma}_{\text{ZF}}\} + \sigma_n^2 / |u_k|^2$ , the condition (b) holds since  $\frac{1}{K} \text{tr}\{(\mathbf{B}\boldsymbol{\Phi}^2\mathbf{B}^*)^{-1}\} - \frac{1}{K} \text{tr}\{(\mathbf{B}\mathbf{B}^*)^{-1}\} \frac{1}{K} \text{tr}\{\boldsymbol{\Phi}^{-2}\} \rightarrow 0$  when  $K$  is large [36].

Since the SINDR is large, we can approximate  $\log(1 + \text{SINDR})$  with  $\log(\text{SINDR})$ . By substituting (36) into  $\log(\text{SINDR})$ , the achievable rate at high SINDR regime can be given by

$$\begin{aligned} R_k^{\text{LSNDR}} &= \log(\gamma_{\text{ZF},k}) \\ &= \underbrace{\log\left(\frac{(M-K)}{\text{tr}\{\boldsymbol{\Phi}^{-2}\}} \cdot \frac{a_0 \rho_t}{\sigma_n^2}\right)}_{R_{\text{Ideal}}^{\text{LSNR}}} - \underbrace{\log\left(\frac{|b_k|^2}{K} \text{tr}\{(\mathbf{B}\mathbf{B}^*)^{-1}\}\right)}_{\Delta R_{\text{UE},k}} \\ &\quad - \underbrace{\log\left(\frac{\frac{M-K}{M} a_0 \rho_t \phi_k^2 \varepsilon_2 + \sigma_{\text{eq},k}^2}{\varepsilon_3 \sigma_n^2}\right)}_{\Delta R_{\text{BS},k}}. \end{aligned} \quad (37)$$

According to (14),  $R_{\text{ideal}}^{\text{LSNR}}$  denotes the ideal achievable rate at the high SNR regime,  $\Delta R_{\text{BS},k}$  denotes the performance degradation caused by the reciprocity at the BS side, and  $\Delta R_{\text{UE},k}$  means the performance degradation caused by the reciprocity at the UE side. Hence, Proposition 2 holds. ■

From (34), the sum achievable rate degradation caused by the nonlinear reciprocity mismatch at the UE side can be further denoted as

$$\begin{aligned} \Delta R_{\text{UE}} &= \sum_{k=1}^K \Delta R_{\text{UE},k} = \sum_{k=1}^K \log \left( \frac{|b_k|^2}{K} \sum_{q=1}^K |b_q|^{-2} \right) \\ &= K \log \left( \frac{1}{K} \sum_{k=1}^K \frac{1}{|b_k|^2} \right) - \sum_{k=1}^K \log \left( \frac{1}{|b_k|^2} \right) \stackrel{(a)}{\geq} 0, \end{aligned} \quad (38)$$

where (a) is conditioned on the Jensen's inequality as  $\log(\frac{1}{K} \sum_{k=1}^K x_k) \geq \frac{1}{K} \sum_{k=1}^K \log(x_k)$ , and the equality holds if and only if  $x_1 = \dots = x_K$ .

**Remark 2** (Impact of the nonlinear reciprocity mismatch at the UE side). From (38),  $\Delta R_{\text{UE}}$  is always larger than 0, which reveals that the nonlinear reciprocity mismatch at the UE side can degrade the performance of the multi-user massive MIMO system. Further, the performance loss caused by the nonlinear mismatch at the UE side is not related to the downlink transmit power.

According to (33), the performance degradation caused by the nonlinear reciprocity mismatch at the BS side increases with the average transmit power  $\rho_t$ . Due to the complex expression of  $g_{\text{ZF},m}$  and  $\sigma_{\text{ZF},m}$ , it is difficult to thoroughly analyze the impact of the average power and the saturation level  $A_{\text{sat},m}$  from (33). In practice, the power of the input signal of the HPA is less than a maximum acceptable input power, and the IBO is larger than a threshold. Therefore, to further analyze the impact of the nonlinear amplification and the reciprocity mismatch at the BS side, we consider a special case where the IBO is large and the hardware of UEs is perfect. Based on the assumption, the SINDR can be further denoted as follows.

**Proposition 3** (SINDR with large IBO). Suppose that the hardware and HPAs of UEs are perfect, i.e.,  $u_k = v_k = 1$  and  $B_{\text{sat}} \gg |x_{u,k}|$ . When the saturation level of the HPA is larger than the average power of the input signal of the BS, the downlink SINDR of the multi-user massive MIMO system with ZF in the presence of the nonlinear reciprocity mismatch can be given by

$$\gamma_{\text{ZF},k}^{\text{LIBO}} = \frac{a_0 \varepsilon_4 \frac{(M-K)}{\text{tr}\{\Phi^{-2}\}} \left( 1 - \frac{2\rho_t e^{\delta_a^2}}{MA_{\text{sat}}^2} \right) \rho_t}{\frac{M-K}{M} a_0 \phi_k^2 \varepsilon_3 \left( 1 - \frac{2\rho_t e^{\delta_a^2}}{MA_{\text{sat}}^2} \right) \rho_t + \sigma_n^2}, \quad (39)$$

where  $\varepsilon_3 = e^{2\delta_t^2} + (e^{2\delta_r^2} - 2)e^{\delta_t^2 + \delta_r^2} \text{sinc}^2(\theta_t) \text{sinc}^2(\theta_r)$ , and  $\varepsilon_4 = \text{sinc}^2(\theta_t) \text{sinc}^2(\theta_r) e^{\delta_t^2 - \delta_r^2}$ .

*Proof:* Proof see Appendix B. ■

From (39), since the average transmit power  $\rho_t/M$  is less than  $A_{\text{sat}}^2$ , the SINDR always increases with  $\rho_t$ . In addition, the SINDR also increases with the saturation level  $A_{\text{sat}}$ .

When the sum power of the fluctuant-channel interference and multi-user interference dominates the denominator of SINDR, SINDR would approach an upper limit subject to the ratio of  $\varepsilon_3$  and  $\varepsilon_4$ . Moreover, increasing the number of the antennas  $M$  at the BS can also increase SINDR, which can be explained from two aspects. The antenna array gain of BS increases as more antennas are deployed at the BS. Increasing the number of antennas means that the average power allocated to each antenna decreases, which reduces the nonlinearity of the HPA.

**Remark 3** (Impact of the nonlinear reciprocity at the BS side). The reciprocity mismatch at the BS side generates the fluctuant-channel interference and multi-user interference, which dramatically degrades the system performance. Besides, the performance loss due to the mismatch at the BS side increases with the average transmit power. Further, the nonlinearity of the HPA can exacerbate the reciprocity mismatch and the performance loss.

From the above analytical results, the performance loss due to the mismatch at the BS side increases with the transmit power, while the performance loss due to the mismatch at the UE side remains constant. This shows that the impact of the nonlinear reciprocity mismatch at the BS side is much more severe than that at the UE side. Therefore, the reciprocity calibration at the BS side is essential for the multi-user massive MIMO system to improve the system performance.

#### IV. CALIBRATION OF NONLINEAR RECIPROACITY MISMATCH

Based on the analytical results in the above section, the nonlinear reciprocity mismatch causes dramatic performance degradation. In this section, we propose a nonlinear reciprocity calibration approach for the BS to mitigate the performance loss due to the nonlinear reciprocity mismatch.

##### A. Existing Work on Linear Reciprocity Calibration

The reciprocity calibration aims at making the ratio of downlink and uplink channel equal, i.e.,

$$c_m \frac{h_{\text{DL},m,k}}{h_{\text{UL},m,k}} = c_i \frac{h_{\text{DL},i,k}}{h_{\text{UL},i,k}}, \quad \forall m, i \in [1 : M], k \in [1 : K], \quad (40)$$

where  $c_m$  is the calibration coefficient of the  $m$ -th antenna at the BS.

In the massive MIMO system with the linear reciprocity mismatch,  $h_{\text{DL},m,k}/h_{\text{UL},m,k} = t_m/r_m$ . Hence, (40) can be further denoted as

$$c_m \frac{t_m}{r_m} = c_i \frac{t_i}{r_i}, \quad \forall m, i \in [1 : M]. \quad (41)$$

From (41), it can be seen that the reciprocity calibration requires the knowledge of the ratio of  $t_m$  and  $r_m$ . In the conventional reciprocity calibration,  $f_m = t_m/r_m$  ( $m \in [1 : M]$ ) can be estimated by using the OTA training [23], [25], [26]. The training pilot sequence of the  $m$ -th antenna at the BS can be denoted as  $\mathbf{x}_m = [x_m[1], \dots, x_m[Q]]^T$  with  $\mathbb{E}\{|x_m[q]|^2\} = \rho_c$ , where  $Q$  is the pilot length. Let  $\mathbf{y}_{m,i}$  denote the training signal transmitted by the  $m$ -th antenna



and received by the  $i$ -th antenna, and let  $\mathbf{y}_{i,m}$  represent the opposite direction signal. The received signals can be denoted as

$$\mathbf{y}_{m,i} = \sqrt{a_0} \omega_{m,i} t_m r_i \mathbf{x}_m + \mathbf{n}_{m,i}, \quad (42a)$$

$$\mathbf{y}_{i,m} = \sqrt{a_0} \omega_{m,i} t_i r_m \mathbf{x}_i + \mathbf{n}_{i,m}, \quad (42b)$$

where  $\omega_{m,i}$  denotes the wireless propagation channel between the  $m$ -th antenna and the  $i$ -th antenna,  $\mathbf{n}_{m,i}$  and  $\mathbf{n}_{i,m}$  are AWGN vector. According to [25], the reciprocity mismatch factors can be computed by

$$\hat{\mathbf{f}} = [1, -\bar{\mathbf{y}}_1^T \bar{\mathbf{Y}}_2^* (\bar{\mathbf{Y}}_2^T \bar{\mathbf{Y}}_2^*)^{-1}], \quad (43)$$

where  $\bar{\mathbf{y}}_1$  is the first column of the matrix  $\bar{\mathbf{Y}}$ ,  $\bar{\mathbf{Y}}_2$  consist of the second column to the last column of  $\bar{\mathbf{Y}}$ , and  $\bar{\mathbf{Y}} \in \mathbb{C}^{M \times M}$  is defined as

$$[\bar{\mathbf{Y}}]_{m,i} = \begin{cases} \sum_{j \neq m}^M |\mathbf{x}_m^T \mathbf{y}_{j,m}|^2, & \text{when } m = i, \\ -\mathbf{y}_{i,m}^H \mathbf{x}_m^* \mathbf{x}_i^T \mathbf{y}_{m,i}, & \text{when } m \neq i. \end{cases} \quad (44)$$

Then, the linear calibration coefficients can be legitimately computed by

$$\hat{\mathbf{c}} = \left[ \frac{c_0}{\hat{f}_1}, \dots, \frac{c_0}{\hat{f}_M} \right]^T, \quad (45)$$

where  $c_0$  can be any non-zero constant. Finally, during the data transmission, the precoding matrix for implementing the linear reciprocity calibration can be given by  $\mathbf{W}_c = \frac{1}{\sqrt{\beta_c}} \text{diag}(\mathbf{c}) \mathbf{W}$  where  $\beta_c = \text{tr} \{ \text{diag}(\mathbf{c}) \mathbf{W} \mathbf{W}^H \text{diag}(\mathbf{c}^*) \}$ .

**Remark 4** (Overhead and computational complexity of conventional calibration). During the conventional calibration, the antennas of BS broadcast  $Q$ -length training pilots in turn. Therefore, the training overhead of conventional calibration is  $MQ$ . For computing the reciprocity mismatch factors, the asymptotic computational complexity can be given by  $\mathcal{O}(M^3 + QM^2)$ . The computational complexity of computing the calibration coefficients can be given by  $\mathcal{O}(M)$ .

In the massive MIMO system with the nonlinear reciprocity mismatch,  $h_{\text{DL},m,k}/h_{\text{UL},m,k} = \frac{t_m}{r_m} \mu(A_{\text{sat},m}/\sigma_{x,m}) = \mu_m(\sigma_{x,m}) = \bar{g}_m$ , and (40) can be rewritten as

$$c_m \mu_m(|c_m| \sigma_{x,m}) = c_i \mu_i(|c_i| \sigma_{x,i}), \quad \forall m, i \in [1 : M]. \quad (46)$$

From the equation, the nonlinear reciprocity calibration requires the knowledge of the function  $\bar{g}_m = \mu_m(\sigma_{x,m})$ . Since the reciprocity mismatch coefficients  $\bar{g}_m$  of the  $m$ -th antenna is a nonlinear function of the average transmit power  $\sigma_{x,m}$ , the calibration coefficients are also related to the power. Therefore, compared with linear reciprocity calibration, the nonlinear reciprocity calibration encounters greater challenges as follows:

- **Challenge 1:** Since the reciprocity mismatch factor varies with the transmit power, it is essential to obtain the relationship  $\bar{g}_m = \mu_m(\sigma_{x,m})$  between the nonlinear reciprocity mismatch factor  $\bar{g}_m$  and the transmit power  $\sigma_{x,m}$ .
- **Challenge 2:** From (16), the function of  $\bar{g}_m$  and  $\sigma_{x,m}$  is complex. Hence, it is difficult to determine the expression of the mismatch function  $\bar{g}_m = \mu_m(\sigma_{x,m})$ .

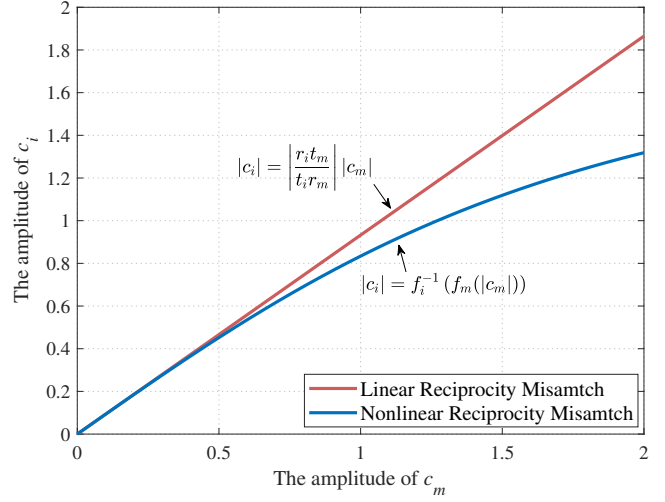


Fig. 2. The relationship of the calibration coefficients  $c_m$  and  $c_i$  ( $i \neq m$ )  $\forall m, i \in [1 : M]$ . The function  $\varphi_i^{-1}(x)$  denotes the inverse function of  $x = \varphi_i(|c_i|) = |c_i| \mu_i(|c_i| \sigma_{x,m})$ , and  $\varphi_m(|c_m|) = |c_m| \mu_m(|c_m| \sigma_{x,m})$ .

- **Challenge 3:** As illustrated in Fig. 2, the relationships between the calibration coefficients are highly nonlinear so that it is difficult to solve the nonlinear calibration coefficients.

To overcome these challenges, we propose three techniques which are the multi-power training pilots, the OTA-polynomial training approach, and the toward optimal nonlinear calibration coefficients for the nonlinear reciprocity calibration.

### B. Pilots Design for Nonlinear Reciprocity Calibration

According to challenge 1, the reciprocity mismatch is a nonlinear function of the transmit power. To sample the nonlinear function versus the transmit power, we propose multi-power training pilots. Suppose the maximum transmit power of the  $m$ -th antenna is  $\sigma_{\text{max},m}^2$ . The power sequences of the calibration signals can be denoted as  $\{\rho_{c,1}, \dots, \rho_{c,N}\}$ , where  $\rho_{c,n} = (n\sigma_{\text{max},m}/N)^2$  ( $n \in [1 : N]$ ). The calibration signal sequence of  $m$ -th antenna at the  $n$ -th power point can be denoted as  $\{x_m^{(n)}[q]\}_{q \in [1:Q]}$ , where  $\mathbb{E} \{ |x_m^{(n)}[q]|^2 \} = \rho_{c,n}$ ,  $\forall q \in [1 : Q]$ .

### C. Nonlinear Reciprocity Mismatch Factors Estimation

According to Challenge 2, it is difficult to determine  $\mu_m(\sigma_{x,m})$  in practice. Inspired by applying polynomials to characterize the HPA, we employ the polynomials to approximate the function  $g_m = t_m \mu(A_{\text{sat},m}/\sigma_{x,m})$  during the reciprocity calibration. Based on [32], the function  $\mu_m(\sigma_{x,m}) = g_m/r_m$  can be expressed by a  $\Pi$ -order<sup>2</sup> polynomial function as

$$\mu_m(\sigma_{x,m}) = \frac{g_m}{r_m} = \sum_{\varpi=0}^{\Pi} \tau_{m,\varpi} \psi_{\varpi}(\sigma_{x,m}), \quad (47)$$

<sup>2</sup>The optimal polynomial order  $\Pi$  depends on the transmit power  $\rho_t$ , the equivalent noise power  $\bar{\sigma}_z^2$ , and the property of the nonlinear function. According to [37], the optimal order can be chose by minimizing the squared error, i.e.,  $\Pi_0 = \arg \min_{\Pi} \sum_{n=1}^N \sum_{q=1}^Q \sum_{m,i}^M (y_{m,i}^{(n)}[q] - y_{\Pi,m,i}^{(n)}[q])^2$ , where  $y_{\Pi,m,i}^{(n)}[q]$  denotes the results via polynomial fitting.

where  $\tau_{m,\varpi}$  is the polynomial coefficient, and  $\psi_\varpi(\sigma_{x,m})$  is the orthogonal polynomial given by

$$\psi_\varpi(\sigma_{x,m}) = \sum_{l=0}^{\varpi} (-1)^{l+\varpi+2} \frac{(\varpi+l+2)!}{l!(l+1)!(\varpi-l)!} \sigma_{x,m}^l. \quad (48)$$

Then, we propose an OTA training approach based the multi-power pilots to estimate the polynomial coefficient  $\tau_{m,\varpi}$ .

Let  $y_{m,i}^{(n)}[q]$  denote the signal received by the  $i$ -th antenna.  $y_{i,m}^{(n)}[q]$  denotes the received signal by the  $m$ -th antenna. They are can be given by

$$y_{m,i}^{(n)}[q] = \sqrt{a_0} r_i \omega_{m,i} g_m x_m^{(n)}[q] + \tilde{z}_{m,i}^{(n)}[q], \quad (49a)$$

$$y_{i,m}^{(n)}[q] = \sqrt{a_0} r_m \omega_{i,m} g_i x_i^{(n)}[q] + \tilde{z}_{i,m}^{(n)}[q], \quad (49b)$$

where  $\tilde{z}_{m,i}^{(n)}[q]$  denotes the equivalent noise consisting of the nonlinear distortion and the AWGN at the  $i$ -th antenna, and  $\tilde{z}_{i,m}^{(n)}[q]$  is the equivalent noise at the  $m$ -th antenna. Then, the polynomial coefficients vector  $\boldsymbol{\tau} = [\tau_{1,0}, \tau_{1,1}, \dots, \tau_{M,\Pi}]^T$  can be computed by the LS approach as follows.

**Proposition 4** (Estimation of the polynomial coefficients). Assuming that the first polynomial coefficient of the first antenna is a constant, e.g.,  $\tau_{1,0} = 1$ , the polynomial coefficients can be computed by

$$\hat{\boldsymbol{\tau}} = \left[ 1, -\bar{\boldsymbol{\psi}}_1^T \boldsymbol{\Psi}_2^* (\boldsymbol{\Psi}_2^T \boldsymbol{\Psi}_2^*)^{-1} \right]^T, \quad (50)$$

where  $\bar{\boldsymbol{\psi}}_1$  is the first column of  $\boldsymbol{\Psi}$ ,  $\boldsymbol{\Psi}_2$  consists of the second to the  $M\Pi$ -th column of  $\boldsymbol{\Psi}$ , and  $\boldsymbol{\Psi}$  consists of the received training signals and denoted as (54).

*Proof:* Based on (50), since the wireless propagation channel is reciprocal, i.e.,  $h_{m,i} = h_{i,m}$ , the equation can be denoted as  $\frac{g_m}{r_m} x_m^{(n)}[q] y_{i,m}^{(n)}[q] = \frac{g_i}{r_i} x_i^{(n)}[q] y_{m,i}^{(n)}[q]$  by ignoring the noise. By substituting (47) into the equation, it can be further denoted as

$$\bar{y}_{m,i}^{(n)}[q] \sum_{\varpi=0}^{\Pi} \tau_{i,\varpi} \psi_\varpi(\rho_{c,n}) - \bar{y}_{i,m}^{(n)}[q] \sum_{\varpi=0}^{\Pi} \tau_{m,\varpi} \psi_\varpi(\rho_{c,n}) = 0, \quad (51)$$

where  $\bar{y}_{m,i}^{(n)}[q] = y_{m,i}^{(n)}[q] x_i^{(n)}[q]$  and  $\bar{y}_{i,m}^{(n)}[q] = y_{i,m}^{(n)}[q] x_m^{(n)}[q]$ . For all  $n \in [1 : N]$  and  $q \in [1 : Q]$ , (51) holds. Thus, by gathering all the signals and ignoring the noise, the equation can be further denoted as

$$\boldsymbol{\Psi}_{m,i} \boldsymbol{\tau}_i - \boldsymbol{\Psi}_{i,m} \boldsymbol{\tau}_m = 0, \quad (52)$$

where  $\boldsymbol{\tau}_i = [\tau_{i,1}, \dots, \tau_{i,\Pi}]^T$ ,  $\boldsymbol{\tau}_m = [\tau_{m,1}, \dots, \tau_{m,\Pi}]^T$ ,  $\boldsymbol{\Psi}_{m,i}^{(1)} = [\boldsymbol{\Psi}_{m,i}^{(1)}, \dots, \boldsymbol{\Psi}_{m,i}^{(N)}]^T$ ,  $\boldsymbol{\Psi}_{i,m}^{(1)} = [\boldsymbol{\Psi}_{i,m}^{(1)}, \dots, \boldsymbol{\Psi}_{i,m}^{(N)}]^T$ ,  $\boldsymbol{\Psi}_{m,i}^{(n)} = [\bar{y}_{m,i}^{(n)}[1] \boldsymbol{\psi}_n, \dots, \bar{y}_{m,i}^{(n)}[Q] \boldsymbol{\psi}_n]$ ,  $\boldsymbol{\Psi}_{i,m}^{(n)} = [\bar{y}_{i,m}^{(n)}[1] \boldsymbol{\psi}_n, \dots, \bar{y}_{i,m}^{(n)}[Q] \boldsymbol{\psi}_n]$ , and  $\boldsymbol{\psi}_n = [\psi_0(\rho_{c,n}), \dots, \psi_\Pi(\rho_{c,n})]^T$ . By stacking the equation of all pairs of antennas into the matrix form, the overall equation can be denoted as

$$\boldsymbol{\Psi} \boldsymbol{\tau} = 0, \quad (53)$$

where  $\boldsymbol{\tau} = [\boldsymbol{\tau}_1^T, \dots, \boldsymbol{\tau}_M^T]^T$ , and  $\boldsymbol{\Psi}$  is defined as

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{1,2} & -\boldsymbol{\Psi}_{2,1} & 0 & \cdots \\ \boldsymbol{\Psi}_{1,3} & 0 & -\boldsymbol{\Psi}_{3,1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & \boldsymbol{\Psi}_{2,3} & -\boldsymbol{\Psi}_{3,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (54)$$

To exclude the trivial all-zero solution to (53),  $\tau_{1,0}$  is assumed to be known previously and set to 1. Then, (53) can be further written as

$$\boldsymbol{\Psi}_2 \boldsymbol{\tau}_c + \bar{\boldsymbol{\psi}}_1 = 0, \quad (55)$$

where  $\boldsymbol{\tau}_c$  consists of the second to the  $\Pi$ -th row of  $\boldsymbol{\tau}$ ,  $\bar{\boldsymbol{\psi}}_1$  is the first column of  $\boldsymbol{\Psi}$  and  $\boldsymbol{\Psi}_2$  consists of the second to the  $M\Pi$ -th column of  $\boldsymbol{\Psi}$ . (55) can be solved by the least square algorithm and its solution can be given by

$$\hat{\boldsymbol{\tau}}_c = -(\boldsymbol{\Psi}_2^H \boldsymbol{\Psi}_2)^{-1} \boldsymbol{\Psi}_2^H \bar{\boldsymbol{\psi}}_1. \quad (56)$$

As a result, the coefficient vector  $\boldsymbol{\tau} = [\tau_1, \hat{\boldsymbol{\tau}}_c^T]^T$  can be denoted as (50). ■

#### D. Computing the Reciprocity Calibration Coefficients

In terms of (46), only the amplitude of the calibration coefficients changes the output of HPA. Hence, the amplitude and phase of the calibration coefficient can be computed independently. By substituting (47) into (46), the amplitude and phase should, respectively, satisfy

$$|c_m| \bar{\mu}_m(|c_m|) = |c_m| \bar{\mu}_i(|c_i|), \quad (57)$$

$$\angle c_m + \angle \mu_m(|c_m|) = \angle c_i + \angle \mu_i(|c_i|), \quad (58)$$

where  $\bar{\mu}_m(|c_m|)$  is the amplitude of  $\mu_m(|c_m| \sigma_{x,m})$  and can be denoted as

$$\bar{\mu}_m(|c_m|) = \sqrt{\mu_{r,m}^2(|c_m| \sigma_{x,m}) + \mu_{i,m}^2(|c_m| \sigma_{x,m})}. \quad (59)$$

In (59),  $\mu_{r,m}(\sigma_{x,m}) = \sum_{\varpi=0}^{\Pi} \hat{\tau}_{r,m,\varpi} \psi_\varpi(\sigma_{x,m})$ ,  $\mu_{i,m}(\sigma_{x,m}) = \sum_{\varpi=0}^{\Pi} \hat{\tau}_{i,m,\varpi} \psi_\varpi(\sigma_{x,m})$ , where  $\hat{\tau}_{r,m,\varpi}$  is the real part of  $\hat{\tau}_{m,\varpi}$ , and  $\hat{\tau}_{i,m,\varpi}$  denotes the imaginary part of  $\hat{\tau}_{m,\varpi}$ . Therefore, the amplitude and phase of calibration coefficients can be computed by solving (57) and (58), respectively.

As mentioned by Challenge 3, it is difficult to solve the calibration coefficients from (57). We can find that there are infinitely many solutions to (57), even including an all-zero solution which can make the system fail. Due to the complex expression of the function  $\mu_m(x)$ , it is difficult to find the calibration coefficients exactly satisfying the total power constraint and the maximum power constraint based on (57). To address the issues, an optimization problem is formulated to solve the equations efficiently and find the optimal calibration coefficients. The problem seeks to maximize the downlink achievable rate with maximum transmit power constraints, i.e.,  $\sum_{m=1}^M |c_m|^2 \sigma_{x,m}^2 \leq \rho_t$  and  $|c_m| \sigma_{x,m} \leq \sigma_{\max,m}$ , and

reciprocity constraints in (57). Thus, the problem can be formulated as

$$\begin{aligned} \mathcal{P}_1 : & \max_{|c_m|} \gamma_k \\ \text{s.t. } \mathcal{C}_1 : & \sum_{m=1}^M |c_m|^2 \sigma_{x,m}^2 \leq \rho_t, \\ \mathcal{C}_2 : & \varphi_m(|c_m|) = \varphi_i(|c_i|), \\ \mathcal{C}_3 : & 0 \leq |c_m| \leq c_{\max,m}, \end{aligned} \quad (60)$$

where  $c_{\max,m} = \sigma_{\max,m}/\sigma_{x,m}$ , and  $\varphi_m(|c_m|) = |c_m| \bar{\varphi}_m(|c_m|)$ .

Since both the objective and  $\mathcal{C}_2$  are non-convex,  $\mathcal{P}_1$  is non-convex and difficult to solve. To solve  $\mathcal{P}_1$  efficiently, we reformulate  $\mathcal{P}_1$  as an equivalent convex problem as follows.

**Proposition 5** (Equivalent convex optimization). By substituting (41) into the objective and relaxing the equation constraint,  $\mathcal{P}_1$  can be transformed into an equivalent convex problem as

$$\begin{aligned} \mathcal{P}_2 : & \max_{|c_m|, g_0} g_0 \\ \text{s.t. } \mathcal{C}_1, \mathcal{C}_3, \\ \mathcal{C}_4 : & 0 \leq g_0 \leq \varphi_m(|c_m|), \quad \forall m \in [1 : M]. \end{aligned} \quad (61)$$

*Proof:* From the theoretical analysis, the distortion is negligible in practical systems and can be regarded as the equivalent noise. By substituting (41) into the SINDR in (26) and focusing on the nonlinear reciprocity mismatch at the BS side, the SINDR can be further denoted as  $\gamma_k = (M - K)\rho_t a_0 g_0^2 |\text{tr}\{\mathbf{R}^*\}|^2 / (\text{tr}\{\mathbf{\Phi}^{-2}\} M \text{tr}\{\mathbf{R}\mathbf{R}^*\} \bar{\sigma}_n^2)$ , where  $g_0 = \varphi_m(|c_m|)$ . Since the quadratic function is monotonically increasing when  $g_0 \geq 0$ , the objective of  $\mathcal{P}_1$  can be replaced by  $g_0$  and the constraint  $\mathcal{C}_2$  can be rewritten as  $g_0 = \varphi_m(|c_m|)$ ,  $\forall m \in [1 : M]$ .

Since  $\varphi_m(|c_m|)$  is a nonlinear function of  $|c_m|$ , the equation constraint  $\mathcal{C}_2$  is nonconvex. Thanks to the concavity of  $\varphi_m(|c_m|)$ , the constraint  $\mathcal{C}_2$  becomes convex by relaxing the equation, i.e.,  $g_0 \leq \varphi_m(|c_m|)$ . Hence,  $\mathcal{P}_1$  can be reformulated as  $\mathcal{P}_2$  as (61). Because both the objective function and the function  $g_0 - \varphi_m(|c_m|)$  are monotonically increasing with  $g_0$ , the problem  $\mathcal{P}_2$  after relaxation is equivalent to the primal problem  $\mathcal{P}_1$  [38]. ■

As  $\mathcal{P}_2$  is a convex optimization problem, it has a unique maximum. The optimal solution can be obtained by some math tools, e.g., the interior-point method [38], but the performance of such method decreases rapidly as the antenna number  $M$  increases. Inspired by the sequential linear programming (SLP) [39], we propose an efficient algorithm to solve  $\mathcal{P}_2$ .

By employing the Taylor expansion centered at  $|c_m^{l-1}|$ , the nonlinear function  $\varphi_m(|c_m|)$  can be approximated as

$$\begin{aligned} \varphi_m(|c_m|) & \approx \tilde{\varphi}_m(|c_m^{l-1}|, |c_m|) \\ & = \varphi_m(|c_m^{l-1}|) + \varphi'_m(|c_m^{l-1}|)(|c_m| - |c_m^{l-1}|), \end{aligned} \quad (62)$$

where  $\{|c_m^{l-1}|\}_{m \in [1:M]}$  is the solution to the  $(l-1)$ -th iteration optimization. Then, the subproblem of the  $l$ -th iteration can be

**Algorithm 1** SLP for computing the calibration coefficients.

**Require:**  $|c_m^0| = 0 \quad \forall m \in [1 : M]$ ,  $\epsilon$ .

**Ensure:** The nonlinear reciprocity calibration coefficients  $\{\bar{c}_m\}_{m \in [1:M]}$ .

- 1: **repeat**
- 2: Compute  $\varphi_m(|c_m^{l-1}|)$ ,  $\varphi'_m(|c_m^{l-1}|)$ , and  $\chi_m^{l-1}$ ,  $\forall m \in [1 : M]$ .
- 3: Solve the equation  $\sum_{m=1}^M \left[ \frac{(g_0 - \chi_m^{l-1})\sigma_{x,m}}{\varphi'_m(|c_m^{l-1}|)} \right]^2 = \rho_t$  by the quadratic formula.
- 4: Set  $g_0^l = \min \{ \tilde{\varphi}_m(|c_m^{l-1}|, c_{\max,m}), \hat{g}_0 \}$ .
- 5: Set  $\bar{c}_m^l = \frac{g_0^l - \chi_m^{l-1}}{\varphi'_m(|c_m^{l-1}|)}$ ,  $\forall m \in [1 : M]$ .
- 6: Step size rule: Set  $\Delta c_m^l = \bar{c}_m^l - |c_m^{l-1}|$  and choose a  $\varrho \in (0, 1)$ . Let  $\varsigma^{(l)}$  be the largest element in  $\{\varrho^j\}_{j=0,1,\dots}$  satisfying  $\min_m \{ \varphi_m(|c_m^{l-1}|) \} \leq \min_m \{ \varphi_m(|c_m^{l-1}| + \varsigma^{(l)} \Delta c_m^l) \}$  and the inequalities denoted in (65).
- 7: Set  $|c_m^l| = |c_m^{l-1}| + \varsigma^{(l)} \Delta c_m^l$ ,  $\forall m \in [1 : M]$ .
- 8: **until**  $\frac{1}{M} \sum_{m=1}^M ||c_m^l| - |c_m^{l-1}|| < \epsilon$ .

denoted as

$$\begin{aligned} \mathcal{P}_3 : & \max_{f_m, g_0} g_0 \\ \text{s.t. } \mathcal{C}_1, \mathcal{C}_3, \\ \mathcal{C}_5 : & 0 \leq g_0 \leq \tilde{\varphi}_m(|c_m^{l-1}|, |c_m|), \quad \forall m \in [1 : M]. \end{aligned} \quad (63)$$

$\mathcal{P}_3$  is a linear programming and the solution can be given as follows.

**Proposition 6** (Solution to the  $l$ -th subproblem). The problem  $\mathcal{P}_3$  is convex and has a unique solution as

$$\begin{cases} \bar{c}_m^l = \frac{g_0^l - \chi_m^{l-1}}{\varphi'_m(|c_m^{l-1}|)}, \\ g_0^l = \min \{ \tilde{\varphi}_m(|c_m^{l-1}|, c_{\max,m}), \hat{g}_0 \}, \end{cases} \quad (64)$$

where  $\hat{g}_0$  is the solution to the quadratic equation  $\sum_{m=1}^M \left[ \frac{(g_0 - \chi_m^{l-1})\sigma_{x,m}}{\varphi'_m(|c_m^{l-1}|)} \right]^2 = \rho_t$  and can be solved by the quadratic formula, and  $\chi_m^{l-1} = \varphi_m(|c_m^{l-1}|) - \tilde{\varphi}_m(|c_m^{l-1}|)|c_m^{l-1}|$ .

*Proof:* The solution to  $\mathcal{P}_3$  can be solved by using the Lagrange method and Karush-Kuhn-Tucker conditions [38]. ■

In (64), to ensure that  $\bar{c}_m^l$  is non-negative,  $g_0^l$  should be larger than  $\max_m \{ \chi_m^{l-1} \}$ , i.e.,

$$\begin{aligned} \min_m \{ \tilde{\varphi}_m(|c_m^{l-1}|, c_{\max,m}) \} & > \max_m \{ \chi_m^{l-1} \}, \\ \sum_{m=1}^M \left[ \frac{\max_m \{ \chi_m^{l-1} \} - \chi_m^{l-1}}{\varphi'_m(|c_m^{l-1}|)\sigma_{x,m}} \right]^2 & < \rho_t. \end{aligned} \quad (65)$$

Therefore, we propose a simple line search approach to choose a step size  $\varsigma \in (0, 1]$  ensuring that the inequalities in (65) are true and  $\min_m \{ \varphi_m(|c_m^l|) \}$  is non-decreasing. Assume that the  $\{|c_m^{l-1}|\}_{m \in [1:M]}$  in the  $(l-1)$ -th iteration can guarantee the inequalities in (65). The step size  $\alpha$  can be the largest element in  $\{\varrho^j\}_{j=0,1,\dots}$  satisfying  $\min_m \{ \varphi_m(|c_m^{l-1}|) \} \leq \min_m \{ \varphi_m(|c_m^l|) \}$  and the inequalities in (65), where  $\varrho \in (0, 1)$ , and  $|c_m^l| = (1 - \varsigma)|c_m^{l-1}| + \varsigma \bar{c}_m^l$ . Finally, the SLP

---

**Algorithm 2** Nonlinear reciprocity calibration
 

---

- **Step 1 (Transmit training pilots):** The antenna  $m$  transmits the training pilots  $\{x_m^{(n)}[1], \dots, x_m^{(n)}[Q]\}$ , and antenna  $i$  ( $i \neq m$ ) receives the training signals  $y_{m,i}^{(n)}[q]$ ,  $\forall m, i \in [1 : M]$ ,  $n \in [1 : N]$ ,  $q \in [1 : Q]$ .
  - **Step 2 (Estimate the polynomial coefficients):** The baseband processor gathers all the received training signals and formulates the matrix  $\Psi$  denoted in (54). Then, the polynomial coefficients can be estimated by (50).
  - **Step 3 (Compute the nonlinear calibration coefficients):** The baseband processor formulates the amplitude function  $\bar{\mu}(\sigma_{x,m})$  denoted in (59). Then, the amplitude of the calibration coefficients can be computed by solving the optimization problem  $\mathcal{P}_2$  with Algorithm 1. The phase of the calibration coefficients can be computed by (66).
- 

approach for computing the nonlinear calibration coefficients is summarized as Algorithm 1.

**Remark 5** (Convergence analysis). Since  $\mathcal{P}_2$  equivalently maximizes  $\min_m \{\varphi_m(|c_m|)\}$  which is non-decreasing during the iterations and is a bounded function,  $(\bar{c}_1, \dots, \bar{c}_M)$  is a limit point of the iterations in Algorithm 1, which yields  $\lim_{l \rightarrow \infty} \min_m \{\varphi_m(|c_m^l|)\} = \min_m \{\varphi_m(\bar{c}_m)\}$ . According to (62), the limit point  $(\bar{c}_1, \dots, \bar{c}_M)$  is likewise the stationary point of  $\mathcal{P}_2$ .

Further, the phases of calibration coefficients can be computed by solving the equations denoted in (58). Since the variables are more than the equations, the equations are underdetermined and have infinitely many solutions. Since the phases of calibration coefficients unrelated to the average transmit power, any particular solution to the equations can make the system work well. Therefore, we can give a particular solution denoted as

$$\begin{aligned} \angle c_m &= -\angle \mu_n(|c_m| \sigma_{x,m}) \\ &= -\arctan \left( \frac{\mu_{r,m}(|c_m| \sigma_{x,m})}{\mu_{i,m}(|c_m| \sigma_{x,m})} \right), \quad \forall m \in [1 : M]. \end{aligned} \quad (66)$$

Thus, the precoding matrix of ZF for implementing the nonlinear reciprocity calibration can be denoted as  $\mathbf{W}_c = \text{diag}(\mathbf{c})\mathbf{W}$ .

### E. Calibration Process, Overhead and Complexity Analysis

The overall process of the nonlinear reciprocity calibration can be described as Algorithm 2. According to Algorithm 2, it can be seen that the overhead is caused by transmitting the training signals, and the computational complexity mainly results from computing the polynomial coefficients and the nonlinear calibration coefficients, respectively. The training overhead is caused by learning the system coefficients and can be defined as the number of transmitting the training pilots. As for the computational complexity, we focus on the complexity resulting from the multiplication. Then, the training overhead and the computational complexity for the polynomial nonlinear reciprocity calibration are given as follows.

**Remark 6** (Overhead and computational complexity). For the polynomial nonlinear reciprocity calibration, the total training overhead is  $MNQ$ . The computational complexity of computing the polynomial coefficients can be given by  $\mathcal{O}(M^3\Pi^3 + M^2Q^2N^2\Pi)$  and the computational complexity of computing calibration coefficients by Algorithm 1 is  $\mathcal{O}(M\Pi)$ .

### F. Non-coherent OTA-polynomial Training

In Section IV-C, the channel remains constant during the whole training process, which may be invalid if the number of antennas at the BS is very large. To cope with this issue, we extend the nonlinear mismatch factors estimation to the non-coherent nonlinear mismatch factors estimation [26], i.e., the non-coherent OTA-polynomial training. In the non-coherent training, only a part of antennas participate in the training process during each coherent slot.

Let  $\mathcal{M} = \{1, 2, \dots, M\}$  denote the index set of all the antennas. Suppose that the number of coherent slots used for training is  $\mathcal{T}_c$ . In each coherent slot, only a subset of antennas participate in the current training process. Hence, the antennas are divided into  $\mathcal{T}_c$  groups  $\mathcal{M}_1, \dots, \mathcal{M}_{\mathcal{T}_c}$  based on two rules. First, each antenna should participate in the training at least once, i.e.,  $\cup_{l_c=1}^{\mathcal{T}_c} \mathcal{M}_{l_c} = \mathcal{M}$ . Second, the calibration in the different groups should be correlated, e.g., the adjacent two groups contains at least one same antenna, i.e.,  $\mathcal{M}_{l_c} \cap \mathcal{M}_{l_c+1} \neq \emptyset$  ( $l_c \in [1 : \mathcal{T}_c - 1]$ ).

During the  $l_c$ -th coherent slot, the antennas in the subset  $\mathcal{M}_{l_c}$  ( $\mathcal{M}_{l_c} \subseteq \mathcal{M}$ ) broadcast training signal in turn. The training signals between antenna  $m$  ( $m \in \mathcal{M}_{l_c}$ ) and antenna  $i$  ( $i \in \mathcal{M}_{l_c}$ ) can be denoted as

$$y_{m,i,l_c}^{(n)}[q] = \sqrt{a_0} r_i \omega_{m,i}^{l_c} g_m x_{m,l_c}^{(n)}[q] + \tilde{z}_{m,i,l_c}^{(n)}[q], \quad (67a)$$

$$y_{i,m,l_c}^{(n)}[q] = \sqrt{a_0} r_m \omega_{i,m}^{l_c} g_m x_{i,l_c}^{(n)}[q] + \tilde{z}_{i,m,l_c}^{(n)}[q], \quad (67b)$$

where  $\omega_{m,i}^{l_c}$  and  $\omega_{i,m}^{l_c}$  denotes the wireless channel between the  $m$ -th antenna and the  $i$ -th antenna in the subset  $\mathcal{M}_{l_c}$  during the coherent slot  $l_c$ . Same as (51)–(53), by stacking these equations into matrix form with  $i, j \in \mathcal{M}_{l_c}$ , we can obtain  $\Psi_{l_c} \boldsymbol{\tau} = \tilde{\mathbf{N}}_{l_c}$ , where  $\Psi_{l_c}$  stacks the received training signal at the  $l_c$  coherent slot and has the same structure with  $\Psi$  denoted as (54). Same as (50), the polynomial coefficients can be computed by

$$\hat{\boldsymbol{\tau}} = [1, -\bar{\boldsymbol{\phi}}_{nc,1}^T \Psi_{nc,2}^* (\Psi_{nc,2}^T \Psi_{nc,2}^*)^{-1}]^T, \quad (68)$$

where  $\bar{\boldsymbol{\psi}}_{nc,1}$  is the first column of  $\Psi_{nc}$ ,  $\Psi_{nc,2}$  consists of the second to the  $M\Pi$ -th columns of  $\Psi_{nc}$ ,  $\Psi_{nc} = [\Psi_1^T, \dots, \Psi_{\mathcal{T}_c}^T]^T$ .

**Remark 7** (Overhead of non-coherent training). To accumulate same  $M(M-1)QN$  training signals as the coherent training, the non-coherent training overhead in each coherent slot can approximated as  $MQN/\sqrt{\mathcal{T}_c}$ , and the whole training overhead of  $\mathcal{T}_c$  slots is  $\sqrt{\mathcal{T}_c}MQN$ . Therefore, the non-coherent training relaxes the requirement of the coherent time at the cost of increasing the number of coherent slots and the overall training overhead. Further, since the the overall calibration overhead of the coherent training process is smaller, the coherent training is worth being adopted unless the coherence time cannot support it.



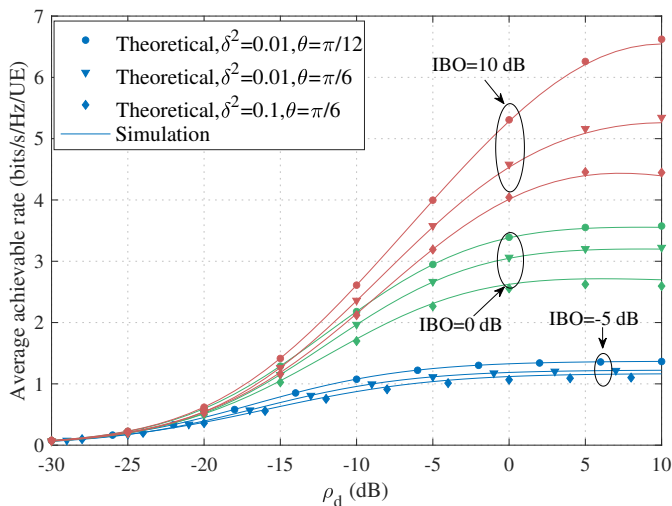


Fig. 3. Average achievable rate versus the downlink transmit SNR  $\rho_d$ .

## V. SIMULATION RESULTS AND DISCUSSIONS

In this section, we provide simulation results for the multi-user massive MIMO system in the presence of the nonlinear reciprocity mismatch to verify the analytical results and to show the performance of the proposed calibration approach. For easy presentation, RC denotes reciprocity calibration, and NRC represents nonlinear reciprocity calibration.

The system parameters for simulations are set as follows. The cell radius is normalized to 1 and the minimum distance between BS and UE is set to 0.01 [12]. The BS is equipped with  $M = 256$  antennas and is deployed at the center of the cell. There are  $K = 20$  single-antenna UEs served by the BS simultaneously. The large-scale path loss between the  $k$ -th UE and BS is modeled as  $\phi_k = \sqrt{\zeta d_{h,k}^{-\xi}}$ , where  $\zeta$  is the path gain at the reference distance of the far-field area,  $d_{h,k}$  is the distance between the BS and UE  $k$ , and  $\xi$  is the path loss exponent [27]. In the simulation,  $\zeta$  is set to  $-20$  dB and  $\xi$  equals to 3.7. We use  $\rho_d$  to denote the downlink transmit SNR, i.e.  $\rho_d = \rho_t/\sigma_n^2$ , and the variance  $\sigma_n^2$  of AWGN is set to 1 during the simulation. The small amplification gain  $a_0$  is equal to 20 dB. Further, both the IBO and the mismatch coefficients are differently set in each simulation.

### A. Impact of Nonlinear Reciprocity Mismatch

Fig. 3 illustrates the downlink average achievable rate as a function of the average transmit SNR  $\rho_d$  for different values of IBO,  $\delta^2$ , and  $\theta$ .  $a_m$ ,  $|t_m|$ ,  $|r_m|$ ,  $|u_k|$ , and  $|v_k|$  are distributed as  $\{\ln a_m, \ln |t_m|, \ln |r_m|, \ln |u_k|, \ln |v_k|\} \sim \mathcal{N}(0, \delta^2)$ , and  $\angle t_m, \angle r_m, \angle u_k$ , and  $\angle v_k$  are distributed as  $\{\angle t_m, \angle r_m, \angle u_k, \angle v_k\} \sim \mathcal{U}(-\theta, \theta)$ . From the figure, it can be seen that the theoretical results are closed to the simulations, which verifies the theoretical results. The achievable rate increases with the transmit SNR and approaches an upper limit when the SNR is large. Further, as the IBO increases or the mismatch parameters decrease, the upper limit of the achievable rate increases. This implies that the rate limitation is caused by both the limitation of the HPA and the reciprocity mismatch, which is consistent with analytical results.

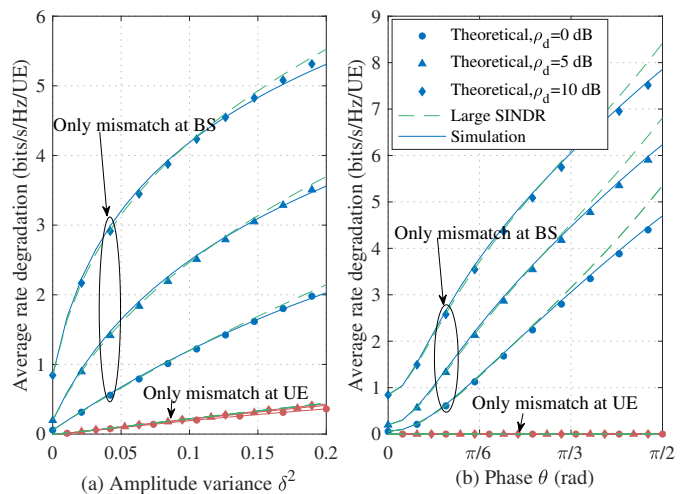


Fig. 4. Average achievable rate versus reciprocity mismatch coefficients  $\delta^2$  and  $\theta$ .

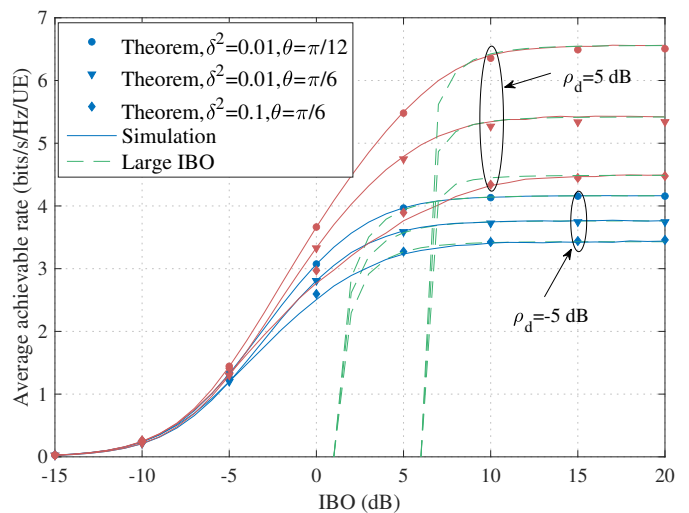


Fig. 5. Average achievable rate versus IBO.

The impacts of the amplitude and phase reciprocity mismatch on the average achievable rate degradation are shown in Fig. 4a and Fig. 4b, respectively, where the IBO is set to 10 dB. From the figures, it can be seen that both the amplitude mismatch at the BS side and the UE side degrade the achievable rate, and only the phase mismatch at the BS side can degrade the achievable rate. Further, the performance degradation caused by the mismatch at the BS side is much larger than that at the UE side and increases with the downlink transmit SNR  $\rho_d$ . This implies that the reciprocity calibration at the BS side is very essential for the TDD multi-user massive MIMO system, especially at the high SNR regime. Besides, the dashed curves are closed to the solid curves when the mismatch parameters are small. This is because when the mismatch is serious, the SINDR is not large enough, and the approximation  $\log(1 + \text{SINDR})$  with  $\log(\text{SINDR})$  is not accurate.

Further, the average achievable rate versus the IBO of the BS is demonstrated in Fig. 5 with the transmit SNR  $\rho_d$  set to  $-5$  dB and  $5$  dB. As seen from the figure, the achievable rate

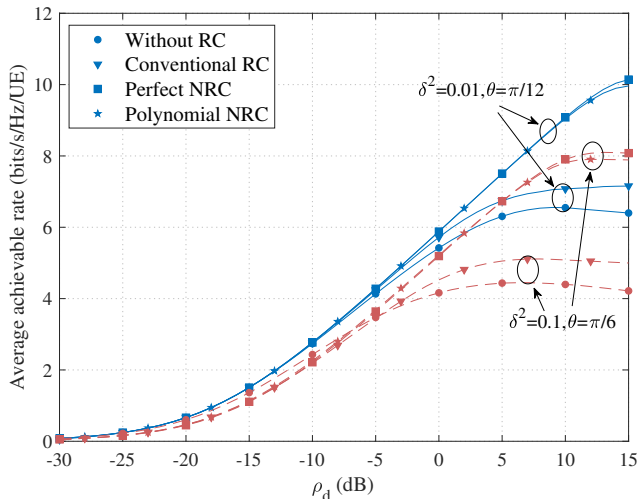


Fig. 6. Average achievable rate after RC versus the downlink transmit SNR  $\rho_d$ .

increases when IBO increases. In the small IBO region, the achievable rates are almost the same for the different system coefficients. This indicates that the poor amplification ability of the HPA greatly limits system performance. In the large IBO region, the achievable rate approaches an upper limit. As the transmit SNR  $\rho_d$  increases or the mismatch parameters decrease, the upper limit of the achievable rate increases, which implies that the transmit power and the reciprocity mismatch limit the system performance. Further, the impact of the reciprocity mismatch on the achievable rate becomes greater when both the transmit SNR and IBO are large. These results indicate that the nonlinearity of HPAs intensifies the reciprocity mismatch, which is consistent with the theoretical results in (39).

### B. Performance of Nonlinear Reciprocity Calibration

Fig. 6 demonstrates the average achievable rate versus the transmit SNR  $\rho_d$  with different calibration approaches. Note that the perfect NRC is the performance benchmark of the nonlinear reciprocity calibration. In the simulation, the polynomial order is set to  $\Pi = 5$  and the IBO is 10 dB. The training pilot length  $N$  and  $Q$  are set to 6 and 10, respectively. From the figure, it can be seen that the achievable rate increases with the transmit SNR. In the small SNR regime, the system with any reciprocity calibration approach almost has the same performance as the system without the reciprocity calibration. In the high SNR regime, the achievable rate with the nonlinear reciprocity calibration is much larger than both the rate without calibration and the rate with conventional calibration. The results imply that the reciprocity calibration is more essential to the TDD system working at the high SNR regime. The performance of the polynomial nonlinear reciprocity calibration is very closed to the perfect nonlinear reciprocity calibration. When the transmit SNR is large, the performance of the polynomial NRC is less than the perfect NRC. This performance loss is caused by the calibration error due to the OTA-polynomial training.

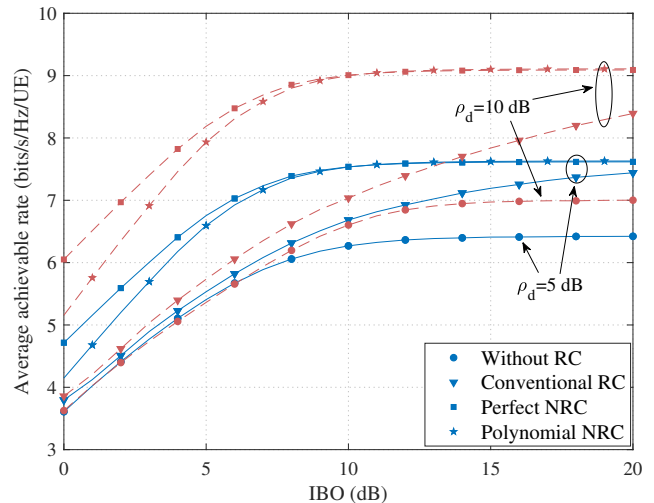


Fig. 7. Average achievable rate after the RC versus IBO.

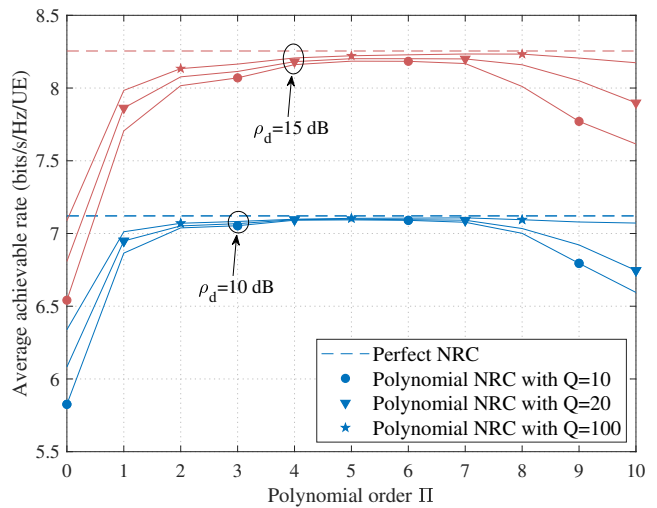


Fig. 8. Average achievable rate after RC versus the polynomial order  $\Pi$ .

The average achievable rate after the reciprocity calibration versus the IBO is illustrated in Fig. 7 with the SNR  $\rho_d$  set to 10 dB and 15 dB. The mismatch coefficients  $\{\delta^2, \theta\}$  are set to  $\{0.05, \pi/6\}$ . From the figure, the average achievable rate after the reciprocity calibration increases with the IBO. In the small IBO regime, the achievable rate of the polynomial NRC is smaller than the perfect NRC, and it gradually approaches the rate of the perfect NRC. At the large IBO regime, the polynomial NRC has the same performance as the perfect NRC. This is because the estimation error of the polynomial coefficients decreases with the increase of IBO. Further, the conventional calibration performs poorly when IBO is small, and it performs better as the IBO increases. At the very large IBO regime, the performance of the conventional calibration approaches the NRC. This is because the nonlinearity of HPA disappears when IBO is very large, and the TDD system suffers the linear reciprocity mismatch.

The relationship between the achievable rate and the polynomial order is shown in Fig. 8. The IBO is set to 5 dB, and the mismatch parameters  $\{\delta^2, \theta\}$  are set to  $\{0.1, \pi/6\}$ .

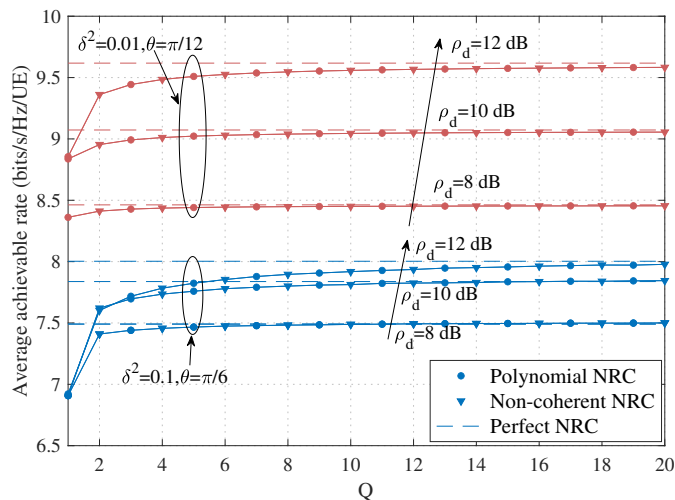


Fig. 9. Average achievable rate after the RC versus the pilot length  $Q$ .

The order  $\Pi = 0$  denotes the conventional RC approach. The figure shows that when  $Q$  is small, as the polynomial order increases, the achievable rate first increases and then decreases. This is because the received training signal is noisy when  $Q$  is small. The high-order polynomial fitting by using the noisy data suffers over-fitting, which makes the learned polynomial function perfectly fit the noisy training data, but fail to fit the actual function. When the pilot length  $Q$  is large, the performance of the polynomial NRC always approaches the perfect NRC as the polynomial order is large. This is because the impact of the noise on the polynomial fitting gradually decreases and even vanishes, as the length of the training pilots increases. Consequently, to reduce the computational complexity and achieve a good performance of the polynomial nonlinear reciprocity calibration, the polynomial order should not be very large.

Fig. 9 shows the average achievable rate versus the pilot length. The IBO is set 10 dB and polynomial order  $\Pi$  is set to 5. In the non-coherent NRC, to gathering  $M(M-1)QN$  training signals, the antennas are divided into  $M(M-1)/2$  groups, and each groups contains two antennas. From the figure, it can be seen that the achievable rate increases with the pilot length  $Q$ , and gradually approaches to the performance of the perfect NRC. This is because the estimation errors of the polynomial coefficients decreases with the pilot length  $Q$ . Further, we can also observe that the achievable rate of the non-coherent NRC is same as that of the coherent polynomial NRC which requires less training overhead.

Finally, the computational complexity of the conventional RC and the polynomial NRC versus the antenna number is illustrated in Fig. 10. During the simulation,  $N$  is set to  $\Pi + 1$ . Fig. 10a compares the complexity of the conventional RC training and the polynomial NRC training. We find that the polynomial NRC training takes more time than the conventional RC training. Then, in Fig. 10b, we compare the computational complexity of computing the calibration coefficients. From Fig. 10b, it can be seen that the conventional linear RC is much faster than the NRC. Further, the SLP algorithm is faster than the interior point algorithm, and the

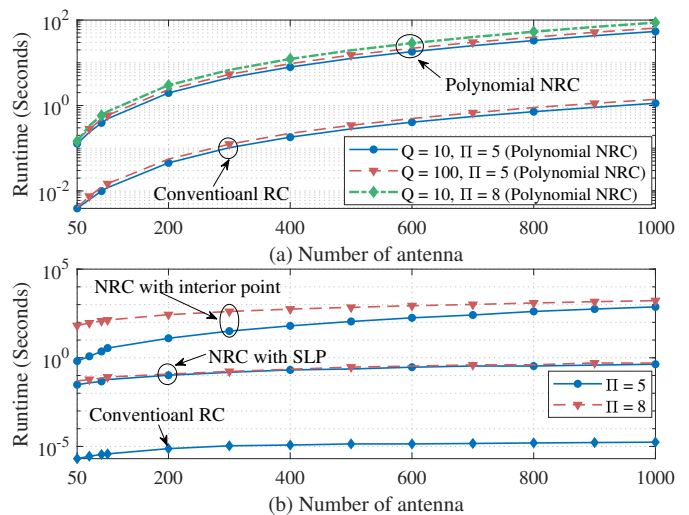


Fig. 10. Complexity of the conventional NRC and the polynomial NRC versus the antenna number.

advantages of the SLP algorithm becomes more obvious when the antennas number increases.

## VI. CONCLUSIONS

In this paper, we have studied the nonlinear reciprocity mismatch of the multi-user massive MIMO system, including the impact analysis and the nonlinear reciprocity calibration. By modeling the transmit RF gain as a nonlinear function of the transmit power, we derived the closed-form expression of the ergodic achievable rate with the nonlinear reciprocity mismatch. Based on the closed-form achievable rate, the performance loss caused by the nonlinear mismatch at the BS side and the UE side was presented, respectively. The analytical results revealed that the impact of the mismatch at the BS side was much severer than that at the UE side. To further analyze the impact of the nonlinear mismatch at the BS side, we considered a special case where the IBO was larger than zero, which demonstrated that the nonlinearity exacerbated the reciprocity mismatch. Then, we proposed a novel nonlinear reciprocity calibration approach for the BS. During the calibration, the nonlinear relationship between the mismatch factor and the transmit power was approximated by the polynomial fitting, and the polynomial coefficients were estimated by the OTA training. Finally, to compute the nonlinear calibration coefficients efficiently, we formulated an auxiliary optimization problem and proposed a fast algorithm to solve it. Due to the low complexity of the algorithm, the nonlinear NRC was easy to implement in the actual system.

### APPENDIX A PROOF OF PROPOSITION 1

According to (7), the closed-form expression of the normalization scalar  $\beta_{ZF}$  of the ZF precoding scheme can be given by

$$\beta_{ZF} = \mathbb{E} \left\{ \text{tr} \left[ (\mathbf{H}_{UL}^T \mathbf{H}_{UL}^*)^{-1} \right] \right\} \quad (69)$$

$$\stackrel{(a)}{=} \frac{M \text{tr} \{ (\mathbf{B} \Phi^2 \mathbf{B}^*)^{-1} \}}{\text{tr} \{ \mathbf{R} \mathbf{R}^* \} (M - K)},$$

where the step (a) is because  $(\mathbf{H}_{\text{UL}}^T \mathbf{H}_{\text{UL}}^*)^{-1}$  obeys the inverse-Wishart distribution denoted as  $(\mathbf{H}_{\text{UL}}^T \mathbf{H}_{\text{UL}}^*)^{-1} \sim \mathcal{W}^{-1}(M(\mathbf{B}\Phi^2\mathbf{B}^* \text{tr}\{\mathbf{R}\mathbf{R}^*\})^{-1}, K, M)$ .

According to [24, Eq. (14)] and [27, Eq. (40)], we further approximate  $(\mathbf{H}_{\text{UL}}^T \mathbf{H}_{\text{UL}}^*)^{-1}$  into a diagonal matrix as

$$(\mathbf{H}_{\text{UL}}^T \mathbf{H}_{\text{UL}}^*)^{-1} \approx \frac{1}{\text{tr}\{\mathbf{R}\mathbf{R}^*\}} \text{diag}\left(\frac{1}{|b_1|^2 \phi_1^2}, \dots, \frac{1}{|b_K|^2 \phi_K^2}\right). \quad (70)$$

Then, the transmitted signal  $x_{b,m}$  of the  $m$ -th antenna with the ZF precoding can be rewritten as

$$x_{b,m} = [\mathbf{W}]_m \mathbf{s} = \frac{r_m^*}{\text{tr}\{\mathbf{R}\mathbf{R}^*\} \sqrt{\beta_{\text{ZF}}}} \sum_{i=1}^K \frac{h_{m,i}^* s_i}{b_i \phi_k}. \quad (71)$$

The variance of  $x_{b,m}$  can be derived as

$$\begin{aligned} \sigma_{x,m}^2 &= \mathbb{E}\{|x_{b,m}|^2\} \\ &= \frac{\mathbb{E}\left\{\left|\sum_{i=1}^K h_{m,i}^* s_i b_i^{-1} \phi_i^{-1}\right|^2\right\} |r_m|^2}{(\text{tr}\{\mathbf{R}\mathbf{R}^*\})^2 \beta_{\text{ZF}}} \\ &= \frac{(M-K)\rho_t |r_m|^2 \sum_{i=1}^K |h_{m,i}|^2 |b_i \phi_i|^{-2}}{M \text{tr}\{\mathbf{R}\mathbf{R}^*\} \sum_{i=1}^K |b_i \phi_i|^{-2}} \\ &\stackrel{(b)}{=} \frac{|r_m|^2 \rho_t}{\text{tr}\{\mathbf{R}\mathbf{R}^*\}}, \end{aligned} \quad (72)$$

where (b) holds due to LLN. By substituting (72) into  $\mu(x)$ , the closed-form expression of the linear scalar  $g_{\text{ZF},m}$  of ZF can be denoted as

$$g_{\text{ZF},m} = t_m \mu\left(\frac{A_{\text{Ast},m} \sqrt{\text{tr}\{\mathbf{R}\mathbf{R}^*\}}}{\sqrt{|r_m|^2 \rho_t}}\right). \quad (73)$$

The closed-form expression of the variance  $\sigma_{d,m}^2$  of the nonlinear distortion  $d_m$  of ZF can be denoted as

$$\sigma_{\text{ZF},m}^2 = |t_m|^2 \lambda_m \left(\sqrt{\frac{|r_m|^2 \rho_t}{\text{tr}\{\mathbf{R}\mathbf{R}^*\}}}\right). \quad (74)$$

The effective downlink channel gain  $\mathbf{H}_{\text{eq}} = \mathbf{H}_{\text{DL}} \mathbf{W}$  for ZF can be denoted as

$$\begin{aligned} \mathbf{H}_{\text{eq}} &= \frac{1}{\sqrt{\beta_{\text{ZF}}}} \mathbf{H}_{\text{DL}} \mathbf{H}_{\text{UL}}^* (\mathbf{H}_{\text{UL}}^T \mathbf{H}_{\text{UL}}^*)^{-1} \\ &= \frac{1}{\sqrt{\beta_{\text{ZF}}}} \mathbf{U} \bar{\mathbf{H}} \mathbf{R}^{-1} \mathbf{G}_{\text{ZF}} \bar{\mathbf{H}}^H (\bar{\mathbf{H}} \bar{\mathbf{H}}^H)^{-1} \mathbf{B}^{-1}, \end{aligned} \quad (75)$$

where  $\bar{\mathbf{H}} = \mathbf{H}\mathbf{R}$ , and  $\mathbf{G}_{\text{ZF}} = \text{diag}(g_{\text{ZF},1}, \dots, g_{\text{ZF},M})$ . To accurately derive the SINDR, we further denoted  $\mathbf{R}^{-1} \mathbf{G}_{\text{ZF}} = \alpha \mathbf{I}_M + \mathbf{\Delta}_{\text{rg}}$ , where  $\alpha$  is the arithmetic mean of  $g_m/r_m$ , i.e.,  $\alpha = \frac{1}{M} \sum_{m=1}^M g_{\text{ZF},m}/r_m$ ,  $\mathbf{\Delta} = \text{diag}(\Delta_1, \dots, \Delta_M)$ , the arithmetic mean of  $\Delta_m$  is zero and the variance of  $\Delta_m$  is equal to the variance of  $g_{\text{ZF},m}$ . Then, the effective downlink channel can be further given by

$$\begin{aligned} \mathbf{H}_{\text{eq}} &= \frac{1}{\sqrt{\beta_{\text{ZF}}}} \mathbf{U} [\bar{\mathbf{H}} \mathbf{\Delta}_{\text{rg}} \bar{\mathbf{H}}^H (\bar{\mathbf{H}} \bar{\mathbf{H}}^H)^{-1} + \alpha] \mathbf{B}^{-1} \\ &= \frac{1}{\sqrt{\beta_{\text{ZF}}}} \mathbf{U} \left[ \frac{1}{\text{tr}\{\mathbf{R}\mathbf{R}^*\}} \bar{\mathbf{H}} \mathbf{\Delta}_{\text{rg}} \bar{\mathbf{H}}^H \right. \\ &\quad \left. \cdot \text{diag}\left(\frac{1}{\phi_1^2}, \dots, \frac{1}{\phi_K^2}\right) + \alpha \mathbf{I}_K \right] \mathbf{B}^{-1}, \end{aligned} \quad (76)$$

In light of (76), the statistical effective channel gain  $\mathbb{E}\{h_{\text{eq},k,k}\}$  can be derived as

$$\begin{aligned} \mathbb{E}\{h_{\text{eq},k,k}\} &= \frac{u_k}{\sqrt{\beta_{\text{ZF}} b_k}} \mathbb{E}\left\{\frac{\bar{\mathbf{h}}_k \mathbf{\Delta}_{\text{rg}} \bar{\mathbf{h}}_k^H}{\text{tr}\{\mathbf{R}\mathbf{R}^*\} \phi_k^2} + \alpha\right\} \\ &= \frac{u_k \text{tr}\{\mathbf{G}_{\text{ZF}} \mathbf{R}^* - \alpha \mathbf{R}\mathbf{R}^*\}}{\sqrt{\beta_{\text{ZF}} b_k} \text{tr}\{\mathbf{R}\mathbf{R}^*\}} + \frac{\alpha u_k}{\sqrt{\beta_{\text{ZF}} b_k}} \\ &= \frac{u_k \text{tr}\{\mathbf{G}_{\text{ZF}} \mathbf{R}^*\}}{\sqrt{\beta_{\text{ZF}} b_k} \text{tr}\{\mathbf{R}\mathbf{R}^*\}}. \end{aligned} \quad (77)$$

Then, the power of the effective signal received at the  $k$ -th UE can be given by

$$\begin{aligned} \Upsilon_{\text{ZF},k}^{\text{ES}} &= a_0 \rho_t |\mathbb{E}\{h_{\text{eq},k,k}\}|^2 \\ &= \frac{a_0 \rho_t (M-K) |u_k \text{tr}\{\mathbf{G}_{\text{ZF}} \mathbf{R}^*\}|^2}{M |b_k|^2 \text{tr}\{(\mathbf{B}\Phi^2\mathbf{B}^*)^{-1}\} \text{tr}\{\mathbf{R}\mathbf{R}^*\}}. \end{aligned} \quad (78)$$

The second moment of the statistical effective channel gain of ZF can be denoted as

$$\begin{aligned} &\mathbb{E}\{|h_{\text{eq},k,k}|^2\} \\ &= \frac{|u_k|^2}{\beta_{\text{ZF}} |b_k|^2} \mathbb{E}\left\{\left|\frac{\bar{\mathbf{h}}_k \mathbf{\Delta}_{\text{rg}} \bar{\mathbf{h}}_k^H}{\text{tr}\{\mathbf{R}\mathbf{R}^*\} \phi_k^2} + \alpha\right|^2\right\} \\ &= \frac{|u_k|^2}{\beta_{\text{ZF}} |b_k|^2} \left[\frac{\mathbb{E}\{|\bar{\mathbf{h}}_k \mathbf{\Delta}_{\text{rg}} \bar{\mathbf{h}}_k^H|^2\}}{(\text{tr}\{\mathbf{R}\mathbf{R}^*\})^2 \phi_k^4} + \frac{\mathbb{E}\{\bar{\mathbf{h}}_k \mathbf{\Delta}_{\text{rg}} \bar{\mathbf{h}}_k^H\}}{\text{tr}\{\mathbf{R}\mathbf{R}^*\} \phi_k^2} \right. \\ &\quad \left. + \frac{\mathbb{E}\{\bar{\mathbf{h}}_k \mathbf{\Delta}_{\text{rg}}^* \bar{\mathbf{h}}_k^H\}}{\text{tr}\{\mathbf{R}\mathbf{R}^*\} \phi_k^2} + |\alpha|^2\right] \\ &= \frac{|u_k|^2}{\beta_{\text{ZF}} |b_k|^2} \left[\frac{\text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R}) \mathbf{R} \mathbf{R}^* (\mathbf{G}_{\text{ZF}}^* - \alpha \mathbf{R}^*)\}}{(\text{tr}\{\mathbf{R}\mathbf{R}^*\})^2} \right. \\ &\quad \left. + \frac{\text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R}) \mathbf{R}^*\} \text{tr}\{(\mathbf{G}_{\text{ZF}}^* - \alpha \mathbf{R}^*) \mathbf{R}\}}{(\text{tr}\{\mathbf{R}\mathbf{R}^*\})^2} \right. \\ &\quad \left. + \frac{2 \text{Re}\{\text{tr}\{(\mathbf{G}_{\text{ZF}}^* - \alpha \mathbf{R}^*) \mathbf{R}\}\}}{\text{tr}\{\mathbf{R}\mathbf{R}^*\}} + |\alpha|^2\right] \\ &= \frac{|u_k|^2 \text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}}^* - \alpha \mathbf{R}^*)\}}{M \beta_{\text{ZF}} |b_k|^2 \text{tr}\{\mathbf{R}\mathbf{R}^*\}} \\ &\quad + \frac{|u_k \text{tr}\{\mathbf{G}_{\text{ZF}} \mathbf{R}^*\}|^2}{M \beta_{\text{ZF}} |b_k|^2 (\text{tr}\{\mathbf{R}\mathbf{R}^*\})^2}. \end{aligned} \quad (79)$$

Hence, the power of the self-interference can be given by

$$\begin{aligned} \Upsilon_{\text{ZF},k}^{\text{SI}} &= a_0 \rho_t \text{Var}\{h_{\text{eq},k,k}\} \\ &= a_0 \rho_t [\mathbb{E}\{|h_{\text{eq},k,k}|^2\} - |\mathbb{E}\{h_{\text{eq},k,k}\}|^2] \\ &= \frac{a_0 \rho_t |u_k|^2 \text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}}^* - \alpha \mathbf{R}^*)\}}{(M-K)^{-1} M^2 |b_k|^2 \text{tr}\{(\mathbf{B}\Phi^2\mathbf{B}^*)^{-1}\}}. \end{aligned} \quad (80)$$

Further, the power of the multi-user interference of ZF can be derived as

$$\begin{aligned} \Upsilon_{\text{ZF},k}^{\text{MUI}} &= a_0 \rho_t \sum_{i \neq k}^K \mathbb{E}\{|h_{\text{eq},k,i}|^2\} \\ &= \frac{a_0 \rho_t |u_k|^2}{\beta_{\text{ZF}}} \sum_{i \neq k}^K \mathbb{E}\left\{\left|\frac{\bar{\mathbf{h}}_k \mathbf{\Delta}_{\text{rg}} \bar{\mathbf{h}}_i^H}{\text{tr}\{\mathbf{R}\mathbf{R}^*\} b_i \phi_i^2}\right|^2\right\} \\ &= \sum_{i \neq k}^K \frac{a_0 \rho_t |u_k|^2 \phi_k^2 \text{tr}\{(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})(\mathbf{G}_{\text{ZF}} - \alpha \mathbf{R})^*\}}{(M-K)^{-1} M^2 |u_k^{-1} b_i|^2 \phi_i^2 \text{tr}\{(\mathbf{B}\Phi^2\mathbf{B}^*)^{-1}\}}. \end{aligned} \quad (81)$$



Finally, the power of the nonlinear distortion can be evaluated as

$$\Upsilon_{ZF,k}^{\text{NLI}} = a_0 |u_k|^2 \phi_k^2 \mathbb{E} \{ \mathbf{h}_k \boldsymbol{\Sigma}_{ZF} \mathbf{h}_k^H \} = a_0 |u_k|^2 \phi_k^2 \text{tr} \{ \boldsymbol{\Sigma}_{ZF} \}, \quad (82)$$

where  $\boldsymbol{\Sigma}_{ZF} = \text{diag}(\sigma_{ZF,1}^2, \dots, \sigma_{ZF,M}^2)$ .

#### APPENDIX B PROOF OF PROPOSITION 3

As the HPA works with the large IBO state, the average power of the input signal is less than the saturation level, i.e.,  $\frac{\sigma_{x,m}^2}{A_{\text{sat},m}^2} < 1$ . Hence, we use Taylor series centered at  $\epsilon$  to approximate the function  $f(x) = \mu(1/x)$ , where  $\epsilon$  is a very small positive constant. By utilizing the Taylor expansion, the function  $f(x)$  denoted as

$$f_\epsilon(x) = f(\epsilon) - f'(\epsilon)(x - \epsilon) + \frac{f''(\epsilon)}{2}(x - \epsilon)^2 + \frac{f'''(\epsilon)}{6}(x - \epsilon)^3 + \mathcal{O}((x - \epsilon)^3). \quad (83)$$

When  $\epsilon$  approaches to 0,  $f(x)$  can be given by

$$f(x) = \lim_{\epsilon \rightarrow 0} f_\epsilon(x) \approx 1 - x^2 + \mathcal{O}(x^3). \quad (84)$$

By substituting  $\sigma_{x,m}^2$  denoted in (72) into  $f_0(x)$ ,  $g_{ZF,m}$  can be approximated as

$$g_{ZF,m} = t_m \mu \left( \frac{A_{\text{Ast},m} \sqrt{\text{tr} \{ \mathbf{R} \mathbf{R}^* \}}}{\sqrt{|r_m|^2 \rho_t}} \right) \approx t_m \left( 1 - \frac{|r_m|^2 \rho_t A_{\text{sat},m}^{-2}}{\text{tr} \{ \mathbf{R} \mathbf{R}^* \}} \right). \quad (85)$$

Similarly, by exploiting the Taylor expansion of  $\lambda_m(1/x)$ , the variance of the nonlinear distortion can be approximated as

$$\sigma_{ZF,m}^2 \approx \frac{|r_m|^6 \rho_t^3 A_{\text{sat},m}^{-6}}{2(\text{tr} \{ \mathbf{R} \mathbf{R}^* \})^3}. \quad (86)$$

By substituting (85) into (27), the power of effective signal can be further given by

$$\begin{aligned} \tilde{\Upsilon}_{ZF,k}^{\text{ES}} &= a_0 \rho_t \frac{\left| \sum_{m=1}^M t_m r_m^* \left( 1 - \frac{|r_m|^2 \rho_t A_{\text{sat},m}^{-2}}{\text{tr} \{ \mathbf{R} \mathbf{R}^* \}} \right) \right|^2}{(M - K)^{-1} M \text{tr} \{ \boldsymbol{\Phi}^{-2} \} \sum_{m=1}^M |r_m|^2} \\ &\stackrel{(a)}{=} a_0 \rho_t \frac{(M - K) |\mathbb{E} \{ t_m r_m^* \}|^2}{\text{tr} \{ \boldsymbol{\Phi}^{-2} \} \mathbb{E} \{ |r_m|^2 \}} \left( 1 - \frac{\mathbb{E} \{ |a_m|^{-2} \}}{M \rho_t^{-1} A_{\text{sat}}^2} \right)^2 \\ &\stackrel{(b)}{\approx} a_0 \rho_t \frac{(M - K) |\mathbb{E} \{ t_m r_m^* \}|^2}{\text{tr} \{ \boldsymbol{\Phi}^{-2} \} \mathbb{E} \{ |r_m|^2 \}} \left( 1 - \frac{2 \mathbb{E} \{ |a_m|^{-2} \}}{M \rho_t^{-1} A_{\text{sat}}^2} \right), \end{aligned} \quad (87)$$

where (a) holds due to LLN, and (c) is conditioned on ignoring the high-order infinitesimality of  $\rho_t/(M A_{\text{sat}})$ . Substituting

(85) into (28) and (29), the sum of self-interference and multi-user interference can be further denoted as

$$\begin{aligned} \tilde{\Upsilon}_{ZF,k}^{\text{SI}} &= a_0 \rho_t \frac{M - K}{M^2 \text{tr} \{ \boldsymbol{\Phi}^{-2} \}} \sum_{m=1}^M |g_{ZF,m} - \alpha r_m|^2 \\ &= a_0 \rho_t \frac{M - K}{M^2 \text{tr} \{ \boldsymbol{\Phi}^{-2} \}} \sum_{m=1}^M [ |g_{ZF,m}|^2 + |\alpha r_m|^2 \\ &\quad - 2 \text{Re} \{ g_{ZF,m} \alpha^* r_m^* \} ] \\ &\stackrel{(c)}{=} \frac{a_0 \rho_t (M - K)}{M \text{tr} \{ \boldsymbol{\Phi}^{-2} \}} \left[ \mathbb{E} \{ |t_m|^2 \} + \left| \mathbb{E} \left\{ \frac{t_m}{r_m} \right\} \right|^2 \mathbb{E} \{ |r_m|^2 \} \right. \\ &\quad \left. - 2 \text{Re} \left( \mathbb{E} \left\{ \frac{t_m}{r_m} \right\} \mathbb{E} \{ t_m^* r_m \} \right) \right] \left( 1 - \frac{2 \rho_t \mathbb{E} \{ a_m^{-2} \}}{M A_{\text{sat}}^2} \right), \end{aligned} \quad (88)$$

where (c) holds due to LLN and ignoring the high-order infinitesimality of  $\rho_t/(M A_{\text{sat}})$ . Similarly, the power of multi-user interference can be rewritten as

$$\begin{aligned} \Upsilon_{ZF,k}^{\text{MUI}} &= \frac{a_0 \rho_t (M - K) \phi_k^2 \sum_{i \neq k}^K \phi_i^{-2}}{M \text{tr} \{ \boldsymbol{\Phi}^{-2} \}} [\mathbb{E} \{ |t_m|^2 \} \\ &\quad - 2 \text{Re} \left( \mathbb{E} \left\{ \frac{t_m}{r_m} \right\} \mathbb{E} \{ t_m^* r_m \} \right) \\ &\quad + \left| \mathbb{E} \left\{ \frac{t_m}{r_m} \right\} \right|^2 \mathbb{E} \{ |r_m|^2 \} ] \left( 1 - \frac{2 \rho_t \mathbb{E} \{ a_m^{-2} \}}{M A_{\text{sat}}^2} \right). \end{aligned} \quad (89)$$

According to (86),  $\sigma_{d,m}^2$  is a higher-order infinitesimality of  $\frac{\rho_t}{M A_{\text{sat}}^2}$  when  $\frac{\rho_t}{M A_{\text{sat}}^2} < 1$ , the impact of  $\sigma_{d,m}^2$  is slight and can be ignored. Then, by substituting (87), (88), and (89) into (26) and exploiting the statistic properties of  $t_m$ ,  $r_m$ , and  $a_m$ , SINR with the large IBO can be given by (39). Therefore, Proposition 3 holds.

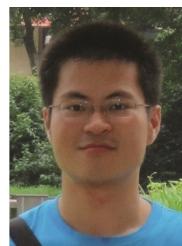
#### REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [2] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, Nov. 2016.
- [3] I. F. Akyildiz, S. Nie, S. C. Lin, and M. Chandrasekaran, "5G roadmap: 10 key enabling technologies," *Comput. Networks*, vol. 106, pp. 17–48, Sep. 2016.
- [4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [5] C. Shan, L. Chen, X. Chen, and W. Wang, "A general matched filter design for reciprocity calibration in multiuser massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8939–8943, Sep. 2018.
- [6] C. Shan, Y. Zhang, L. Chen, X. Chen, and W. Wang, "Performance Analysis of Large Scale Antenna System with Carrier Frequency Offset, Quasi-Static Mismatch and Channel Estimation Error," *IEEE Access*, vol. 5, pp. 26 135–26 145, Nov. 2017.
- [7] D. Mi, M. Dianati, L. Zhang, S. Muhaidat, and R. Tafazolli, "Massive MIMO Performance With Imperfect Channel Reciprocity and Channel Estimation Error," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3734–3749, Sep. 2017.
- [8] W. Zhang, H. Ren, C. Pan, M. Chen, R. C. De Lamare, B. Du, and J. Dai, "Large-scale antenna systems with UL/DL hardware mismatch: Achievable rates analysis and calibration," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1216–1229, Apr. 2015.

- [9] O. Raeesi, A. Gokceoglu, Y. Zou, E. Björnson, and M. Valkama, "Performance Analysis of Multi-User Massive MIMO Downlink under Channel Non-Reciprocity and Imperfect CSI," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2456–2471, Jun. 2018.
- [10] H. Wei, D. Wang, J. Wang, and X. You, "Impact of RF mismatches on the performance of massive MIMO systems with ZF precoding," *Sci. China Inf. Sci.*, vol. 59, no. 2, pp. 1–14, Jan. 2016.
- [11] X. Jiang, M. Cirkic, F. Kaltenberger, E. G. Larsson, L. Deneire, and R. Knopp, "MIMO-TDD reciprocity under hardware imbalances: Experimental results," *IEEE Int. Conf. Commun.*, vol. 2015-Sept, pp. 4949–4953, Jun. 2015.
- [12] R. Nie, L. Chen, C. Shan, and X. Chen, "A decentralized reciprocity calibration approach for cooperative MIMO," *IEEE Access*, vol. 7, pp. 1560–1569, Dec. 2019.
- [13] K. Nishimori, K. Cho, Y. Takatori, and T. Hori, "Automatic calibration method using transmitting signals of an adaptive array for TDD systems," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1636–1640, Nov. 2001.
- [14] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation," in *Proc. IEEE Radio Wirel. Conf. (RAWCON)*, Aug. 2003, pp. 183–186.
- [15] J. Liu, A. Bourdoux, J. Craninckx, P. Wambacq, B. Côme, S. Donnay, and A. Barel, "OFDM-MIMO WLAN AP front-end gain and phase mismatch calibration," in *Proc. 2004 IEEE Radio Wirel. Conf. (RAWCON)*, Feb. 2004, pp. 151–154.
- [16] Jian Liu, G. Vandersteven, J. Craninckx, M. Libois, M. Wouters, F. Petre, and A. Barel, "A novel and low-cost analog front-end mismatch calibration scheme for MIMO-OFDM WLANs," in *Proc. 2006 IEEE Radio Wireless Symp.*, Apr. 2006, pp. 219–222.
- [17] A. Benzin and G. Caire, "Internal Self-Calibration Methods for Large Scale Array Transceiver Software-Defined Radios," in *Proc. Int. ITG Workshop Smart Antennas (WSA)*, Mar. 2017, pp. 49–56.
- [18] X. Luo, F. Yang, and H. Zhu, "Massive MIMO Self-Calibration: Optimal Interconnection for Full Calibration," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10357–10371, Nov. 2019.
- [19] M. Guillaud, D. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. 8th Int. Symp. Signal Process. Applic. (ISSPA)*, Intergovernmental Panel on Climate Change, Ed., vol. 1, no. 9, Jan. 2005, pp. 403–406.
- [20] F. Kaltenberger, J. Haiyong, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Proc. Futur. Netw. Mob. Summit*, Mar. 2011, pp. 1–10.
- [21] B. Kouassi, I. Ghauri, and L. Deneire, "Estimation of Time-Domain Calibration Parameters to Restore MIMO-TDD Channel Reciprocity," in *Proc. 7th Int. ICST Conf. Cogn. Radio Oriented Wirel. Networks Commun.*, Jul. 2012, pp. 254–258.
- [22] R1-092359, "Hardware calibration requirement for dual layer beamforming," Huawei, Los Angeles, CA, USA, 3GPP RAN1 57, Jun. 2009.
- [23] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: practical many-antenna base stations," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Networking (Mobicom)*, Aug. 2012, pp. 53–64.
- [24] H. Wei, D. Wang, H. Zhu, J. Wang, S. Sun, and X. You, "Mutual Coupling Calibration for Multiuser Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 606–619, Jan. 2016.
- [25] R. Rogalin, O. Y. Bursalioğlu, H. C. Papadopoulos, G. Caire, and A. F. Molisch, "Hardware-impairment compensation for enabling distributed large-scale MIMO," in *Proc. 2013 Inf. Theory Appl. Work. (ITA)*, Feb. 2013, pp. 304–313.
- [26] X. Jiang, A. Decurninge, K. Gopala, F. Kaltenberger, M. Guillaud, S. Member, D. Slock, and L. Deneire, "A Framework for Over-the-air Reciprocity Calibration for TDD Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5975–5990, Sep. 2018.
- [27] R. Nie, L. Chen, N. Zhao, Y. Chen, F. R. Yu, and G. Wei, "Relaying Systems With Reciprocity Mismatch: Impact Analysis and Calibration," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4035–4049, Jul. 2020.
- [28] J. Guerreiro, R. Dinis, and P. Montezuma, "Analytical Performance Evaluation of Precoding Techniques for Nonlinear Massive MIMO Systems with Channel Estimation Errors," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1440–1451, Apr. 2018.
- [29] E. Balti and M. Guizani, "Impact of Non-Linear High-Power Amplifiers on Cooperative Relaying Systems," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4163–4175, Oct. 2017.
- [30] C. Yu, J. Jing, H. Shao, Z. H. Jiang, P. Yan, X. W. Zhu, W. Hong, and A. Zhu, "Full-Angle Digital Predistortion of 5G Millimeter-Wave Massive MIMO Transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2847–2860, Jul. 2019.
- [31] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for Hybrid Beamforming Array Utilizing Embedded Over-the-Air Diversity Feedbacks," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 5235–5248, Dec. 2019.
- [32] R. Raich, H. Qian, and G. T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1468–1479, Sep. 2004.
- [33] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot Contamination and Precoding in Multi-Cell TDD Systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [34] A. Minasian, S. ShahbazPanahi, and R. S. Adve, "Distributed massive MIMO systems with non-reciprocal channels: Impacts and robust beamforming," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5261–5277, Nov. 2018.
- [35] H. E. Rowe, "Memoryless Nonlinearities With Gaussian Inputs: Elementary Results," *Bell Syst. Tech. J.*, vol. 61, no. 7, pp. 1519–1526, Sep. 1982.
- [36] R. Krishnan, M. R. Khanzadi, N. Krishnan, Y. Wu, A. GraellAmat, T. Eriksson, and R. Schober, "Linear Massive MIMO Precoders in the Presence of Phase Noise - A Large-Scale Analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3057–3071, May 2016.
- [37] R. L. Eubank and B. R. Jayasuriya, "The asymptotic average squared error for polynomial regression," *Statistics (Ber.)*, vol. 24, no. 4, pp. 311–319, 1993.
- [38] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [39] M. Powell, "A fast algorithm for nonlinearly constrained optimization calculations," in *Numerical analysis*. Berlin, Germany: Springer-Verlag, 1978, vol. 630, pp. 144–157.



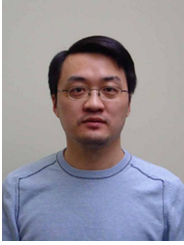
**Rongjiang Nie** received the B.S. degree in electronic information science and technology from Southwest Jiaotong University, Chengdu, China, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His research interests include reciprocity calibration and cooperative multiple-input multiple-output.



**Li Chen** received the B.E. in electrical and information engineering from Harbin Institute of Technology, Harbin, China, in 2009 and the Ph.D. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2014. He is currently an Associate Professor with the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His research interests include integrated computation and communication, integrated sensing and communication.



**Nan Zhao** (Senior Member, IEEE) is currently a Professor at Dalian University of Technology, China. He received the Ph.D. degree in information and communication engineering in 2011, from Harbin Institute of Technology, Harbin, China. Dr. Zhao is serving on the editorial boards of IEEE Wireless Communications and IEEE Wireless Communications Letters. He won the best paper awards in IEEE VTC 2017 Spring, ICNC 2018, WCSP 2018 and WCSP 2019. He also received the IEEE Communications Society Asia Pacific Board Outstanding Young Researcher Award in 2018.



**Yunfei Chen** (Senior Member, IEEE) received his B.E. and M.E. degrees in electronics engineering from Shanghai Jiaotong University, Shanghai, P.R.China, in 1998 and 2001, respectively. He received his Ph.D. degree from the University of Alberta in 2006. He is currently working as an Associate Professor at the University of Warwick, U.K. His research interests include wireless communications, cognitive radios, wireless relaying and energy harvesting.



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

Society of Astronautics.



**Xianbin Wang** (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the National University of Singapore in 2001.

He is currently a Professor and Tier 1 Canada Research Chair with Western University, Canada. Prior to joining Western, he was with the Communications Research Centre Canada (CRC) as a Research Scientist/Senior Research Scientist from July 2002 to December 2007. From January 2001 to July 2002, he was a System Designer with STMicroelectronics. His current research interests include

5G and beyond, the Internet-of-Things, communications security, machine learning and intelligent communications. He has over 400 peer-reviewed journals and conference papers, in addition to 30 granted and pending patents and several standard contributions.

Dr. Wang is a fellow of Canadian Academy of Engineering, a fellow of Engineering Institute of Canada, and an IEEE Distinguished Lecturer. He has received many awards and recognitions, including Canada Research Chair, CRC President's Excellence Award, Canadian Federal Government Public Service Award, Ontario Early Researcher Award and six IEEE Best Paper Awards. He currently serves as an Editor/Associate Editor of the IEEE Transactions on Communications, the IEEE Transactions on Broadcasting, and the IEEE Transactions on Vehicular Technology. He was also an Associate Editor of the IEEE Transactions on Wireless Communications from 2007 to 2011, and the IEEE Wireless Communications Letters from 2011 to 2016. He was involved in many IEEE conferences including GLOBECOM, ICC, VTC, PIMRC, WCNC and CWIT, in different roles such as symposium chair, tutorial instructor, track chair, session chair, and TPC co-chair. He is currently serving as the Vice Chair of the IEEE London Section and the Chair of ComSoc Signal Processing and Computing for Communications Technical Committee.