Self-Calibration for Massive MIMO with Channel Reciprocity and Channel Estimation Errors

De Mi*, Lei Zhang†, Mehrdad Dianati ‡, Sami Muhaïdat§, Pei Xiao* and Rahim Tafazolli*

*Institute for Communication Systems, University of Surrey, United Kingdom
†School of Engineering, University of Glasgow, United Kingdom
‡Warwick Manufacturing Group, University of Warwick, United Kingdom
§Department of Electrical and Computer Engineering, Khalifa University, United Arab Emirates

Email:d.mi@surrey.ac.uk, lei.zhang@glasgow.ac.uk, m.dianati@warwick.ac.uk, muhaïdat@ieee.org, p.xiao@surrey.ac.uk, r.tafazolli@surrey.ac.uk

Abstract—In time-division-duplexing (TDD) massive multiple-input multiple-output (MIMO) systems, channel reciprocity is exploited to overcome the overwhelming pilot training and the feedback overhead. However, in practical scenarios, the imperfections in channel reciprocity, mainly caused by radio-frequency mismatches among the antennas at the base station side, can significantly degrade the system performance and might become a performance limiting factor. In order to compensate for these imperfections, we present and investigate two new calibration schemes for TDD-based massive multi-user MIMO systems, namely, relative calibration and inverse calibration. In particular, the design of the proposed inverse calibration takes into account a compound effect of channel reciprocity error and channel estimation error. We further derive closed-form expressions for the ergodic sum rate, assuming maximum error and channel estimation error. We demonstrate that the inverse calibration scheme outperforms the traditional relative calibration scheme. The proposed analytical results are also verified by simulated illustrations.

I. INTRODUCTION

Massive MIMO (multiple-input-multiple-output) is identified as a promising technological paradigm which has been proposed in order to meet some of 5G requirements, including data rates of 10-20 Gbps and a latency of less than 1 msec [1]-[4]. Exploiting channel reciprocity, time-division-duplexing (TDD) operation enables the channel state information (CSI) acquisition in massive MIMO with an affordable overhead that is independent of the number of base station (BS) antennas [5]. Most prior studies assume the perfect channel reciprocity by constraining the time delay from the uplink (UL) to the downlink (DL) within the coherence time of the channel [6], [7]. However, the assumption of the perfect reciprocity is unrealistic in practical systems even within the coherence time, due to the fact that radio-frequency (RF) transceivers introduce amplitude and phase mismatches between the UL and the DL [8]. The imperfect channel reciprocity contaminates the estimate of the effective channel response. This causes a significant degradation in the performance of linear precoding schemes, due to their sensitivity to the CSI accuracy. Our prior work [9] has thoroughly investigated such performance degradation for two typical linear precoders, i.e., maximum radio transmission (MRT) and zero-forcing (ZF), with considerations of the imperfect channel estimation. The results in [9] show that both MRT and ZF are severely affected by the compound effect of the reciprocity and estimation errors.

In principle, calibration schemes for a precoded TDD MIMO system contain the estimation of transmit (Tx) and receive (Rx) RF frontends’ responses or equivalently calibration coefficients and the design of the calibration matrix to calibrate the precoders. For the calibration coefficients estimation, we explicitly focus on the so-called self-calibration in massive MIMO systems since it can be implemented at the BS side only and without exchanging calibration pilots between the BS and user terminals (UTs) [10]. Prior studies proposed different methods to realise self-calibration in massive MIMO systems [11], [12]. In [11], the use of antenna coupling at the BS has been proposed to measure the calibration coefficients. This method is very sensitive to the scatterings near the BS antennas. A practical study in [12] presented a method where an additional RF transceiver is used as a reference to exchange calibration pilots with other BS antennas’ transceivers. However, the method in [12] is sensitive to the placement of the reference transceiver. In order to obtain reliable estimates of the calibration coefficients, additional calibration circuits can be applied at the BS, as one example shown in [13]. The self-calibration scheme in [12], [13] is known as relative calibration, which has been widely considered in the context of the massive MIMO system [14], [15].

Although a number of recent progress has been developed from the perspective of the reciprocity calibration, various issues still remain open, such that, to the best knowledge of the authors, there have been no calibration schemes taking into account the compound effect of the reciprocity and channel estimation errors. In order to fill these research gaps among the existing literature, we consider a low-cost calibration...
A TDD massive MU-MIMO System with calibration circuits.

Fig. 1. A TDD massive MU-MIMO System with calibration circuits.

circuit that presented in [13], and expand it into the TDD massive MU-MIMO system. Such design enables the BS to estimate the RF responses reliably, also to perform the relative calibration or what we call ‘inverse calibration’ (in the sense that it is based on the inverse of the calibration coefficients). More importantly, with considerations of the compound effect of the reciprocity and estimation errors, we provide an in-depth analysis of the performance of a precoded massive MIMO system with the aforementioned two calibration algorithms. The proposed analytical results are verified via Monte-Carlo simulations. The rest of the paper is organised as follows.

In Section II, we describe the TDD massive MIMO system model with imperfect channel estimation and the reciprocity error model due to the RF mismatch. In Section III, the description of the relative and inverse calibration algorithms comes after a discussion of the considered calibration circuit. The performance evaluation of these two algorithms is given in Section IV. Simulation results and conclusions are provided in Section V and Section VI respectively. Some of the detailed derivations are given in the appendices.

II. SYSTEM MODEL

A TDD massive MU-MIMO system is considered in this paper as illustrated in Fig. 1. It consists of $M$ antennas at the BS, each antenna is connected with an individual RF chain. In addition, $K$ single-antenna UTs ($M \gg K$) are served in the same time and frequency resources. We assume that the time delay from the UL channel estimation to the DL transmission is less than the coherence time of the channel, ensuring that the propagation channels on the UL and DL are equal.

As shown in Fig. 1, we denote the UL and DL propagation channels by $\mathbf{H} \in \mathbb{C}^{M \times K}$ and $\mathbf{H}^T$ respectively, whose entries follow independent identically distributed (i.i.d.) $CN(0, 1)$. We consider the same model of the BS RF frontends response as that set up in [9]. Briefly speaking, $M \times M$ diagonal matrices $\mathbf{H}_{br}$ and $\mathbf{H}_{bt}$ represent the effective response matrices of the Rx and Tx RF frontends at the BS, whose $i$-th diagonal entries are $h_{br,i} = A_{br,i} \exp(j\varphi_{br,i})$ and $h_{bt,i} = A_{bt,i} \exp(j\varphi_{bt,i})$, respectively, where $A$ denotes amplitude, $\varphi$ denotes phase. We model both amplitude and phase reciprocity errors as independent truncated Gaussian random variables [9]. Without loss of generality, the statistical magnitudes of these truncated Gaussian distributed variables are assumed to be static within the considered coherence time of the channel or even longer period, e.g., minutes [12].

We then consider the UL training protocol based on the minimum mean-square error (MMSE) channel estimation as in [6], by taking into account the effect of $\mathbf{H}_{br}$ and $\mathbf{H}_{bt}$. More specifically, in TDD massive MIMO systems, UTs first transmit the orthogonal UL pilots of length $\tau_u$ to BS, where $\tau_u \geq K$. Therefore, the MMSE estimate of the actual UL channel response $\hat{\mathbf{H}}_u$ can be given by [6]

$$\hat{\mathbf{H}}_u = a \mathbf{H}_b \mathbf{R} + b \mathbf{N}_u,$$

where the estimation-error-related parameters are given by

$$a = \frac{\tau_u \rho_u}{\tau_u \rho_u + 1}, \quad b = \frac{\sqrt{\tau_u \rho_u}}{\tau_u \rho_u + 1}.$$  

In addition, the $M \times K$ noise matrix $\mathbf{N}_u$ is the channel estimation noise matrix with i.i.d. $CN(0, 1)$ elements and is independent of $\mathbf{H}_d$, and $\rho_u$ denotes the expected UL transmit SNR. Then the BS uses the transpose of $\hat{\mathbf{H}}_u$ as the estimate of the DL channel $\mathbf{H}_d$, i.e., $\hat{\mathbf{H}}_d = \mathbf{H}_d^H$, whereas the actual effective DL channel is $\mathbf{H}_d = \mathbf{H}_d^H \mathbf{H}_b$. By comparing $\hat{\mathbf{H}}_d$ and $\mathbf{H}_d$, we can rewrite the DL channel estimate $\hat{\mathbf{H}}_d$ as

$$\hat{\mathbf{H}}_d = a \mathbf{H}_b \mathbf{W} + b \mathbf{N}_u^T,$$

where $\mathbf{E} = \mathbf{H}_d^{-1} \mathbf{H}_{br}$ denotes the channel reciprocity error. We can see from (3) a compound effect of an additive distortion, $\mathbf{N}_u$, caused by the imperfect channel estimation, and a multiplicative distortion, $\mathbf{E}$, caused by the imperfect channel reciprocity.

Let a $K \times 1$ vector $\mathbf{s} = [s_1, \ldots, s_k, \ldots, s_K]^T$ denote the symbol to be transmitted to $K$ UTs, where the normalised symbol power per user is assumed, i.e., $\mathbb{E} \{ |s_k|^2 \} = 1$, for $k = 1, 2, \ldots, K$. We also assume that the symbols of different users are independent. The BS applies an $M \times K$ linear precoding matrix $\mathbf{W}$ to map the symbol vector $\mathbf{s}$ into an $M \times 1$ transmit signal vector to the BS antennas. We use $\mathbf{x}$ to denote this transmit signal vector, which is given by

$$\mathbf{x} = \sqrt{\rho_d} \lambda \mathbf{W} \mathbf{s},$$

where $\rho_d$ denotes the average transmit power at the BS (note that the expression $\mathbf{W}$ implies that the power is equally allocated to each UT in this work), and $\lambda$ is a normalisation parameter to satisfy the transmission power constraint at the BS such that

$$\mathbb{E} \{ ||\mathbf{x}||^2 \} = \mathbb{E} \{ ||\sqrt{\rho_d} \lambda \mathbf{W} \mathbf{s}||^2 \} = \rho_d.$$  

Hence, $\lambda$ can be calculated as follows:

$$\lambda = \sqrt{\frac{1}{\mathbb{E} \{ \text{tr} (\mathbf{WW}^H) \}}}.$$  

Based on (4), the received signals of all $K$ UTs can be expressed in a vector form as

$$\mathbf{y} = \mathbf{H}_d \mathbf{x} + \mathbf{n} = \sqrt{\rho_d} \lambda \mathbf{H}_d^H \mathbf{W} \mathbf{s} + \mathbf{n},$$
where the $K \times 1$ vector $n$ denotes the DL received noise for all $K$ UTs, whose $k$-th element $n_k \sim CN(0, \sigma_k^2)$. We assume that $\sigma_k^2 = 1, \forall k$. Therefore, $\rho_d$ can also be treated as the average input SNR for the DL transmission. For the $k$-th UT, we have

$$y_k = \sqrt{\rho_d} h_k^H H_{bt} w_k s_k + \sum_{i=1, i \neq k}^K \sqrt{\rho_d} h_k^H H_{bt} w_i s_i + n_k,$$

where the $M \times 1$ vectors $h_k$ and $w_k$ are the $k$-th column of $H$ and $W$ respectively. Note that the received signal $y_k$ in (8) is decomposed into three terms. The first two terms accounts for the desired signal for the $k$-th UT and the inter-user interference from other $K-1$ UTs, respectively, while the last term represents the receiver noise. Since $W$ is a function of the DL channel estimate instead of the actual DL channel response, the performance of linear precoding schemes is affected by the imperfect channel reciprocity. To address this issue, we introduce the calibration scheme in the following section.

### III. Self-Calibration

To compensate for the imperfection of channel reciprocity, an $M \times M$ pre-precoding calibration matrix $B$ can be introduced to compensate for the non-reciprocity [13], such that $H_d$ in (3) becomes

$$H_{d,CL} = a H_d E B + b N^H \epsilon B,$$

where $H_{d,CL}$ represents the estimate of the DL channel response after applying calibration, which could be used to calculate the DL precoding matrix $W$. The majority of the reported results on the reciprocity calibration has proposed the design concept of the calibration matrix $B$ without considering the effect of channel estimation error [12], [14], [15], e.g., in the case with $a \approx 1$ and $b \approx 0$. In such case, the minimum requirement to calibrate the BS antennas is that $EB = d_{cMY}$, where the scalar $c \in \mathbb{C}_{\neq 0}$ is multiplied by all calibration factors. The scalar $c$ is arbitrary. $\mathbb{C}_{\neq 0}$ denotes the set of non-zero complex numbers. Thus, this does not change the direction of the precoding beamformer [15].

The acquisition of the calibration matrix $B$ contains two steps: 1) the estimation of $H_{bt}, H_{br}$, and 2) finding $B$ based on the estimates of $H_{bt}$ and $H_{br}$. In the following, we first present a calibration circuit used for the first step. We then discuss two calibration algorithms obtained based on the considered circuit design, and their relationship with the channel estimation error.

#### A. RF Frontend Response Measurement

Motivated by the calibration method for conventional MIMO systems in [13], we present a circuit design as shown in Fig. 1 to achieve the measurement of the effective response matrix of the BS RF frontend for the massive MIMO system. Particularly, switching units attached to each antenna have three modes: “Tx/Rx” mode (the antenna connects to Tx or Rx RF frontend), “Link” mode (Tx and Rx RF frontends are connected) and “Null” mode (no connection); A reference signal source is split and equally injected at each Rx RF frontend by couplers. Then the measurement of $H_{bt}$ and $H_{br}$ can be carried out, which contains two steps which we call “self connection” and “half connection”.

During the self connection, all switching units that attached to the BS antennas are set to be in “Link” mode. The reference signal source is disconnected. The individual baseband of each BS antenna, taking $i$-th antenna element as an example, estimates the product of $h_{bt,i}$ and $h_{br,i}$ by sending a known signal $p_i$ simultaneously, such that $r_i = h_{bt,i} h_{br,i} p_i + n_i$ received at the baseband of the $i$-th antenna, where the thermal noise $n_i$ has a negligible value due to the fact that the calibration SNR is usually sufficiently high, e.g., 20 dB in [15]. Thus the estimate of $H_{bt} H_{br}$, denoted by $R_{self}$, is given by

$$R_{self} = \text{diag} (r_1/p_1, \cdots, r_i/p_i, \cdots, r_K/p_K).$$

During the half connection, all switching units that attached to the BS antennas are set to be in “Null” mode. The reference signal source is equally injected at all Rx RF frontends. Let an $M \times 1$ vector $p_{ref}$ be the reference signal vector with duplicate entries. Then the collective received signal vector at all the BS antenna baseband is given by $r_i = H_{br}(p_{ref} + u_i)$, where $r_i$ contains the received signals at each baseband that are sampled at the same time [13]. Again, the effect of the measurement noise $u_i$ is assumed to be trivial in this work due to the assumption on the high calibration SNR. Hence, the estimate of $H_{br}$, denoted by $R_{half}$, is given by

$$R_{half} = \text{diag} (r_i) (\text{diag}(p_{ref}))^{-1}. \tag{11}$$

Note that, as discussed in Section II, the reciprocity-error-related parameters, or equivalently $H_{bt}$ and $H_{br}$, are relatively static, i.e., they change in a much slower rate compared to the variations of the channel state. Hence, once the measurement of the BS RF responses, i.e., $R_{self}$ and $R_{half}$, is reliably obtained, it can be applied within the the coherence time of the channel or even longer period [12].

#### B. Design of the Calibration Matrix

Based on the measurement in (10) and (11), the calibration matrix $B$ can be calculated. The study in [13] considers the relative calibration scheme, where the calibration matrix, denoted by $B_{RC}$, is given by $B_{RC} = R_{self} (R_{half})^{-1}$. As discussed at the beginning of this section, the widely-used relative calibration ignores the effect of the imperfect channel estimation, which can result in the estimation error amplification, and additionally cause the enhancement of the inter-user interference. More specifically, it can be seen from (9) that the use of the calibration matrix $B_{RC}$ may amplify the power of the estimation noise (equivalently, channel estimation error), which can even outweigh the benefit of calibration in certain cases, such as in the low region of $\rho_d$. This motivates us to design a calibration matrix without amplifying the estimation error.

In this work, we present a calibration scheme to compensate for the effect of the channel reciprocity error, as well as to reduce the noise power of the UL channel estimation, or equivalently reduce the estimation noise variance. To this end,
we consider a calibration matrix, denoted by $B_{IC}$, which is given by
\[ B_{IC} = R_{\text{half}}^* (R_{\text{half}} R_{\text{self}}^*)^{-1}. \tag{12} \]
In the ideal scenario, e.g., in the sufficiently high calibration SNR regime, the calibration matrix $B_{IC}$ is equivalent to the inverse of the product of $H_{bt}^*$ and $H_{br}$. Thus, we name this calibration scheme as “Inverse Calibration”. Such a scheme can ensure that noise power of UL channel estimation after calibration is equal to or even less than that before calibration, since the expected value of the product of $H_{bt}^*$ and $H_{br}$ is greater than or equal to one \cite{9}. We shall evaluate the performance of the additive channel estimation error and multiplicative channel reciprocity error.

Recall that in \cite{5}, we denote the desired signal power of the $k^{th}$ UT by $P_s$, and the its inter-user interference by $P_l$, where $P_s$ and $P_l$ are given by
\[ P_s = \left| \sqrt{\rho_d} \lambda h_{bt}^* H_{bt} w_{k,b} k \right|^2, \tag{13} \]
\[ P_l = \left| \sum_{j=1,j \neq k}^K \sqrt{\rho_d} \lambda h_{bt}^* H_{bt} w_{j,b} j \right|^2, \tag{14} \]
respectively. Thus, the ergodic rate for the $k^{th}$ UT, denoted by $R_k$, can be given by
\[ R_k = \mathbb{E} \left( \log_2 \left( 1 + \frac{P_s}{P_l + \sigma_k^2} \right) \right). \tag{15} \]
Let $R_K$ denote the ergodic sum rate of all $K$ UTs. Considering the approximation derived in \cite{9, 17}, $R_K$ is given by
\[ R_K = K R_k \approx K \log_2 \left( 1 + \mathbb{E} \left( P_s \right) \mathbb{E} \left( \frac{1}{P_l + \sigma_k^2} \right) \right), \tag{16} \]
where
\[ \mathbb{E} \left( \frac{1}{P_l + \sigma_k^2} \right) = \mathbb{E} \left( \frac{1}{P_l + \sigma_k^2} \right) + O \left( \frac{\text{var}(P_l + \sigma_k^2)}{(P_l + \sigma_k^2)^2} \right). \tag{17} \]

In this paper, we take the simplest precoding algorithm \cite{5}, i.e., MRT, as an example. Recall \cite{9}, when the MRT is used at the BS, the precoding matrix $W$ can be given by
\[ W = H_{IC}^H. \tag{18} \]
Let $\lambda_{mrt}$ represent the normalisation parameter of the MRT precoding scheme, to meet the power constraint at the BS. With no calibration (NC), i.e., $B = I_M$, the $k^{th}$ UT’s output SINR (signal-to-interference-plus-noise ratio) with uncalibrated MRT precoder has been derived in \cite{9}. The analytical result of the output SINR in \cite{9} is obtained based on \cite{17}, thus, it can be used to obtain the ergodic sum rate in \cite{16}, as follows:

**Lemma 1.** Consider a TDD massive MIMO system modelled in Section II with uncalibrated MRT precoder at the BS. The closed-form expression of the ergodic sum rate for $K$ UTs, $R_{\text{NC}, mrt}$, is given by
\[ R_{\text{NC}, mrt} \approx K \log_2 \left( 1 + \frac{1}{\rho_d A_t \left( \frac{a^2 A_r ((M - 1) A_t + 2) + b^2}{a^2 A_r + b^2} \right)} \right) \times \left( \frac{K^2 + \rho_d K (K - 1) (\rho_d A_t^2 + 2 A_t)}{(\rho_d (K - 1) A_t + K)^3} \right), \tag{19} \]
where the estimation-error-related parameters $a$ and $b$ are given by \cite{2}, and the reciprocity-related parameters $A_t$, $A_r$ and $A_t$ are given in \cite{7}.

**Proof.** See the proof of Theorem 1 in \cite{9}. \hfill $\Box$

The result in \cite{19} quantifies the compound effect of the reciprocity and estimation errors on the ergodic achievable sum rate of the MRT precoded system without calibration. When the inverse calibration is applied at the BS for the MRT precoder, the precoding matrix $W_{mrt}$ can be expressed as
\[ W_{mrt}^H = \left( a H_{bt} E B_{IC} + b N_{bt}^T B_{IC} \right)^H = a H_{bt}^{-1} H^* + b H_{bt}^{-1} (H_{br})^{-1} N_{bt}^*, \tag{20} \]
Then the corresponding normalisation parameter $\lambda_{mrt}$, the expected values of the desired signal power and the interference power of the $k^{th}$ UT, i.e., $\mathbb{E} \left\{ \rho_{IC} \right\}$ and $\mathbb{E} \left\{ 1/ \left( P_{IC}^k + \sigma_k^2 \right) \right\}$, can be derived. Details are provided in Appendix B. These values are used to calculate the ergodic sum rate in the following proposition:

**Proposition 1.** Assuming that the same conditions are held as in the Lemma \cite{7} while the inverse calibration is applied at the BS. The ergodic sum rate for $K$ UTs, $R_{\text{IC}, mrt}$, is given by
\[ R_{\text{IC}, mrt} \approx K \log_2 \left( 1 + \rho_d \left( \frac{a^2 (M - 1) + b^2 E_2}{a^2 + b^2 E_2^2} \right) \right) \times \left( \frac{(K E_2)^2 + \rho_d K (K - 1) (\rho_d + 2 E_2)}{(\rho_d (K - 1) + K E_2)^3} \right), \tag{21} \]
where the reciprocity-related parameters $E_2$ and $E_2$ are given by \cite{27} and \cite{28} respectively, in Appendix A.

**Proof.** See Appendix B. \hfill $\Box$

When the traditional relative calibration is applied at the BS, the precoding matrix $W_{mrt}$ becomes
\[ W_{mrt}^H = \left( a H_{bt} E B_{IC} + b N_{bt}^T B_{IC} \right)^H = a H_{bt}^{-1} H^* + b H_{bt}^{-1} (H_{br})^{-1} N_{bt}^*. \tag{22} \]
Similar to the case of the inverse calibration, $\lambda_{mrt}$, $\mathbb{E} \left\{ \rho_{IC} \right\}$ and $\mathbb{E} \left\{ 1/ \left( P_{IC}^k + \sigma_k^2 \right) \right\}$ can be derived (see Appendix B). Then the corresponding closed-form expression of the ergodic sum rate can be given as follows:
Proposition 2. Assuming that the same conditions are held as in the Lemma 7 while the relative calibration is applied at the BS. The ergodic sum rate for $K$ UTs, $R_{K,mrt}^{RC}$, is given by

\[ R_{K,mrt}^{RC} \approx K \log_2 \left( 1 + \rho_d \left( \frac{\alpha^2((M-1)A_t^2 + 2E_t^4) + b^2 E_t^4}{\alpha^2 + b^2 E_t^2} \right) \right) \times \left( \frac{K^2 A_t^2 + \rho_d K (K-1) E_t^4 (\rho_d E_t^4 + 2A_t)}{\rho_d (K-1) E_t^4 + K A_t^4} \right)^{1/4}, \]  

where the reciprocity-error-related parameters $E_t^4$ is given by (25) in Appendix A.

Proof. See Appendix B.

Comparing (21) with (23), it can be seen that the inverse calibration completely removes phase reciprocity error, and leaves the MRT precoder with the negligible amplitude-error-related parameters, i.e., $E_t^2$ and $E_t^4$. On the contrary, the relative calibration introduces the residual amplitude error into the MRT system, i.e., $A_t^2$ and $E_t^4$ in (23). Consequently, the compound effect of the residual amplitude reciprocity error and the “amplified” channel estimation error may result in a significant performance loss in the system with the relative calibration.

V. SIMULATION RESULTS

In this section, we perform Monte-Carlo simulations to corroborate the analysis presented in Section IV and compare the performance of the calibration schemes discussed in the paper under different scenarios, in order to provide valuable insights into the practical system design. Unless otherwise specified, the statistical magnitudes of both amplitude and phase reciprocity errors are identical to that defined in [9], [18], with quadruple notations, such that $(\alpha_{br,0}, \sigma_{br}^2, [\alpha_r, b_r]) = (\alpha_{br,0}, \sigma_{br}^2, [\alpha_r, b_r]) = (0, 1, [-4 dB, 4 dB])$, and $(\theta_{br,0}, \sigma_{br}^2, [\theta_r, 1, \theta_{r,2}]) = (0, 1, [-50^\circ, 50^\circ])$, as considered in [9]. We also consider a reference scenario, namely, “Perfect Channel Reciprocity”, for the case that $\alpha_{br}^2 = \sigma_{br}^2 = \sigma_{\alpha}^2 = \sigma_{\theta}^2 = 0$. The orthogonal UL pilots are of length $\tau_u = K$, and the ergodic sum rate is measured in bits/s/Hz.

First, we investigate the ergodic sum rate of MRT with IC and RC, for different DL SNR regimes. As a benchmark, we consider perfect channel reciprocity and no calibration, NC. We use $M = 100$, $K = 10$, and reciprocity-error-related parameters as mentioned before. Fig. 2 illustrates that, for the MRT precoded system, IC nearly eliminates the effect of reciprocity error, and outperforms RC. Thus, we conclude that IC is more efficient than RC for MRT. More specifically, increasing the DL transmit power 10 times, e.g., $\rho_d$ from 0 dB to 10 dB, the performance of IC increases by 36% (i.e., 9 bits/s/Hz), while only 12% improvement (i.e., 3 bits/s/Hz) for RC. This can also be approved analytically on the comparison between (21) and (23). As mentioned earlier, both IC and RC can remove the phase reciprocity error. However, RC suffers from the strong residual amplitude error, which can cause estimation error amplification that could outweigh the benefit of using RC. We further analyse this effect of estimation error amplification in RC in Fig. 3.

In Fig. 3 we assume $M = 100$, $K = 10$, $\rho_d = 10$ dB, and a low level estimation error with $\rho_u = 10$ dB. Let $\sigma_A^2 = \sigma_{\theta}^2$ be the x-axis that captures the aggregated effect of both amplitude and phase reciprocity errors, i.e., $(\alpha_{br,0}, \sigma_{br}^2, [\alpha_r, b_r]) = (0, 1, [-4 dB, 4 dB])$, and the phase error has $(\theta_{br,0}, \sigma_{br}^2, [\theta_r, 1, \theta_{r,2}]) = (0, 1, [-50^\circ, 50^\circ])$. We consider the case with the phase error only by setting $\sigma_A^2 = 0$. Fig. 3 shows that: a) both IC and RC eliminate the phase error, with a 12.5% increase in the sum rate compared with NC (i.e., 4 bits/s/Hz more than 32 bits/s/Hz of NC at $\sigma_A^2 = 0$ and $\sigma_{\theta}^2 = 0.2$); b) the performance gain of RC over NC decreases by 50% in the case with the aggregated effect
of the amplitude and phase error (i.e., only 2 bits/s/Hz improvement compared to NC at $\sigma^2_A = \sigma^2_P = 0.2$, which is 50% of the previous 4 bits/s/Hz improvement in (a)), whereas the gain of IC over NC increases by 50% in this case (i.e., around 6 bits/s/Hz improvement compared to NC at $\sigma^2_A = \sigma^2_P = 0.2$); c) RC may not work properly in the presence of the compound effect of both reciprocity and estimation errors, see $\sigma^2_A$, $\sigma^2_P < 0.02$. We now take a closer look at the third observation, c, as follows.

Similar parameters are considered as in Fig. 3. In addition, higher levels of channel estimation error are introduced, e.g., $\rho_u = 0$ dB in Fig. 4(a) and $\rho_u = -5$ dB in Fig. 4(b). It can be seen that the gain of the relative calibration vanishes in the case of the severe estimation error, whereas the inverse calibration, due to its greater robustness to the compound effect of the reciprocity error and estimation error, still works effectively with only a minor performance degradation.

VI. CONCLUSION

In this paper, we have presented and investigated two calibration schemes, i.e., inverse calibration and relative calibration, in the TDD massive MIMO system. A low-cost calibration circuit has been considered, which enables the BS to select these two calibration algorithms. The performance of both calibration algorithms has been evaluated, by taking into account the compound effect of the multiplicative reciprocity error and the additive estimation error. Particularly, we have derived closed-form expressions for the ergodic sum rate of the considered system, assuming MRT with the compound effect of both errors. We have demonstrated that the inverse calibration in general outperforms the traditional relative calibration. Analytical results perfectly match the simulated results for different scenarios, including large or practical number of BS antennas, different number of UTs and different combinations of the reciprocity error and estimation error. The comprehensive performance analysis has been given to provide a useful guidance for the selection of the calibration schemes in the massive MIMO system, which inevitable in practice.

For the way forward, one can consider the application of the inverse calibration in different scenarios, e.g., with different precoding schemes. It can also be of interest to carry out a field trial of the presented calibration circuits and algorithms, thanks to the simplicity of their design principles.

APPENDIX A

Recall the probability density function of a truncated Gaussian distributed variable $x \sim N(l, \sigma^2)$, $x \in [a, b]$ (also $\alpha, \beta$ and $Z$) in (9), the $l$th $(l \geq 0)$ non-central moment of $x$, denoted by $E^l_x$, is given by

$$E^l_x = \mathbb{E}\{x^l\} = \sum_{i=0}^{l} \binom{l}{i} \sigma^i \mu^{l-i} L_i. \quad (24)$$

We can now calculate the $l$th $(l \geq 0)$ non-central moment of a truncated Gaussian distributed variable, such as

$$E^l_4 = \mathbb{E}\{A^4_{bt,i}\} \quad \sigma^2 = 4 \alpha^2 \beta^2 \sigma^2 + 4 \alpha \beta \sigma^4$$

To calculate the expected value of the inverse square of the truncated Gaussian distributed variable $x$, denoted by $E^{-2}_x$ where the subscript $(\cdot)_2$ is used for the 2-nd inverse moment of $x$, we have

$$E^{-2}_x = \mathbb{E}\left\{\frac{1}{x^2}\right\} = \int_a^b \frac{1}{x^2} \frac{1}{\sigma Z} \phi\left(\frac{x - \mu}{\sigma}\right) dx$$

$$= \frac{1}{\sqrt{2\pi} \sigma Z} \int_a^b x^{-2} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx. \quad (26)$$

The integral in (26) is a nonelementary antiderivative and not able to be further simplified when $\mu \neq 0$. Consider the particular case in this work, the value of $E^{-2}_x$ can be measured based on the measurement of the response of the BS RF frontends, i.e., $R_{self}$ and $R_{half}$, such that

$$E^{-2}_x = \frac{1}{M} \text{tr}\left(R_{half}^{-1}(R_{half}^*)^{-1}\right). \quad (27)$$

Similarly, we have

$$E^{-2}_2 = \frac{1}{M} \text{tr}\left(R_{self}^2(R_{half}^*)^{-1}\right). \quad (28)$$

APPENDIX B

A. Inverse Calibration

1. $\lambda_{IC}^M$. Recall (3) and (20), the denominator inside of the square root in $\lambda_{IC}^M$ can be given by

$$\mathbb{E}\left\{\text{tr}\left(W_{IC}^M(W_{IC}^M)^H\right)\right\} = \mathbb{E}\left\{\text{tr}\left(a^2 H_{bt}^1 H^* H^T(H_{bt}^* H_{bt})^{-1}\right)\right\}$$

$$+ \mathbb{E}\left\{\text{tr}\left(b^2 (H_{br} H_{br})^{-1} N_{br}^T (H_{br}^* H_{br})^{-1}\right)\right\}$$

$$= ME_2 \left(\sigma^2 + b^2 E_2^*\right). \quad (29)$$

$$= ME_2 \left(\sigma^2 + b^2 E_2^*\right). \quad (30)$$
where (a) is conditioned on the independence between $H$ and $N_u$. Thus we have

$$\lambda_{\text{mrt}}^{\text{IC}} = \sqrt{\frac{1}{MKE_2^2(a^2 + b^2E_2^2)}}. \quad \text{(31)}$$

(2) $E \{ P_{\text{s,mrt}}^{\text{IC}} \}$: Recall (13) and we have $E \{ |s_k|^2 \} = 1$, we have

$$E \{ P_{\text{s,mrt}}^{\text{IC}} \} = E \left\{ \sqrt{\rho_d \lambda_{\text{mrt}}^{\text{IC}}}H_k^T H_{bt} w_{\text{k,mrt}}^{\text{IC}} \right\}, \quad \text{(32)}$$

where

$$E \{ H_k^T H_{bt} w_{\text{k,mrt}}^{\text{IC}} \}^2 = E \left\{ |H_k^T H_{bt}(aH_{bt}^{-1}h_k^T + b(H_{bt}H_{bt}^*)^{-1}n_{a,k})|^2 \right\} \quad \text{(33)}$$

$$= a^2 E \left\{ |h_k^T h_k^*|^2 \right\} + b^2 E \left\{ |H_{bt}^T (H_{bt}^*)^{-1}n_{a,k}|^2 \right\} \quad \text{(34)}$$

$$= a^2 M(M+1) + b^2 ME_2^2. \quad \text{(35)}$$

Substituting (31) and (35) into (32), we have

$$E \{ P_{\text{s,mrt}}^{\text{IC}} \} = \frac{\rho_d}{K} \left( a^2(M+1) + b^2E_2^2 \right)^2. \quad \text{(36)}$$

(3) $E \left\{ 1/(P_{\text{l,mrt}}^{\text{IC}} + \sigma_k^2) \right\}$: Consider that the symbols of different users are independent, we first calculate $E \{ P_{\text{l,mrt}}^{\text{IC}} \}$ as

$$E \{ P_{\text{l,mrt}}^{\text{IC}} \} = E \left\{ \sum_{i=1,i\neq k}^K \sqrt{\rho_d \lambda_{\text{mrt}}^{\text{IC}}}H_k^T H_{bt} w_{\text{k,mrt}}^{\text{IC}} \right\}^2 \quad \text{(37)}$$

$$= \sum_{i=1,i\neq k}^K \left\{ \sqrt{\rho_d \lambda_{\text{mrt}}^{\text{IC}}}H_k^T H_{bt} w_{\text{k,mrt}}^{\text{IC}} \right\}^2 \quad \text{(38)}$$

$$\overset{(b)}{=} \frac{\rho_d(K-1)}{KE_2^2}, \quad \text{(39)}$$

where (b) can be carried out in the similar way as that from (32) to (36). We then calculate $\text{var}(P_{\text{l,mrt}}^{\text{IC}} + \sigma_k^2)$ by following the technique as in [9] Theorem 1, such that

$$\text{var}(P_{\text{l,mrt}}^{\text{IC}} + \sigma_k^2) = \frac{\rho_d^2(K-1)}{KE_2^4}. \quad \text{(40)}$$

Substituting (39) and (40) into (17), we have

$$E \left\{ \frac{1}{P_{\text{l,mrt}}^{\text{IC}} + \sigma_k^2} \right\} = \frac{KE_2^2 \left( (KE_2^2)^2 + \rho_dK(K-1)(\rho_d + 2E_2^2) \right)}{(\rho_d(K-1) + KE_2^2)^3}. \quad \text{(41)}$$

Now we can arrive at (21) in Proposition 1 by substituting (36) and (41) in (16).

**B. Relative Calibration**

For the sake of simplicity, we list the main results for the relative calibration scheme as below. The derivation of these results follows the similar technique as in the previous subsection.

$$\lambda_{\text{mrt}}^{\text{RC}} = \sqrt{\frac{1}{MA_1(a^2 + b^2E_2^2)}}, \quad \text{(42)}$$