Reciprocity Calibration for Massive MIMO: Proposal, Modeling and Validation

Vieira, Joao; Rusek, Fredrik; Edfors, Ove; Malkowsky, Steffen; Liu, Liang; Tufvesson, Fredrik

Published in:
IEEE Transactions on Wireless Communications

DOI:
10.1109/TWC.2017.2674659

2017

Document Version:
Peer reviewed version (aka post-print)

Reciprocity Calibration for Massive MIMO: Proposal, Modeling and Validation

Joao Vieira, Fredrik Rusek, Ove Edfors, Steffen Malkowsky, Liang Liu, Fredrik Tufvesson
Dept. of Electrical and Information Technology, Lund University, Sweden
firstname.lastname@eit.lth.se

***Abstract*—This paper presents a mutual coupling based calibration method for time-division-duplex massive MIMO systems, which enables downlink precoding based on uplink channel estimates. The entire calibration procedure is carried out solely at the base station (BS) side by sounding all BS antenna pairs. An Expectation-Maximization (EM) algorithm is derived, which processes the measured channels in order to estimate calibration coefficients. The EM algorithm outperforms current state-of-the-art narrow-band calibration schemes in a mean squared error (MSE) and sum-rate capacity sense. Like its predecessors, the EM algorithm is general in the sense that it is not only suitable to calibrate a co-located massive MIMO BS, but also very suitable for calibrating multiple BSs in distributed MIMO systems.**

The proposed method is validated with experimental evidence obtained from a massive MIMO testbed. In addition, we address the estimated narrow-band calibration coefficients as a stochastic process across frequency, and study the subspace of this process based on measurement data. With the insights of this study, we propose an estimator which exploits the structure of the process in order to reduce the calibration error across frequency. A model for the calibration error is also proposed based on the asymptotic properties of the estimator, and is validated with measurement results.

***Index Terms*—Massive MIMO, reciprocity calibration, mutual coupling, Expectation Maximization, validation, calibration error.**

I. Introduction

MASSIVE Multiple-input Multiple-output (massive MIMO) is an emerging technology with the potential to be included in next generation wireless systems, such as fifth-generation (5G) cellular systems. Massive MIMO departs from traditional multi-user MIMO approaches by operating with a large number of base station (BS) antennas, typically in the order of hundreds or even thousands, to serve a relatively small number of mobile terminals [1]. Such a system setup results in a multitude of BS antennas that can be used in an advantageous manner from multiple points of view [2].

One major challenge of operating with a large number BS antennas is that it renders explicit channel estimation in the downlink impractical. Basically, the overhead of channel estimation in the downlink and feeding back the channel estimate to the BS, scales linearly with the number of BS antennas, and quickly becomes unsupportable in mobile time-varying channels [3]. To deal with this challenge, the approach adopted is to operate in time-division-duplex (TDD) mode, rely on channel reciprocity, and use uplink channel state information (CSI) for downlink precoding purposes [4]. However, the presence of the analog front-end circuitry in practical radio units complicates the situation and makes the baseband-to-baseband channel non-reciprocal. Explained briefly, the baseband representation of the received signals [5] experience channels that are not only determined by the propagation conditions, but also by the transceiver front-ends at both sides of the radio link. While it is generally agreed that the propagation channel is reciprocal [6], the transceiver radio frequency (RF) chains at both ends of the link are generally not [7]. Hence, in order to make use of the reciprocity assumption and rely on the uplink CSI to compute precoding coefficients, the non-reciprocal transceiver responses need to be calibrated. Such a procedure is often termed reciprocity calibration, and contains two steps: (i) estimation of calibration coefficients, and (ii) compensation by applying those to the uplink channel estimates.[1]

Reciprocity calibration of small scale TDD MIMO channels has been a matter of study in recent years. Depending on the system setup and requirements, the approach adopted can take many forms. For example, [7] proposed a methodology based on bi-directional measurements between the two ends of a MIMO link to estimate suitable reciprocity calibration coefficients. This calibration approach falls in the class of "over-the-air" calibration schemes where users are involved in the calibration process. A different approach is to rely on dedicated hardware circuitry for calibration purposes, see [8], [9]. Despite the possibilities of extending both mentioned calibration approaches to a massive MIMO context, e.g., [10], [11], recent calibration works suggest this is more difficult than previously thought. For example, [12] questions the feasibility of having dedicated circuits for calibration when the number of transceivers to be calibrated grows large, and [13] argues that the calibration protocols should preferably not rely on mobile units. It thus appears that an increasing trend in massive MIMO systems is to carry out the calibration entirely at the BS side only through over-the-air measurements.

The first proposal in this vein was presented in [14]. The work proposes an estimator for the calibration coefficients, which only makes use of channel measurements between BS antennas. More specifically, bi-directional channel measurements between a given BS antenna, so-called reference antenna, and all other antennas. This estimator was later generalized in order to calibrate large-scale distributed MIMO networks [13], [15]. The estimation problem is formulated as

[1]However, with the term reciprocity calibration, we will interchangeably refer to the estimation step, compensation step, or both. The context will, hopefully, make clear which of the previous cases is being addressed.

constrained least-squares (LS) problem where the objective function uses channel measurements from a set of arbitrary antenna pairs of the network. The generality of this approach spurred many publications dealing with particular cases [16]–[18]. Parallel work in mutual coupling based calibration was also conducted in [12]. An estimator for the calibration coefficients, which enables maximum ratio transmission (MRT), was proposed for BS antenna arrays with special properties.

Although it appears that over-the-air reciprocity calibration only involving the BS side is feasible, some matters need further investigation. Firstly, the approaches available in the literature for co-located BSs are not of great practical convenience. They either rely on antenna elements that need to be (carefully) placed in front of the BS antenna array solely for calibration purposes [14], or are only available for a restrictive case of antenna arrays [12]. Secondly, most estimators for calibration have been derived from empirical standpoints, e.g., [12], [14], and respective extensions [15], [17], [18]. It is not clear how far from fundamental estimation performance bounds, or how close to Maximum likelihood (ML) performance, such estimators are. Thirdly, most available calibration approaches are proposed for narrow-band systems. Such systems bandwidths are usually defined by the frequency selectivity of the propagation channel, which is typically much smaller than the frequency selectivity of the transceiver responses. This results in similar calibration coefficients for adjacent narrowband channels. Thus, it is of interest to model the statistical dependency of such calibration coefficients, and provide means to exploit this dependency in order to reduce the calibration error across frequency. Lastly, there is little publicly available work on validation of massive MIMO calibration schemes. The need for validation is high, as it helps answering many questions of practical nature. For example, [19] raises the question whether the channel reciprocity assumption holds when strong coupling between BS antennas exist, and [20] questions if calibration assumptions similar to the ones used in this work, hold for massive MIMO arrays.

A. Main Contributions of the Paper

Below, we summarize the main contributions of this work.

- We propose a convenient calibration method mainly relying on mutual coupling between BS antennas to calibrate its non-reciprocal analog front-ends. We make no assumptions other than channels due to mutual coupling being reciprocal.
- We show that the narrow-band calibration coefficients can be estimated by solving a joint penalized-ML estimation problem. We provide an asymptotically efficient algorithm to compute the joint solution, which is a particular case of the EM algorithm.
- We validate our calibration method experimentally using a software-defined radio massive MIMO testbed. More specifically, we verify how the measured Error-Vector-Magnitude (EVM) of the downlink equalized signals decreases as the calibration accuracy increases, in a setup where three closely spaced single-antenna users are spatially multiplexed by one hundred BS antennas.
- We propose a non-white Gaussian model for the narrow-band calibration error based on the properties of the proposed estimator, and partially validate this model with measurements.

B. Notation

The operators $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^\dagger$ denote element-wise complex conjugate, transpose, Hermitian transpose, and Moore-Penrose pseudo-inverse, respectively. The element in the nth row and mth column of matrix $\mathbf{A}$ is denoted by $\left[\mathbf{A}\right]_{n,m}$. The operator $\mathrm{E}\{\cdot\}$ denotes the expected value. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ return the real and imaginary part of their arguments. The matrix $\mathbf{I}$ denotes the identity matrix, and $\mathrm{diag}\{a_1, a_2, \ldots a_M\}$ denotes an $M \times M$ diagonal matrix with diagonal entries given by $a_1, a_2, \ldots, a_M$. The operator $\ln$ denotes the natural logarithm. The set of the complex numbers and the set containing zero and the real positive numbers are denoted by $\mathbb{C}$ and $\mathbb{R}_{\geq 0}$, respectively. The operator $\setminus$ denotes the relative set complement. Finally, $||\cdot||$ denotes the Frobenius norm.

C. Paper Outline

The remaining sections of the paper are as follows. Section II presents the signal models. Section III introduces the state-of-the-art estimator for the calibration coefficients, proposes a novel estimator, and provides a comparative analysis by means of MSE and downlink sum-rate capacities. Section IV validates the proposed calibration method experimentally. Using the estimated calibration coefficients obtained from the experiments, the purpose of Section V is twofold: *i)* it studies several aspects of the calibration coefficients across 4.5 MHz of transceiver bandwidth, *ii)* it proposes a model for the calibration error of a narrowband system. Lastly, Section VI summarizes the key takeaways from this work.

II. Signal Models

This section starts by introducing the uplink and downlink signal models, and shows how downlink precoding can be performed using calibrated uplink channel estimates. Finally, it models the channels between BS antennas which we use for calibration purposes.

A. Uplink and Downlink Signal models

Let K single-antenna users simultaneously transmit a pilot symbol in the uplink of a narrow-band MIMO system (e.g., a particular sub-carrier of an OFDM-MIMO system). Collecting the pilot symbols in the vector $\mathbf{p} = [p_1 \cdots p_K]^T$, the received signal by an M-antenna base station can be written as

$$\begin{aligned}\mathbf{y}_{\mathrm{UP}} &= \mathbf{H}_{\mathrm{UP}}\,\mathbf{p} + \mathbf{w} \\ &= \mathbf{R}_{\mathrm{B}}\mathbf{H}_{\mathrm{P}}\mathbf{T}_{\mathrm{U}}\,\mathbf{p} + \mathbf{w}. \end{aligned} \tag{1}$$

In (1), the matrix $\mathbf{R}_{\mathrm{B}} = \mathrm{diag}\left\{r_1^{\mathrm{B}}, \cdots, r_M^{\mathrm{B}}\right\}$ models the hardware response of M BS receive RF chains (one RF chain per antenna), and the matrix $\mathbf{T}_{\mathrm{U}} = \mathrm{diag}\left\{t_1^{\mathrm{U}}, \cdots, t_K^{\mathrm{U}}\right\}$ models the hardware response of K transmit RF chains (one chain per

user). $\mathbf{H}_{\mathrm{P}}$ is the propagation channel matrix, $\mathbf{H}_{\mathrm{UP}}$ is the, so-called, uplink radio channel, and $\mathbf{w}$ is a vector modeling uplink noise. Under the reciprocal assumption of the propagation channel, the received downlink signal can be written as

$$\begin{aligned}\mathbf{y}_{\mathrm{DL}} =&\mathbf{H}_{\mathrm{DL}}\,\mathbf{z}' + \mathbf{w}' \\ =&\mathbf{R}_{\mathrm{U}}\mathbf{H}_{\mathrm{P}}^{T}\mathbf{T}_{\mathrm{B}}\,\mathbf{z}' + \mathbf{w}'. \end{aligned} \tag{2}$$

In (2), the matrix $\mathbf{R}_{\mathrm{U}} = \mathrm{diag}\left\{r_1^{\mathrm{U}}, \cdots, r_K^{\mathrm{U}}\right\}$ models the hardware response of the receive RF chains of the K users, and the matrix $\mathbf{T}_{\mathrm{B}} = \mathrm{diag}\left\{t_1^{\mathrm{B}}, \cdots, t_M^{\mathrm{B}}\right\}$ models the hardware response of M BS transmit RF chains. The entries of $\mathbf{w}'$ model downlink noise, $\mathbf{H}_{\mathrm{DL}}$ is the downlink radio channel, and $\mathbf{z}'$ is a vector with linearly precoded QAM symbols. In particular, $\mathbf{z}' = \mathbf{P}\mathbf{x}$, where $\mathbf{P}$ is the precoding matrix, and the entries of $\mathbf{x}$ contain QAM symbols.

B. Calibration Coefficients

Assume that an error free version of the uplink radio channel, $\mathbf{H}_{\mathrm{UP}}$, is available at the BS. The transpose of the result of pre-multiplying $\mathbf{H}_{\mathrm{UP}}$ with the matrix $\alpha\mathbf{T}_{\mathrm{B}}\mathbf{R}_{\mathrm{B}}^{-1}$, where $\alpha \in \mathbb{C} \setminus 0$ and $r_m \neq 0, \forall\, m$, is a matrix $\mathbf{G}$ that, if used for precoding purposes by means of a linear filtering, is sufficient for spatially multiplexing terminals in the downlink with reduced crosstalk. This can be visualized by expanding $\mathbf{G}$ as

$$\begin{aligned}\mathbf{G} &= \left(\left(\alpha\mathbf{T}_{\mathrm{B}}\mathbf{R}_{\mathrm{B}}^{-1}\right)\mathbf{H}_{\mathrm{UP}}\right)^{T} \\ &= \alpha\mathbf{T}_{\mathrm{U}}\mathbf{H}_{\mathrm{P}}^{T}\mathbf{T}_{\mathrm{B}} \\ &= \alpha\mathbf{T}_{\mathrm{U}}\mathbf{R}_{\mathrm{U}}^{-1}\mathbf{H}_{\mathrm{DL}}. \end{aligned} \tag{3}$$

From (3) we have that $\mathbf{G}$ is effectively the *true* downlink radio channel $\mathbf{H}_{\mathrm{DL}}$ pre-multiplied with a diagonal matrix with unknown entries accounting for the user terminals responses $\mathbf{T}_{\mathrm{U}}\mathbf{R}_{\mathrm{U}}^{-1}$, and α. The row space of $\mathbf{G}$ is thus the same as of the downlink radio channel $\mathbf{H}_{\mathrm{DL}}$. This is a sufficient condition to cancel inter-user interference if, for example, ZF precoding is used (i.e., $\mathbf{H}_{\mathrm{DL}}\mathbf{G}^{\dagger}$ is a diagonal matrix).

From (3), it can also be seen that any non-zero complex scalar α provides equally good calibration.[2] Thus, the matrix

$$\begin{aligned}\mathbf{C} =&\mathrm{diag}\{\mathrm{c}_1, \cdots, \mathrm{c}_{\mathrm{M}}\} \\ =&\mathbf{T}_{\mathrm{B}}\mathbf{R}_{\mathrm{B}}^{-1} \end{aligned} \tag{4}$$

is the, so-called, calibration matrix, and $\{c_m\}$ are the calibration coefficients which can be estimated up to a common complex scalar α. We remark that, although not strictly necessary to build estimators, the concept of a reference transceiver [14] can be used to deal with the ambiguity of estimating $\{c_m\}$ up to α.[3] The remainder of the paper deals with estimation aspects of $c_m = t_m^{\mathrm{B}}/r_m^{\mathrm{B}}$. Thus, for notational simplicity, we write $t_m = t_m^{\mathrm{B}}$, $r_m = r_m^{\mathrm{B}}$, $\mathbf{R} = \mathbf{R}_{\mathrm{B}}$, and $\mathbf{T} = \mathbf{T}_{\mathrm{B}}$. Also, we stack $\{c_m\}$ in the vector $\mathbf{c} = [c_1 \cdots c_M]^T$, for later use.

[2]This follows since both magnitude and phase of α are not relevant in this calibration setup. The former holds since any real scaled channel estimate provides the same precoder matrix $\mathbf{P}$, if the precoder has a fixed norm. The latter follows from (3), since the (uniform phases of the) diagonal entries of $\mathbf{T}_{\mathrm{U}}\mathbf{R}_{\mathrm{U}}^{-1}$ are unknown to the precoder in this calibration setup.

[3]Explained briefly, assuming $c_{ref} = 1$ and solving for $\{c_m\} \setminus c_{ref}$, where c_{ref} is the calibration coefficient associated with a reference transceiver.

C. Inter-BS Antennas Signal model

To estimate the calibration coefficients c_m we sound the M antennas one-by-one by transmitting a sounding signal from each one and receiving on the other $M-1$ silent antennas. Let the sounding signal transmitted by antenna m be $s_m = 1, \forall\, m$, unless explicitly said otherwise. Also, let $y_{n,m}$ denote the signal received at antenna n when transmitting at antenna m. It follows that the received signals between any pair of antennas can be written as

$$\begin{bmatrix} y_{n,m} \\ y_{m,n} \end{bmatrix} = h_{n,m} \begin{bmatrix} r_n t_m & 0 \\ 0 & r_m t_n \end{bmatrix} \begin{bmatrix} s_m \\ s_n \end{bmatrix} + \begin{bmatrix} n_{n,m} \\ n_{m,n} \end{bmatrix}, \tag{5}$$

where

$$h_{n,m} = \bar{h}_{n,m} + \tilde{h}_{n,m} \tag{6}$$

$$= |\bar{h}_{n,m}| \exp(j2\pi\phi_{n,m}) + \tilde{h}_{n,m} \tag{7}$$

models the (reciprocal) channels between BS antennas. The first term $\bar{h}_{n,m}$ describes a channel component due to mutual coupling between antenna elements, often stronger for closely spaced antennas, which we lay down a model for in Sec. II-D. The terms $|\bar{h}_{n,m}|$ and $\phi_{n,m}$ denote the magnitude and phase of $\bar{h}_{n,m}$, respectively. The term $\tilde{h}_{n,m}$, which absorbs all other channel multipath contributions except for the mutual coupling (e.g., reflections by scatterers in front of the BS) is modeled by an i.i.d. zero-mean circularly symmetric complex Gaussian random variable with variance σ^2. Non-reciprocal channel components are modeled by r_m and t_m which materially map to the cascade of hardware components, mainly in the analog front-end stage of the receiver and transmitter, respectively. We assume i.i.d. circularly symmetric zero-mean complex Gaussian noise contributions $n_{m,n}$ with variance N_0. Letting $\left[\mathbf{Y}\right]_{m,n} = y_{m,n}$, the received signals can be expressed more compactly as

$$\mathbf{Y} = \mathbf{R}\mathbf{H}\mathbf{T} + \mathbf{N}. \tag{8}$$

Note that $\mathbf{H} = \mathbf{H}^T$ is assumed, and the diagonal entries in the $M \times M$ matrix $\mathbf{Y}$ are undefined.

D. Modeling Mutual Coupling

The purpose of this section is to provide a model for the mutual coupling between antenna elements, i.e. $\bar{h}_{m,n}$, as a function of their distance. Instead of pursuing a circuit theory based approach to model the effect of mutual coupling [19], our modeling approach uses S-parameter measurements from a massive MIMO BS antenna array [21]. We note that this model is used only for simulation purposes, and not to derive any of the upcoming estimators of $\mathbf{c}$.

1) Test Array Description: The antenna array considered for modeling is a 2-dimensional planar structure with dual-polarized patch elements spaced by half a wavelength. More information about the antenna array can be found in [22]. The dimensional layout of the array adopted for this work corresponds to the 4×25 rectangular grid in the upper part of the array shown in Fig. 1. Only one antenna port is used per antenna element. For a given antenna, the polarization port is chosen such that its adjacent antennas - the antennas spaced by half wavelength - are cross-polarized. This setting provides,

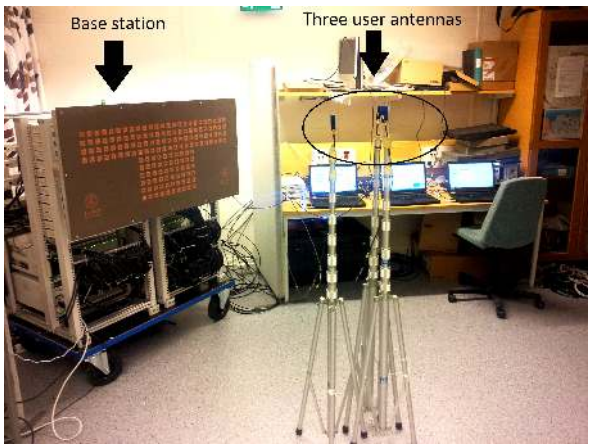


Fig. 1. The massive MIMO lab setup used throughout this work. The BS is on the left side where a "T" shaped antenna array can be seen. Three closely spaced user antennas stand the middle of the picture.

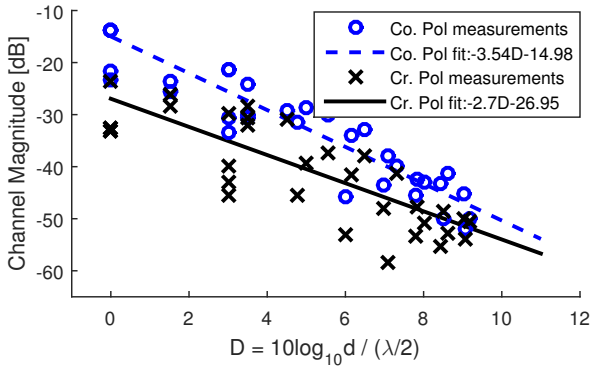


Fig. 2. Measured coupling magnitudes $|\bar{h}_{n,m}|$ between different antenna pairs. The circles corresponds to measurements between co-polarized antenna elements, and the crosses between cross polarized antenna elements. The variable d corresponds to the physical distance between antenna elements. The straight lines represent the corresponding linear LS fits.

so-called, polarization diversity, and reduces mutual coupling effects between adjacent antennas since co-polarized antennas couple stronger [21].

2) Modeling coupling gains between antennas: The channel magnitude $|\bar{h}_{n,m}|$ between several pairs of cross and co-polarized antennas were measured in an anechoic chamber using a Vector Network Analyzer, at 3.7 GHz - the center frequency of the array. Fig. 2 shows the measured channel magnitudes. Different channel magnitudes for the very same measured distance and polarization cases, are due mostly to the relative orientation of the antenna pair with respect to their polarization setup. For example, vertically (co-)polarized antennas couple more strongly when they are oriented horizontally. A linear LS fit was performed to model the coupling gain $|\bar{h}_{n,m}|$ as a function of antenna distance. The phase $\phi_{m,n} = \phi_{n,m}$ is modeled uniformly in $[0, 1]$, as a clear dependence with distance was not found.

III. Estimation of the Calibration Coefficients

In this section we deal with estimation aspects of the calibration matrix $\mathbf{C} = \mathbf{T}\mathbf{R}^{-1}$. We introduce the state-of-art estimator of $\mathbf{C}$ [13], [15], and propose a novel iterative penalized-ML estimator.[4] A comparative numerical analysis is made by means of MSE and sum-rate capacity. We conclude the section with two interesting remarks.

A. The Generalized Method of Moments estimator

Calibration of large-scale distributed MIMO systems using a similar system model to (8) was performed in [13] and [15].[5] Based on the structure of the system model, the authors identified that

$$\mathrm{E}\left\{y_{n,m}c_n - y_{m,n}c_m\right\} = 0. \tag{9}$$

Define $g_{m,n} \triangleq y_{n,m}c_n - y_{m,n}c_m$, and $\boldsymbol{g}(\mathbf{c}) = [g_{1,2} \cdots g_{1,M}\, g_{2,3} \cdots g_{2,M} \cdots g_{M-1,M}]^T$.[6] An estimator for $\mathbf{c}$ was proposed by solving

$$\hat{\boldsymbol{c}}_{\mathrm{GMM}} = \arg \min_{\substack{\boldsymbol{c} \\ s.t.\ f_c(\boldsymbol{c})=1}} \boldsymbol{g}^H(\mathbf{c})\boldsymbol{W}\boldsymbol{g}(\mathbf{c}) \tag{10}$$

with $\mathbf{W} = \mathbf{I}$. Two constraints were suggested to avoid the all-zero solution, namely $f_c(\boldsymbol{c}) = c_1$ or $f_c(\boldsymbol{c}) = ||\mathbf{c}||^2$. By setting the gradient with respect to $\mathbf{c}$ to zero, an estimator in closed-form was given. Next, we provide a few remarks on this estimation approach.

A fact not identified in [13] and [15], is that this estimator is an instance of a estimation framework widely used for statistical inference in econometrics, namely the generalized method of moments (GMM). The variable $g_{m,n}$ - whose expectation is zero - is termed a moment condition within GMM literature [23]. With a proper setting of the weighting matrix $\mathbf{W}$, it can be shown that the solution to (10) provides an estimator that is asymptotically efficient [23]. However, no such claim can be made in the low signal-to-noise (SNR) regime, where an optimal form of $\mathbf{W}$ is not available in the literature. This typically leads to empirical settings of $\mathbf{W}$, e.g., $\mathbf{W} = \mathbf{I}$. As a result, moment conditions comprising measurements with low SNR constrain the performance since they are weighted equally. It thus appears that an inherent problem of the GMM estimator is the selection of $\mathbf{W}$. Nevertheless, it provides a closed-form estimator based on a cost function where nuisance parameters for calibration, as $h_{m,n}$, are conveniently left out.

B. Joint Maximum Penalized-Likelihood estimation

Here we address joint maximum penalized-likelihood estimation for $\mathbf{c}$ and for the equivalent channel $\boldsymbol{\Psi} \triangleq \mathbf{RHR}$. Noting that (8) can be written as

$$\begin{aligned} \mathbf{Y} &= \mathbf{RHRC} + \mathbf{N} \\ &= \boldsymbol{\Psi}\mathbf{C} + \mathbf{N}, \end{aligned} \tag{11}$$

[4]We note that the only assumption used to derive the estimators is $\mathbf{H} = \mathbf{H}^T$. The generality of this assumption allows the estimators to be used in other calibration setups than those of co-located MIMO systems, as it will be pointed out later.

[5]In their work, $h_{m,n}$ denotes the propagation channel between antennas of different BSs. The reciprocal model adopted for $h_{m,n}$ accounts for large-scale and small-scale fading.

[6]The dependency of $\boldsymbol{g}(\mathbf{c})$ on $y_{n,m}$ is explicitly left out, for notational convenience.

the optimization problem can be put as

$$[\hat{\mathbf{c}},\ \hat{\mathbf{\Psi}}] = \arg\max_{\mathbf{c},\mathbf{\Psi}} \ln p(\mathbf{Y}|\mathbf{C},\mathbf{\Psi}) + \mathrm{Pen}(\mathbf{C},\mathbf{\Psi},\epsilon') = \arg\min_{\mathbf{c},\Psi} J_{\mathrm{ML}}(\mathbf{Y},\mathbf{C},\mathbf{\Psi},\epsilon) \quad (12)$$

with $J_{\mathrm{ML}}(\mathbf{Y},\mathbf{C},\mathbf{\Psi},\epsilon) = ||\mathbf{Y}-\mathbf{\Psi}\mathbf{C}||^2 + \mathrm{Pen}(\mathbf{C},\mathbf{\Psi},\epsilon)$. Here, $p(\mathbf{Y}|\mathbf{C},\mathbf{\Psi})$ denotes the probability density function (PDF) of $\mathbf{Y}$ conditioned on $\mathbf{C}$ and $\mathbf{\Psi}$, and $\mathrm{Pen}(\mathbf{C},\mathbf{\Psi},\epsilon)$ is a penalty term parametrized by $\epsilon = \epsilon' N_0$ with $\epsilon \in \mathbb{R}_{\geq 0}$.

There are many uses for the penalty term in ML formulations [24]. Here, we use it mainly to control the convergence rate of the algorithm (presented in Sec. III-C), and use ϵ as a tuning parameter. With this in mind, we pursue Ridge Regression and set the penalty term as[7]

$$\mathrm{Pen}(\mathbf{C},\mathbf{\Psi},\epsilon) = \epsilon(||\mathbf{C}||^2 + ||\mathbf{\Psi}||^2). \quad (13)$$

After some re-modeling, a vectorized version of (11) can be written as

$$\widetilde{\mathbf{Y}} = \mathbf{\Psi}_{\mathrm{eq}}(\tilde{\mathbf{\Psi}})\mathbf{c} + \widetilde{\mathbf{N}}, \quad (14)$$

or as

$$\mathbf{Y}' = \mathbf{C}_{\mathrm{eq}}(\mathbf{c})\tilde{\mathbf{\Psi}} + \mathbf{N}', \quad (15)$$

where $\tilde{\mathbf{\Psi}}$ stacks all $\psi_{n,m} = [\mathbf{\Psi}]_{n,m}$ into an $(M^2-M)/2 \times 1$ vector, and $\mathbf{\Psi}_{\mathrm{eq}}(\tilde{\mathbf{\Psi}})$ and $\mathbf{C}_{\mathrm{eq}}(\mathbf{c})$ are equivalent observation matrices which are constructed from $\tilde{\mathbf{\Psi}}$ and $\mathbf{c}$, respectively. The structure of these matrices is shown in Appendix A, but it can be pointed out that $\mathbf{\Psi}_{\mathrm{eq}}(\tilde{\mathbf{\Psi}})$ and $\mathbf{C}_{\mathrm{eq}}(\mathbf{c})$ are a block diagonal, where each block is a column vector.

From (15), it is seen that for a given $\mathbf{C}_{\mathrm{eq}}(\mathbf{c})$, the penalized-ML estimator of $\tilde{\mathbf{\Psi}}$ is given by[8]

$$\tilde{\mathbf{\Psi}}_{\mathrm{ML}} = \left(\mathbf{C}_{\mathrm{eq}}^H(\mathbf{c})\mathbf{C}_{\mathrm{eq}}(\mathbf{c}) + 2\epsilon\mathbf{I}\right)^{-1}\mathbf{C}_{\mathrm{eq}}^H(\mathbf{c})\mathbf{Y}'. \quad (16)$$

If in (15), we replace $\tilde{\mathbf{\Psi}}$ by its estimate $\tilde{\mathbf{\Psi}}_{\mathrm{ML}}$, then the penalized ML solution for $\mathbf{c}$ is

$$\hat{\mathbf{c}}_{\mathrm{ML}} = \arg\min_{\mathbf{c}} ||\mathbf{Y}' - \mathbf{C}_{\mathrm{eq}}(\mathbf{c})\left(\mathbf{C}_{\mathrm{eq}}^H(\mathbf{c})\mathbf{C}_{\mathrm{eq}}(\mathbf{c}) + 2\epsilon\mathbf{I}\right)^{-1} \times \mathbf{C}_{\mathrm{eq}}^H(\mathbf{c})\mathbf{Y}'||^2. \quad (17)$$

It is possible to further simplify (17) for the case of unpenalized ML estimation ($\epsilon = 0$) and attack the optimization problem with gradient-based methods [26]. We have implemented the conjugate gradient method in a Fletcher-Reeves setting with an optimized step-size through a line-search. However, this turns out to be far less robust than, and computationally more expensive to, the method provided next. Therefore we omit to provide the gradient in closed form.

[7]Ridge Regression [25] is an empirical regression approach widely used in many practical fields, e.g., Machine Learning [24], as it provides estimation robustness when the model is subject to a number of degeneracies. This turns out to the case in this work, and we point out why this occurs later. However, we emphasize that the main reason of adding the penalty terms is to control the convergence of the algorithm, which we also point out later why this is the case. To finalize, we parametrize the penalty term (13) with a single parameter in order simplify the convergence analysis and be able to extract meaningful insights.

[8]The factor 2 in the regularization term of (16) appears since $\psi_{m,n} = \psi_{n,m}$. Note that ϵ is considered as a constant during the optimization, otherwise it is obvious that $\epsilon = 0$ minimizes (13).

C. An EM Algorithm to find the joint Penalized-ML Estimate

Here we provide a robust and computational efficient algorithm to find the joint penalized-ML estimate of $\mathbf{c}$ and $\mathbf{\Psi}$. Instead of pursuing an approach similar to the one used to reach (17), the algorithm has its roots in the joint solution found by setting the gradient of $J_{\mathrm{ML}}(\mathbf{Y},\mathbf{C},\mathbf{\Psi},\epsilon)$ to zero. Before presenting the algorithm, we therefore briefly address this gradient approach.

Each entry of (11) is given by $y_{n,m} = \psi_{n,m}c_m + n_{n,m}$. The derivative of $J_{\mathrm{ML}}(\mathbf{Y},\mathbf{C},\mathbf{\Psi},\epsilon)$ with respect to c_m^* is given by

$$\frac{\partial J_{\mathrm{ML}}(\mathbf{Y},\mathbf{C},\mathbf{\Psi},\epsilon)}{\partial c_m^*} = \epsilon c_m + \sum_{\substack{n=1 \\ n\neq m}}^{M} |\psi_{n,m}|^2 c_m - y_{n,m}\psi_{n,m}^*. \quad (18)$$

Setting (18) to zero and solving for c_m yields

$$c_m = \left(\epsilon + \sum_{\substack{n=1 \\ n\neq m}}^{M} |\psi_{n,m}|^2\right)^{-1} \sum_{\substack{n=1 \\ n\neq m}}^{M} \psi_{n,m}^* y_{n,m}, \quad (19)$$

which can be expressed in a vector form as

$$\hat{\mathbf{c}}_{\mathrm{ML}} = \left(\mathbf{\Psi}_{\mathrm{eq}}^H(\tilde{\mathbf{\Psi}})\mathbf{\Psi}_{\mathrm{eq}}(\tilde{\mathbf{\Psi}}) + \epsilon\mathbf{I}\right)^{-1}\mathbf{\Psi}_{\mathrm{eq}}^H(\tilde{\mathbf{\Psi}})\widetilde{\mathbf{Y}}. \quad (20)$$

In a similar fashion, setting the derivative of $J_{\mathrm{ML}}(\mathbf{Y},\mathbf{C},\mathbf{\Psi},\epsilon)$ with respect to $\psi_{n,m}^*$ to zero and solving for $\psi_{n,m}$ provides

$$\psi_{n,m} = \left(|c_n|^2 + |c_m|^2 + 2\epsilon\right)^{-1}\left(y_{m,n}c_n^* + y_{n,m}c_m^*\right), \quad (21)$$

which can be expressed in a vector form as (16). Equations (19) and (21) show the analytical form for each entry of the penalized-ML vector estimates, which will prove to be useful during the complexity analysis. Combining the results from (20) and (16) yield the joint solution

$$\begin{bmatrix} \hat{\mathbf{c}}_{\mathrm{ML}} \\ \tilde{\mathbf{\Psi}}_{\mathrm{ML}} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{\Psi}_{\mathrm{eq}}^H(\tilde{\mathbf{\Psi}}_{\mathrm{ML}})\mathbf{\Psi}_{\mathrm{eq}}(\tilde{\mathbf{\Psi}}_{\mathrm{ML}}) + \epsilon\mathbf{I}\right)^{-1}\mathbf{\Psi}_{\mathrm{eq}}^H(\tilde{\mathbf{\Psi}}_{\mathrm{ML}})\widetilde{\mathbf{Y}} \\ \left(\mathbf{C}_{\mathrm{eq}}^H(\hat{\mathbf{c}}_{\mathrm{ML}})\mathbf{C}_{\mathrm{eq}}(\hat{\mathbf{c}}_{\mathrm{ML}}) + 2\epsilon\mathbf{I}\right)^{-1}\mathbf{C}_{\mathrm{eq}}^H(\hat{\mathbf{c}}_{\mathrm{ML}})\mathbf{Y}' \end{bmatrix} \quad (22)$$

The particular structure of (22) suggests that a pragmatic approach for solving can be pursued. More specifically, (22) can be separated into two sub-problems, i.e., solving for $\hat{\mathbf{c}}_{\mathrm{ML}}$ and $\tilde{\mathbf{\Psi}}_{\mathrm{ML}}$ separately. Since each of the solutions depend on previous estimates, the joint solution can be computed iteratively, by sequentially solving two separate regularized LS problems, given an initial guess. Since each iteration estimates $\mathbf{c}$ and $\tilde{\mathbf{\Psi}}$ separately, this approach can be seen as an instance of the EM algorithm [27], where the - often challenging - *Expectation step* is performed by estimating only the first moment of the nuisance parameters $\{\psi_{m,n}\}$. The convergence of the algorithm can be analyzed using standard methods, such as a distance between consecutive point estimates. The GMM estimator can be used to compute a reliable initial guess for iteration - in contrast to a purely random initialization. This is often good practice to ensure convergence to a *suitable* local optimum since $J_{\mathrm{ML}}(\mathbf{Y},\mathbf{C},\mathbf{\Psi},\epsilon)$ is not a convex function of its joint parameter space. For sake of clarity, Algorithm 1 summarizes the proposed iterative procedure.

Algorithm 1 Expectation-Maximization

Require: Measurement matrix $\mathbf{Y}$, convergence threshold Δ_{ML}, penalty parameter ϵ, initial guess $\hat{\mathbf{c}}$

1: **Initialization**: set $\Delta = \delta$ where $\delta > \Delta_{\mathrm{ML}}$
2: **while** $\Delta \geq \Delta_{\mathrm{ML}}$ **do**
3: $\quad \tilde{\mathbf{\Psi}}_{\mathrm{ML}} = \left(\mathbf{C}_{\mathrm{eq}}^{H}(\hat{\mathbf{c}})\mathbf{C}_{\mathrm{eq}}(\hat{\mathbf{c}}) + 2\epsilon\mathbf{I}\right)^{-1}\mathbf{C}_{\mathrm{eq}}^{H}(\hat{\mathbf{c}})\mathbf{Y}'$
4: $\quad \hat{\mathbf{c}}_{\mathrm{ML}} = \left(\mathbf{\Psi}_{\mathrm{eq}}^{H}(\tilde{\mathbf{\Psi}}_{\mathrm{ML}})\mathbf{\Psi}_{\mathrm{eq}}(\tilde{\mathbf{\Psi}}_{\mathrm{ML}}) + \epsilon\mathbf{I}\right)^{-1}\mathbf{\Psi}_{\mathrm{eq}}^{H}(\tilde{\mathbf{\Psi}}_{\mathrm{ML}})\tilde{\mathbf{Y}}$
5: $\quad \Delta = ||\hat{\mathbf{c}}_{\mathrm{ML}} - \hat{\mathbf{c}}||^2$
6: $\quad \hat{\mathbf{c}} = \hat{\mathbf{c}}_{\mathrm{ML}}$
7: **end while**

Output: Calibration coefficients estimate $\hat{\mathbf{c}}_{\mathrm{ML}}$

Observe that ϵ, i.e. the penalty term parameter in (13), ends up regularizing both matrix inversions. This is of notable importance from two points-of-view: *i)* from an estimation (robustness) point-of-view, since the matrices to be inverted are constructed from parameter estimates (and thus are subject to estimation errors) and no favorable guarantee exists on their condition number, e.g., see (35). *ii)* from a convergence point-of-view, as it is well-known that the convergence rate of regularized LS adaptive filters is inversely proportional to their eigenvalue spread [28]; This property combo justifies why Ridge Regression was pursued in the first place.

A side remark regarding an application of the EM algorithm follows. We highlight that the calibration coefficients $\mathbf{c}$ and the equivalent channels $\psi_{m,n} = r_m h_{m,n} r_n$ are jointly estimated. As previously mentioned, this a feature is not present in the GMM estimator. Noticeably, this feature makes the EM algorithm robust and hence very suitable to calibrate distributed MIMO systems since channel fading (i.e., high variations of $|h_{m,n}|$) often occurs [13]. As mentioned in Sec. III-A, the system model used can be also representative to that of distributed systems.

D. Complexity Analysis

The complexity of each iteration of Algorithm 1 is dominated by steps 3 and 4. Fortunately the block diagonal structure of the equivalent matrices allows for the inversions to be of reduced complexity, as detailed next. From (21), each calculation of $\psi_{m,n}$ requires a few multiplications and additions. Since $\left(M^2 - M\right)/2$ such calculations are needed to compute (16), the complexity order of step 3 is $\mathcal{O}(M^2)$. Similarly, the complexity of step 4 is $\mathcal{O}(M^2)$ which can be seen directly from (19). The explanation of the $\mathcal{O}(M^2)$ behavior is that the complexity of each calibration coefficient c_m is $\mathcal{O}(M)$, and M such calibration coefficients need to be computed. Overall, each iteration of the EM algorithm is of complexity $\mathcal{O}(M^2)$, and the algorithm's complexity is $\mathcal{O}(N_{\mathrm{ite}}\, M^2)$, with N_{ite} being the number of iterations needed for convergence. The number of iterations needed for convergence is studied in Sec. III-E3.

As for the GMM estimator, the closed-form solutions presented in [13] and [15] have complexity orders of $\mathcal{O}(M^3)$, as they consist of an inverse of a Hermitian matrix of size $M-1$, and of the eigenvector associated with the smallest eigenvalue of a Hermitian matrix of size M.

On a practical note, we remark that the computational complexity of both approaches does not stand as a prohibitive factor for BS arrays using hundreds or even several thousands of antennas. This is because calibration typically needs to be performed on a hourly basis [14], [22].

E. Performance Assessment

1) Simulation setup for the MSE analysis: We simulate reciprocity calibration over a 4×25 rectangular array as the one in Fig. 1. The linear regression parameters obtained in Sec. II-D2 are used to model the coupling gains $\bar{h}_{m,n}$. The mth transceiver maps to the antenna in row a_{row} and column a_{col} of the array as $m = 25(a_{\mathrm{row}} - 1) + a_{\mathrm{col}}$. The reference transceiver index is set to *ref* $= 38$, as it is associated with one of the most central antenna elements of the 2-D array.

The Cramér-Rao Lower Bound (CRLB) is computed to verify the asymptotical properties of the estimators' error [27]. From (6) and (8), it can be seen that if $\bar{h}_{m,n}$ is assumed to be known, the PDF of $\mathbf{Y}$ conditioned on $\mathbf{R}$ and $\mathbf{T}$ is a multivariate Gaussian PDF. This makes the CRLB of $\mathbf{c}$ to have a well known closed-form, which is computed in Appendix B.

The transmitter t_m and receiver r_m gains are set to $t_m = (0.9 + \frac{0.2m}{M}\exp(-j2\pi m/M))/t_{ref}$ and $r_m = (0.9 + \frac{0.2(M-m)}{M}\exp(j2\pi m/M))/r_{ref}$, respectively. We used this deterministic setting for the transceivers, as it allows for a direct comparison of the parameter estimates' MSE with the CRLB. Moreover, this setting incorporates eventual mismatches within the transceivers complex amplitude which are in line with the magnitude variations measured from the transmitters/receivers of our testbed, i.e., spread of around 10-percent around the mean magnitude (and uniform phase). This spread is in line with transceiver models adopted in other calibration works [13].

The variance σ^2 of the multipath propagation contribution during calibration is set to -60 dB. Our motivation for this value is as follows. If the closest physical scatter to the BS is situated, say, 15 meters away, then by Friis' law [29] we have a path loss of around $10\log_{10}(\frac{4\pi d}{\lambda}) = 10\log_{10}(\frac{4\pi(2\times 15m)}{3\times 10^8/(3.7\times 10^9)}) = 73$ dB per path. This number does not account for further losses due to reflections and scattering. Based on this, we use -60 dB as the power (variance) of the resulting channel stemming from a large number of such uncorrelated paths.

For consistency with the reference antenna concept used in the CRLB computations, the MSE of the EM algorithm output $\hat{\mathbf{c}}_{\mathrm{ML}}$, is defined as

$$\mathrm{MSE}_m = \mathrm{E}\left\{ |c_m - [\hat{\mathbf{c}}_{\mathrm{ML}}]_{m,1} / [\hat{\mathbf{c}}_{\mathrm{ML}}]_{ref,1}|^2 \right\}, \qquad (23)$$

since the estimated "reference" coefficient $[\hat{\mathbf{c}}_{\mathrm{ML}}]_{ref,1}$ is not necessarily equal to 1. This is because the concept of reference antenna is not used by the EM algorithm. As for the GMM estimator, the constraint provided in [15] is adopted, i.e., $c_{ref} = 1$ in (10), which is already coherent with the computed CRLB. The results are averaged over 1000 Monte-Carlo simulations, and the threshold Δ_{ML} is set to 10^{-6} which, based on our experience, ensures that convergence is reached

in many parameter settings. The initial guess for the EM algorithm is produced by the GMM estimator.

2) Estimators' MSE vs CRLB: Fig. 3 compares the MSE of the estimators with the CRLB for two transceiver cases. Both estimators appear to be asymptotically efficient. Noticeably, the performance gains of the EM algorithm can be grossly superior to the GMM (up to 10 dB), as it approaches the CRLB at much smaller values of N_0. As mentioned previously, this is mainly because the GMM estimator does not appropriately weight moment conditions with less quality.

Two remarks about the CRLB itself are now in place. *i)* As mentioned in Appendix B, the assumptions used during the CRLB computations, could result in an underestimated CRLB. Indeed, the results in Fig. 3 suggest that the assumptions used during the CRLB computations do not affect its final value since the estimators' MSE asymptotically converges to the computed CRLB. This is convenient since (asymptotically) efficient estimators can still be built with limited information. *ii)* It was assumed that $\phi_{m,n}$ - the phase of $\bar{h}_{m,n}$ - is known during the CRLB computations, although it is originally modeled as a random variable in Sec.II-D2. However, if $\phi_{m,n}$ is assumed to be known, the CRLB is independent of the value of $\phi_{m,n}$. This is because a phase rotation in $\boldsymbol{\mu}_{n,m}$, does not influence (42), due to the structure of $\boldsymbol{\Sigma}^{-1}$. Thus, any realization of $h_{m,n}$ - from the model proposed in Sec. II-D - provides the same CRLB result.

From the previous two remarks and standard estimation theory [27], it follows that the (narrowband) calibration error - in the high SNR regime - produced by the studied estimators can be well modeled as a multivariate zero-mean Gaussian distribution with covariance matrix given by the transformed inverse Fisher information matrix, found in (38). The Gaussianity of the calibration error is further verified (experimentally) in Sec. V-D.

3) Convergence of the EM algorithm: The convergence is analyzed for $N_0 = -40$ dB, which from Fig. 3 appears to be a region where EM-based estimation provides significant gains compared to GMM. Fig. 4 illustrates the role played by the regularization constant ϵ in terms of convergence rate and MSE. Noticeably, the higher ϵ the faster the algorithm appears to converge. The number of iterations until convergence N_{ite} is seen to be much smaller than M with large enough ϵ (i.e., around 5 iterations when $\epsilon = 0.1$).[9] However, increasing ϵ indefinitely is not an option as it degrades the performance. Moreover, the results also indicate that proper tuning of ϵ can provide MSE gains compared to the unregularized case which is asymptotically efficient (notice that this does not conflict with the CRLB theorem, as an estimator built with $\epsilon \neq 0$ is not necessarily unbiased). This was - to some extent - expected due the benefits of Ridge Regression as discussed in Sec.III-C.

With that, we identify that a fine tuning of ϵ can provide many-fold improvements. We note that in the literature there is a number of approaches available that deal with optimization of regularization constants in standard (non-iterative) LS problems [24]. However, they are not directly applicable to this work as they typically optimize single error metrics, and are in general computationally expensive. Here, our main use for ϵ is to accelerate the convergence and provide estimation robustness to the algorithm, all achieved at no complexity cost. For this matter, we treat ϵ as a hyperparameter (an approach widely adopted in regularized LS adaptive filtering [28]). Further investigation on fully automatizing the EM algorithm is an interesting matter of future work.

For the remainder of the paper, we set $\epsilon = 0$ and proceed accordingly, for simplicity.

4) Simulation Setup for Sum-rate Capacity Analysis : The same parameter setting as in Sec. III-E1 is kept in this setup, and the remaining simulation framework is defined next.

We assume that the uplink channel $\mathbf{H}_{\text{UP}}$ is perfectly know to the BS, and that there are two noise sources in the system. The first noise source is downlink additive noise modeled by $\mathbf{w}'$, see (2). Here, $\mathbf{w}'$ have i.i.d. zero-mean circularly symmetric complex Gaussian distributed random entries with variance N_w equal to 1. The same model is used for the entries of the downlink channel matrix $\mathbf{H}_{\text{DL}}$. The second noise source is the error during estimation of $\mathbf{c}$ (i.e., calibration error). With that, the precoded signal $\mathbf{z}' = \mathbf{P}\mathbf{x}$ is subject to calibration errors. The transmit power constraint $\mathbb{E}\{||\mathbf{z}'||^2\} = K$ is used. Also, we set $K = 10$ single antenna users, and assume $t_k^U = t_k^B$ and $r_k^U = r_k^B$ for sake of simplicity.

The sum-rate capacities [30] are evaluated for different calibration cases. More specifically, when no calibration is employed (i.e., $\hat{c}_m = 1$), when calibration is performed with the GMM or the EM algorithm, for the case of perfect calibration (i.e., $\hat{c}_m = c_m$), and as a baseline, when precoding is performed using the *true* downlink channel $\mathbf{H}_{\text{DL}}$. The analysis is performed with $N_0 = -40$ dB, for the reasons mentioned during the convergence analysis.

5) Sum-rate Capacity Results: Fig. 5 shows the obtained sum-rates cumulative distribution functions (CDFs) for different precoding schemes [2]. Similarly to the MSE results, EM-based calibration provides significant gains compared to the GMM case. The magnitude of these gains obviously depend on both the calibration (and communication system) setup. For example, there are no sum-rate differences when $N_0 \to 0$ or $N_0 \to \infty$, as both GMM and EM approaches converge to that of perfect calibration, or to the uncalibrated case, respectively. Thus, it in only in a certain region of N_0 values that EM based calibration provides gains.[10]

It is interesting that - for this setup - there is no fundamental loss in capacity between this calibration approach (i.e., precoding with perfectly calibrated uplink CSI) and precoding with the true downlink CSI. Quantifying this loss is out of scope of this work, however, the interested reader is referred to [32] for an overview on the loss of different types of reciprocity calibration. We now finalize the section with two interesting remarks.

[9] If, instead, the initial guess is chosen randomly (e.g., calibration coefficients with unit-norm and i.i.d. uniform phases) then our simulations indicate that the order of N_{ite} is $\mathcal{O}(M)$.

[10] Our analysis based on a wide range of parameter values also indicates that, in general, stricter calibration requirements need to be met in order to release the full potential of ZF compared to MRT precoding (i.e., no sum-rate difference compared to the perfect calibrated case). Noticeably, this observation is in line with previous calibration studies [31].

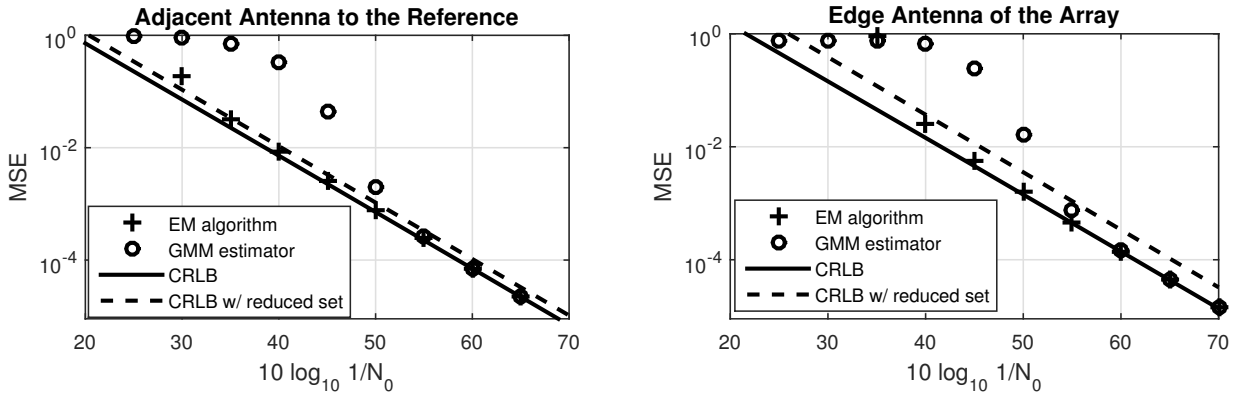


Fig. 3. MSE of the GMM estimator and the EM algorithm (with $\epsilon = 0$), versus their CRLB (solid line), for 2 extreme transceiver cases. Namely, a transceiver associated with an antenna at the edge of the array, and a transceiver associated with an antenna adjacent to the reference. The CRLB plotted by a dashed line is discussed in Sec. III-F.

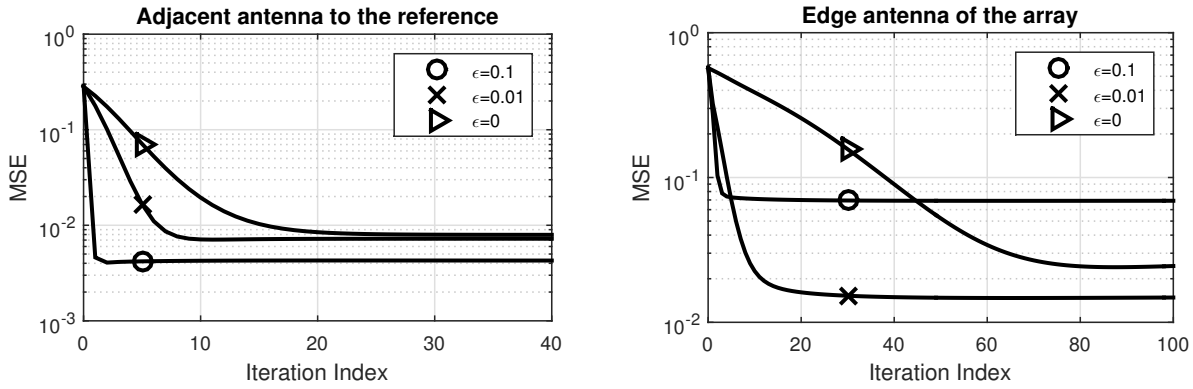


Fig. 4. MSE per iteration of the EM algorithm, for different regularization constants ϵ. The plots are for $N_0 = -40$ dB, and the remaining simulation settings are the same as Fig. 3. Note the different scales of the plots.

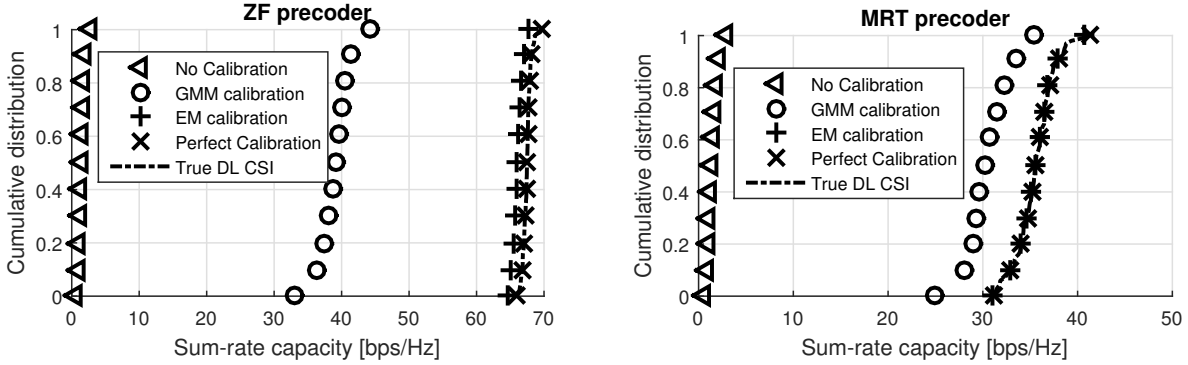


Fig. 5. CDFs of the sum-rates capacities for different calibration cases. Left) ZF precoder; Right) MRT precoder.

F. Remark 1: Calibration with Reduced Measurement Sets

There are several benefits of using a reduced measurement set for calibration (e.g., by only relying on high quality measurements). This is possible as long as (11) is not underdetermined. As an illustrative example, the dashed line in Fig. 3 shows the CRLB when a reduced measurements set - comprising the measurements between antenna pairs whose elements are distanced by at most $1/\sqrt{2}$ wavelengths - is used. The number of measurement signals in this case drops from $M(M-1)$ to less than $8M$, since one antenna signals to, at most, 8 other antennas. The performance loss turns out to be insignificant, i.e. 2 dB for the neighbor case and 4 dB for the edge case, considering the number of signals discarded. This indicates that the channels between neighbor antennas, which are dominated by mutual coupling, are the most important for calibration. Thus, there is an interesting trade-off between the asymptotic performance of an estimator and its computational complexity (proportional to the number of measurements).

Another benefit of using reduced measurement sets is a possible reduction of resource overhead dedicated for calibration. This can be very important from a system deployment point-of-view. To finalize, we remark that ML closed form estimators

can be also reached when reduced measurement sets are used. This can be the case for the current (general) calibration setup when a reduced set of measurements is used, or for the case of working with a full set of measurements when the calibration setup is a special case. An example of the latter is given next.

G. Remark 2: Closed-form Unpenalized ML Calibration for Linear arrays

Consider an M-antenna linear array, and let m index the antennas in ascending order starting at one edge of the linear array. Assume that mutual coupling only exists between adjacent antenna elements, and that the channel between any other antenna pairs is weak enough so that it can be neglected without any noticeable impact on performance. We summarize our findings in Proposition 1.

Proposition 1: Using a reference antenna as a starting point, say $c_1 = 1$, the unpenalized ML solution for any $c_{\ell+1}$, with $1 \leq \ell \leq M-1$, can be obtained sequentially by

$$\hat{c}_{\ell+1} = \hat{c}_{\ell} \frac{y^*_{\ell+1,\ell} y_{\ell,\ell+1}}{|y_{\ell+1,\ell}|^2}. \tag{24}$$

Proof: See Appendix C. ■

We can also deduce the following interesting corollary.

Corollary 1: For any of the two constraints considered in (10), the GMM (vector) estimator coincides with (24) up to a common complex scalar.

Proof: See Appendix C. ■

IV. Validation of the calibration method in a massive MIMO testbed

In this section, we detail the experiment performed to validate the proposed mutual coupling based calibration method. More specifically, we implemented it in a software-defined radio testbed, and performed a TDD transmission from 100 BS antennas to 3 single antenna terminals.

Note that the analysis conducted in this section and in Sec. V is measurement based. As stationarity is assumed in the analysis, we monitored the system temperature throughout the measurements and verified no significant changes. We also made an effort to keep static propagation conditions, and performed the experiments at late hours in our lab with no people around.

A. Brief Description of the Testbed

Here we briefly outline the relevant features of the testbed for this work. Further information can be found in [22].

1) Antenna/Transceiver setup: The BS operates with 100 antennas, each antenna connected to one distinct transceiver. For simplicity, the same transceiver settings (e.g., power amplifier gain and automatic gain control) are used in both calibration and data communication stages for all radio units. This ensures that the analog front-ends yield the same response during both stages, thus the estimated calibration coefficients are valid during the communication stage.

TABLE I
High-level OFDM parameters

Parameter	Variable	Value
Carrier frequency	f_{c}	3.7 GHz
Sampling Rate	F_{s}	7.68 MS/s
FFT Size	N_{FFT}	2048
# Used sub-carriers	N_{SUB}	1200

2) Synchronization of the radios: Time and Frequency synchronization is achieved by distributing reference signals to all radio units. However, this does not guarantee phase alignment between all BS transceiver radio chains which motivates reciprocity calibration.

B. Communication Protocol used

Once the measurements to construct the observation matrix $\mathbf{Y}$ are performed, $\mathbf{c}$ is estimated using the unpenalized EM algorithm. The following sequence of events is then performed periodically:

1) Uplink Channel Estimation and Calibration: Users simultaneously transmit frequency orthogonal pilot symbols. The BS performs LS-based channel estimation, and interpolates the estimates between pilot symbols. Reciprocity calibration is then performed independently per subcarrier, i.e. as in (3), for coherence purposes with Sec. II. This calibrated version of the downlink channel is then used to construct a ZF precoder.

2) Downlink channel estimation and data transmission: Downlink pilot symbols are precoded in the downlink and each user performs LS-based channel estimation. Using the estimates, each user recovers the payload data using a one-tap equalizer.

We note that 4-QAM signaling per OFDM sub-carrier is used for uplink channel estimation and data transmission. The main parameters are shown in Table I. Further information on the signaling protocol (e.g., uplink/downlink frame structure or uplink pilot design) is found on [22].

C. Measurement Description

The setup used in our experiments is shown in Figure 1. Although not being a typical propagation scenario found in cellular systems, this extreme setup - closely located users under strong line-of-sight conditions - requires high calibration requirements to be met if spatial separation of users is to be achieved. In addition, we use ZF precoding as it is known to be very sensitive to calibration errors [32].

The EVM [33] of the downlink equalized received samples at each mobile station was evaluated, and used as performance metric for validation purposes. The rationale is that, with multiple mobile terminals, calibration errors are translated into downlink inter-user interference (and loss of array gain), which increases the EVM. Letting r be the downlink equalized received sample when symbol s is transmitted, the EVM is defined as

$$\mathrm{EVM} = \mathrm{E}\left\{ \frac{|r-s|^2}{|s|^2} \right\}, \tag{25}$$

where the expectation is taken over all system noise sources (e.g., hardware impairments and thermal noise). Our estimate of (25) was obtained by averaging realizations of $|r-s|^2/|s|^2$ over all OFDM sub-carriers and over received OFDM symbols.

We estimated the EVM for different energy values of the uplink pilots and calibration signals. We do so in order to be able to extract insightful remarks for the analysis of the results. In particular, letting $E_{\mathrm{Pilot}} = \mathrm{E}\{p_k p_k^*\}$ in (1) denote the energy of the uplink pilot, which, for simplicity, is the same for all users, and let E_{Cal} denote the energy of the sounding signal s_m in (5), we estimated the EVM for a 2-dimensional grid of E_{Pilot} and E_{Cal}. The results reported next are given with respect to the relative energies $Er_{\mathrm{Pilot}} = E_{\mathrm{Pilot}}/E_{\mathrm{Pilot}}^{\max}$ and $Er_{\mathrm{Cal}} = E_{\mathrm{Cal}}/E_{\mathrm{Cal}}^{\max}$, where $E_{\mathrm{Pilot}}^{\max}$ and $E_{\mathrm{Cal}}^{\max}$ are the maximum energies of the uplink pilot and calibration signal used in the experiments. Other systems parameters (e.g., transmit power in the downlink) were empirically set and kept constant throughout the experiment.

D. Validation Results

Fig. 6 shows the measured EVMs for the 3 user terminals in our experiment. Before discussing the results, we remark that analyzing the EVM when Er_{Cal} is reduced beyond -30 dB is not of fundamental interest, as it approaches the uncalibrated case (where high EVMs are to be expected). Overall, a positive trend is observed with increasing Er_{Cal} until -10 dB. This reflects the BS ability of spatially separating users which increases with increasing the calibration quality. The fact that downlink EVMs down to -10 dB are achieved, which are much smaller than the EVMs when $Er_{\mathrm{Cal}} = -30$ dB, i.e. close to the uncalibrated case, motivates our validation claim.

It is possible to observe a saturation of the EVMs at high enough Er_{Cal} and Er_{Pilot} for all user cases. This is an expected effect in practical systems. Explained briefly, system impairments other than the calibration or the uplink channel estimation error, become the dominant error sources that bound the EVM performance[11]. Remarkably, this saturation effect implies that the calibration SNR - available in a practical array as ours - is sufficiently large not to be the main impairment to constrain the system performance. Mutual coupling channels are thus reliable (and reciprocal enough), so that they can be used for signaling in order to calibrate the system.[12]

V. Aspects of Wideband Calibration and Error Modeling

A short summary of this section follows. Using the measurements from the Sec. IV, we treat the estimated calibration coefficients across OFDM sub-carriers as realizations of a discrete stochastic process. Using low rank approximation theory, we propose a parametrized low dimensional basis that characterizes the subspace spanned by this process accurately. Based on the reduced basis, we propose a wideband estimator that averages out the calibration error across frequency. Using the wideband estimator results, we validate the narrowband calibration error model proposed in Sec. III-E2. We remark that our experiment makes use of a bandwidth of $F_s N_{\mathrm{sub}}/N_{\mathrm{FFT}} = 4.5$MHz.

A. Wideband Remarks for the Calibration Coefficients

Denote the calibration coefficient of BS antenna m at the kth OFDM sub-carrier as $C_m[k] = t_m^k/r_m^k$. The variable $\hat{C}_m[k]$ is the estimate of $C_m[k]$ at sub-carrier k - obtained, e.g., with the EM algorithm - and is modeled as

$$\begin{aligned}\hat{C}_m[k] =& C_m[k] + E_m[k] \\ =& |C_m[k]| \exp(j2\pi\zeta_m[k]) + E_m[k] \end{aligned} \quad (26)$$

where $E_m[k]$ is an i.i.d. random process representing the calibration error which is assumed zero-mean and independent of $C_m[k]$. Let the random phasor process $\exp(j2\pi\zeta_m[k])$ in (26) absorb the phase shift stemming from the arbitrary time that a local oscillator needs to lock to a reference signal. Such phase shift is often modeled as uniformly distributed, and thus

$$\mathrm{E}\{\exp(j2\pi\zeta_m[k])\} = 0. \quad (27)$$

Moreover, since local oscillators associated with different transceivers lock at arbitrary times, it is safe to assume

$$\mathrm{E}\{\exp(j2\pi\zeta_m[k_1])\exp(-j2\pi\zeta_n[k_2])\} = 0,\ m \neq n. \quad (28)$$

Not making further assumptions on the statistics of $\hat{C}_m[k]$, we now proceed with a series expansion, but before doing so we make one last remark. The series expansion conducted next is performed based on measurements from the 100 testbed transceivers, and serves as an example approach to obtain a suitable basis for $\hat{C}_m[k]$. This can well apply to mass-production transceiver manufactures that can reliably estimate the statistical properties of the hardware produced. However, as our testbed operates with relatively high-end transceivers - compared to the ones expected to integrate commercial massive MIMO BSs - the dimensionality of the subspace verified in our analysis might be underestimated. Intuitively, the higher transceiver quality, the less basis functions are needed to accurately describe $\hat{C}_m[k]$. Nevertheless, the upcoming remarks apply for smaller bandwidths - than 4.5MHz - depending on the properties of the transceivers.

B. Principal Component Analysis

From the assumption (27), it follows that the element at the v_1th row and v_2th column of the covariance matrix $\mathbf{K}_m$ of $\hat{C}_m[k]$ is defined as

$$[\mathbf{K}_m]_{[v_1,v_2]} = \mathrm{E}\left\{\hat{C}_m[v_1]\ \hat{C}_m^*[v_2]\right\}. \quad (29)$$

[11]Mobile terminals error sources (e.g., in-phase and quadrature imbalance or thermal noise) qualify for such impairments. For a given downlink transmit power, it is straightforward to understand how such impairments bound the downlink EVMs regardless of the calibration and uplink estimation quality.

[12]We note there exists an interesting theoretical trade-off between the calibration quality and the capacity of downlink channels with respect to the strength of mutual coupling. In practice, the proposed calibration method can be used in compact antenna arrays with very low coupling (say -30 dB between adjacent elements) provided that the transmit power during calibration is sufficient to provide good enough estimation SNR. In such a setup, the impact of coupling in the capacity is negligible.

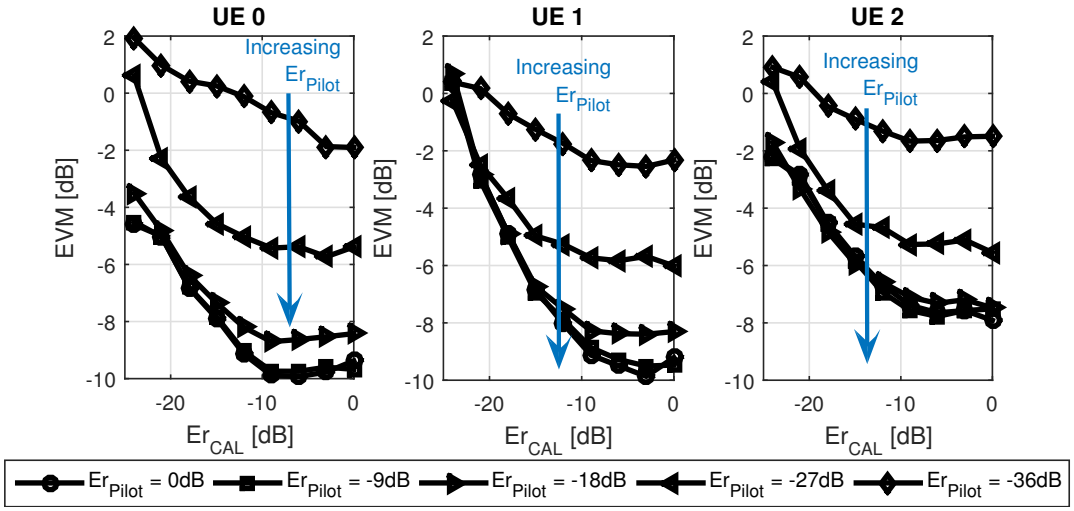


Fig. 6. Measured EVM at each of the three user terminals during a massive MIMO downlink transmission.

From the assumption (28), it follows that the principal components of $\hat{C}_m[k]$ are obtained by singular value decomposition (SVD) of $\mathbf{K}_m$ only [34]. Let the SVD of $\mathbf{K}_m$ be written as

$$\mathbf{K}_m = \sum_{i=1}^{\mathrm{N_{SUB}}} \mathbf{u}_i^m \lambda_i^m (\mathbf{u}_i^m)^H, \tag{30}$$

where $\{\mathbf{u}_i^m\}_{i=1}^{\mathrm{N_{SUB}}}$ are the principal components, and λ_i^m is the power (variance) of the coefficient obtained from projecting $\hat{C}_m[k]$ into $\mathbf{u}_i^m$. We use the convention $\lambda_1^m \geq \lambda_2^m \cdots \geq \lambda_{N_{\mathrm{SUB}}}^m$, and $\mathbf{u}_i^m = \left[[u_i^m[1], \cdots, u_i^m[N_{\mathrm{SUB}}]\right]^{\mathrm{T}}$. Fig. 7 shows several coefficients and basis functions of the expansion, that were estimated based on 100 realizations of $\hat{C}_m[k]$, each measured with $Er_{\mathrm{Cal}} = 5$ dB (which from Fig. 6 provides a relatively high calibration SNR). Noticeably, it appears that all processes (one per transceiver) live mostly in a one-dimensional sub-space and thus can be well described by their first principal component $\mathbf{u}_1^m$. This fact also indicates that the contribution of the calibration error in the expansion is small, and thus the first principal component of $\hat{C}_m[k]$ is also representative for the true coefficients $C_m[k]$.

Visual inspection indicates that both magnitude and phase of the first principal component can be well approximated with a linear slope across frequency. The inherent error of this approximation is very small compared to the magnitude of the process itself. We note that this linear trend holds for any transceiver of the array (not only for the ones shown in Fig. 7).

C. Wideband Modeling and Estimation

The previous analysis indicates that any first principal component can be well described by a linear magnitude slope γ_m, and a linear phase ξ_m across frequency. Such properties are well captured by the Laplace kernel $\exp((\gamma_m + j2\pi\xi_m)k)$, for small values of $|\gamma_m|$ (since the range of k is finite). The final parameter to model a realization of the process is the complex offset A_m. With that, the general model (26) can thus be re-written as

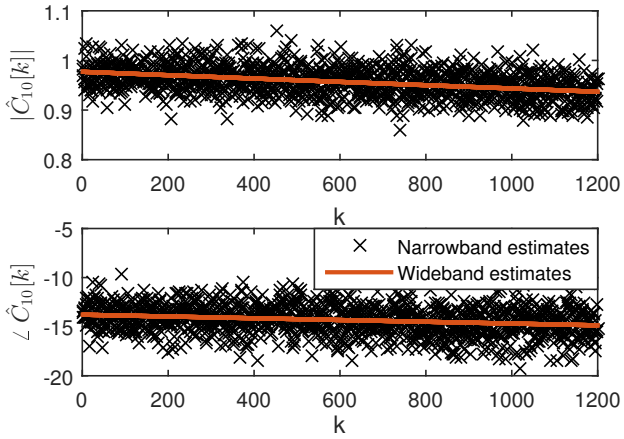


Fig. 8. A realization of the narrow-band estimator $\hat{C}_m[k]$, and the proposed wideband estimator $\hat{C}_i[k]^{\mathrm{WB}}$.

$$\hat{C}_m[k] = A_m \exp((\gamma_m + j2\pi\xi_m)k) + w_m[k], \tag{31}$$

where $w_m[k]$ is a random process that absorbs: the calibration error $E_m[k]$, the error due to the low rank approximation, and the error due to the linear modeling of the first principal component $\mathbf{u}_1^m$. Given an observation $\{\hat{C}_m[k]\}_{k=1}^{N_{\mathrm{SUB}}}$, the ML estimator of A_m, ξ_m and γ_m, namely, $\hat{A}_m$, $\hat{\xi}_m$ and $\hat{\gamma}_m$ is straightforward to derive [27]. Thus, we define the wideband estimator of $\hat{C}_m[k]$ as

$$\hat{C}_m[k]^{\mathrm{WB}} = \hat{A}_m \exp((\hat{\gamma}_m + j2\pi\hat{\xi}_m)k). \tag{32}$$

For illustration purposes, a realization of the ML wideband estimator $\hat{C}_m[k]^{\mathrm{WB}}$ is contrasted with that of the narrow-band estimator $\hat{C}_m[k]$ in Fig. 8. The obtained error reduction is evident.

D. A Model for the Calibration Error

Here, we use the wideband estimator results to verify the Gaussianity of the narrow-band calibration error proposed

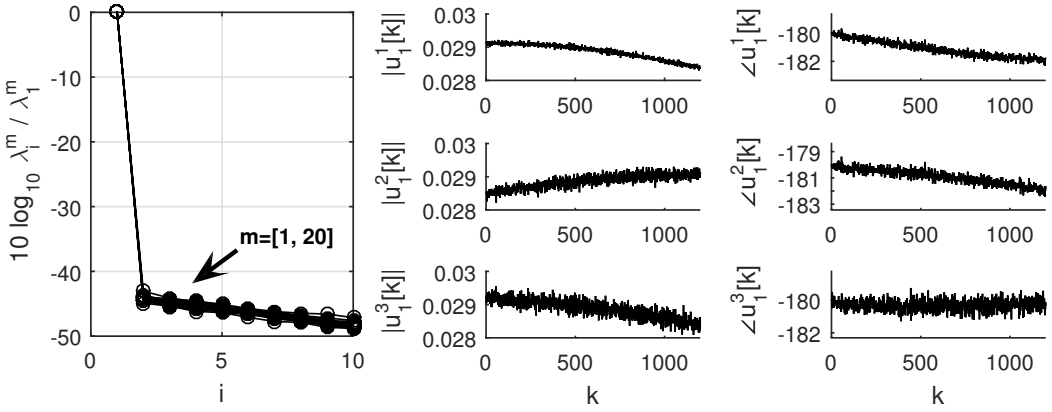


Fig. 7. Principal component and coefficients of $\hat{C}_m[k]$. *Left)* The 10 strongest normalized singular values for 20 transceivers; *Middle)* Magnitude of the principal component for 3 transceivers; *Right)* Phase of the principal component for 3 transceivers.

in Sec. III-E2. This is done under the two following main assumptions.

1) The residual process $E_m[k] = \hat{C}_m[k] - C_m[k]$ is well described by $\hat{E}_m[k] = \hat{C}_m[k] - \hat{C}_m[k]^{\mathrm{WB}}$. This is reasonable if $\mathrm{E}\left\{|\hat{C}_m[k]^{\mathrm{WB}} - C_m[k]|^2\right\} \ll \mathrm{E}\left\{|\hat{C}_m[k] - C_m[k]|^2\right\}$. To justify, the estimation gains scale linearly in the number of realizations [27], which is $N_{\mathrm{SUB}} = 1200$ in this case. Assuming that: the estimation error is independent across realizations, the underlying model (31) describes the first principal component well, and the low rank approximation error is minuscule, there are gains of $10\log_{10} N_{\mathrm{SUB}} \approx 30$ dB which justify the first main assumption.

2) The residual process $E_m[k]$ is ergodic.[13] This is met if $E_m[k]$ is stationary and the ensemble of N_{SUB} samples is representative for statistical modeling. The former holds for small OFDM bandwidths (e.g., 4.5 MHz) as the hardware impairments do not vary significantly across the band. The latter is also met, as we have $N_{\mathrm{SUB}} = 1200$ narrow-band estimators whose estimated errors $\{\hat{E}_m[k]\}_{k=1}^{N_{\mathrm{SUB}}}$ were found to be mutually uncorrelated.

Fig. 9 shows the empirical CDF of both real and imaginary parts of $\{\hat{E}_m[k]\}_{k=1}^{N_{\mathrm{SUB}}}$ - which we found to the uncorrelated - for two transceiver cases. Each of the empirical CDFs is contrasted with a zero-mean Gaussian distribution of equal variance. Overall, the empirical CDFs for both transceivers resemble a Gaussian CDF extremely well. The Gaussianity of the calibration error was further verified by passing a Kolmogorov-Smirnov test with 0.05 significance level [35]. We note that these observations hold not only for the two transceivers in Fig. 9, but for all transceivers of the array. Noticeably, the empirical distribution of the calibration error is in line with the asymptotic properties of ML estimators, i.e. the error can be modeled by an additive zero-mean Gaussian multivariate. The final element for a full characterization is its covariance matrix, relating the errors across antennas.

[13]Ergodicity is necessary since each (independent) measurement of $\hat{C}_m[k]$ takes about ten minutes with our test system (due to the locking time of the local oscillator to the reference signal). As potential system temperature drifts during the measurements can result in varying statistical properties, it is safer to perform the analysis based on one solely realization of $E_m[k]$.

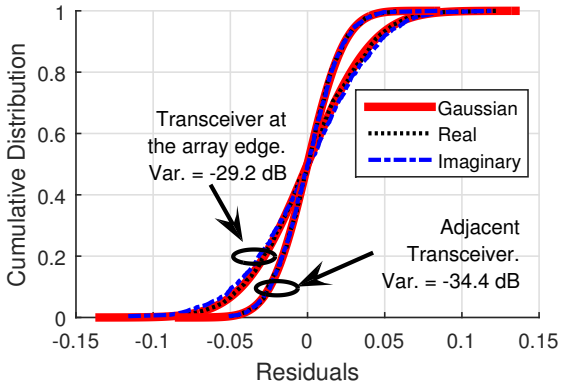


Fig. 9. Empirical CDFs for the real and imaginary parts of the calibration error, for a transceiver at the edge of the array, and for an adjacent transceiver to the reference antenna. A Gaussian CDF of equal variance is plotted for both cases for comparison.

A good approximation (at high SNR) is the inverse of the transformed Fisher Information matrix in (38). Noticeably, future calibration works can benefit from the convenience of safely assuming a non-white Gaussian calibration error.

VI. Conclusions

We have proposed and validated a convenient calibration method which rely on mutual coupling to enable the reciprocity assumption in TDD massive MIMO systems. We verified that in a practical antenna array, the channels due to mutual coupling are reliable and reciprocal enough, so that they can be used for signaling in order to calibrate the array.

The iterative ML algorithm is asymptotically efficient and outperforms current state-of-the-art estimators in an MSE and sum-rate capacity sense. Further improvements - in terms of MSE and convergence rate - can be harvested by proper tuning of its regularization hyperparameter.

The calibration error can be further reduced by proper averaging over the radio bandwidth. More importantly, it did not stand as the main impairment to constraint the performance of the system, from our experiments. Our measurements also

verified that the narrow-band calibration error (at high SNR) is Gaussian distributed, which is coherent with the theory of the estimator proposed. The convenience of safely assuming a non-white Gaussian calibration error can, hopefully, open the door for future analytical studies of calibrated TDD massive MIMO systems.

ACKNOWLEDGMENTS

This work was funded by the Swedish foundation for strategic research SSF, VR, the strategic research area ELLIIT, and the E.U. Seventh Framework Programme (FP7/2007-2013) under grant agreement n 619086 (MAMMOET). We also thank the Comm. Systems group in Bristol University, for letting us replicate several of our results in their testbed.

APPENDIX A: EQUIVALENT CHANNEL MATRICES

Here we show the structure of the equivalent models. Define the column vector $\mathbf{\Psi}_m = [\psi_{1,m} \dots \psi_{m-1,m}\ \psi_{m+1,m} \dots \psi_{M,m}]^T$. The equivalent channel matrix in (14) is written as

$$\mathbf{\Psi}_{\text{eq}}(\tilde{\mathbf{\Psi}}) = \text{diag}\left\{\mathbf{\Psi}_1, \mathbf{\Psi}_2, \dots, \mathbf{\Psi}_M\right\}. \tag{33}$$

Now define

$$\bar{\mathbf{c}}_{n,m} = [c_n\ c_m]^T. \tag{34}$$

Noting that $\psi_{m,n} = \psi_{n,m}$, the equivalent matrix and the parameter vector in (15) are written as

$$\mathbf{C}_{\text{eq}}(\mathbf{c}) = \text{diag}\left\{\bar{\mathbf{c}}_{1,2}, \cdots, \bar{\mathbf{c}}_{1,M}, \bar{\mathbf{c}}_{2,3}, \cdots, \bar{\mathbf{c}}_{2,M}, \cdots\right\}, \tag{35}$$

and

$$\tilde{\mathbf{\Psi}} = \left[\psi_{2,1} \dots \psi_{M,1}\ \psi_{3,2} \dots \psi_{M,2} \dots \psi_{M,M-1}\right]^T. \tag{36}$$

APPENDIX B: THE CRAMÉR-RAO LOWER BOUND

Here we compute the CRLB for the calibration coefficients $\{c_m\} \setminus c_{ref}$. The exclusion of c_{ref} is justified in the end of the calculations. This is achieved by assuming $t_{ref} = r_{ref} = 1$, and treating $c_{ref} = t_{ref}/r_{ref}$ as known for estimation purposes. Define the $(4M-4)\times 1$ vector

$$\boldsymbol{\theta} = [\text{Re}\{t_1\}\ \text{Im}\{t_1\}\ \text{Re}\{r_1\}\ \text{Im}\{r_1\}\ \text{Re}\{t_2\} \dots \text{Im}\{r_M\}]^T, \tag{37}$$

where t_{ref} and r_{ref} do not enter. The CRLB for $\{c_m\} \setminus c_{ref}$ is given by the diagonal entries of the transformed inverse Fisher information matrix [27]

$$\text{var}(\hat{c}_m) \geq \left[\frac{q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{q(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^{\text{H}}\right]_{m,m},\ m \neq ref, \tag{38}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix of $\boldsymbol{\theta}$. The transformation of $\boldsymbol{\theta}$ into the calibration coefficients is given by

$$q(\boldsymbol{\theta}) = \left[\frac{\text{Re}\{t_1\} + j\,\text{Im}\{t_1\}}{\text{Re}\{r_1\} + j\,\text{Im}\{r_1\}} \cdots \frac{\text{Re}\{t_M\} + j\,\text{Im}\{t_M\}}{\text{Re}\{r_M\} + j\,\text{Im}\{r_M\}}\right]^T.$$

We now compute $\mathbf{I}(\boldsymbol{\theta})$. Assuming that $\bar{h}_{m,n}$, σ^2 and N_0 are at hand,[14] the mean $\boldsymbol{\mu}_{n,m}$ and the covariance matrix $\mathbf{\Sigma}_{n,m}$ of $\mathbf{y}_{n,m} = [y_{n,m}\ y_{m,n}]^T$ are given by

$$\boldsymbol{\mu}_{n,m} = \text{E}\left\{\mathbf{y}_{n,m}\right\} = \bar{h}_{n,m}\left[r_n t_m\ r_m t_n\right]^T, \tag{39}$$

$$\begin{aligned}\mathbf{\Sigma}_{n,m} &= \text{E}\left\{(\mathbf{y}_{n,m} - \boldsymbol{\mu}_{n,m})(\mathbf{y}_{n,m} - \boldsymbol{\mu}_{n,m})^H\right\} \\ &= \begin{bmatrix} |r_n|^2|t_m|^2\sigma^2 + N_0 & r_n t_m r_m^* t_n^* \sigma^2 \\ r_m t_n r_n^* t_m^* \sigma^2 & |r_m|^2|t_n|^2\sigma^2 + N_0 \end{bmatrix}.\end{aligned} \tag{40}$$

We can observe that the PDF of $\mathbf{Y}''$, where

$$\mathbf{Y}'' = \left[\mathbf{y}_{1,2}^T \dots \mathbf{y}_{1,M}^T\ \mathbf{y}_{2,3}^T \dots \mathbf{y}_{2,M}^T \dots \mathbf{y}_{M-1,M}^T\right]^T,$$

conditioned on $\boldsymbol{\theta}$, follows a multivariate Gaussian distribution, i.e., $p(\mathbf{Y}''|\boldsymbol{\theta}) \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$, with mean $\boldsymbol{\mu} = \left[\boldsymbol{\mu}_{1,2}^T \dots \boldsymbol{\mu}_{1,M}^T \boldsymbol{\mu}_{2,3}^T \dots \boldsymbol{\mu}_{2,M}^T \dots \boldsymbol{\mu}_{M-1,M}^T\right]^T$ and block diagonal covariance

$$\mathbf{\Sigma} = \text{diag}\left\{\mathbf{\Sigma}_{1,2}, \cdots, \mathbf{\Sigma}_{1,M}, \mathbf{\Sigma}_{2,3}, \cdots, \mathbf{\Sigma}_{2,M}, \cdots, \mathbf{\Sigma}_{M-1,M}\right\}. \tag{41}$$

With that, we have

$$[\mathbf{I}(\boldsymbol{\theta})]_{i,j} = \text{Tr}\left\{\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_i}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\right\} + 2\,\text{Re}\left\{\frac{\partial \boldsymbol{\mu}^H}{\partial \theta_i}\mathbf{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \theta_j}\right\}, \tag{42}$$

with $1 \leq i \leq (4M-4)$ and $1 \leq j \leq (4M-4)$. The remaining computations of $[\mathbf{I}(\boldsymbol{\theta})]_{i,j}$ are straightforward and thus omitted. We note that without the convention of $t_{ref} = r_{ref} = 1$ - and thus $\boldsymbol{\theta}$ is a $4M \times 1$ vector instead - it can be shown that the map $\boldsymbol{\theta} \mapsto \boldsymbol{\mu}$ is not injective which renders $\mathbf{I}(\boldsymbol{\theta})$ not invertible. Thus, the convention of reference antenna is needed to be able to compute the CRLB.

APPENDIX C - CLOSED-FORM UNPENALIZED ML ESTIMATOR FOR LINEAR ARRAYS

Here we derive the closed-form unpenalized (i.e. $\epsilon = 0$) ML estimator for the linear array setup described in Sec. III-G. By leaving out the terms that do not depend on $\mathbf{c}$, it follows that, after a few manipulations, the optimization problem of (17) can be written as

$$\begin{aligned}\{\hat{c}_m\} &= \arg\max_{\mathbf{c}} \mathbf{Y'}^H \mathbf{C}_{\text{eq}}(\mathbf{c})\mathbf{C}_{\text{eq}}^{\dagger}(\mathbf{c})\mathbf{Y'} \\ &= \arg\max_{\{c_m\}} \sum_{\ell=1}^{M-1} f_{\text{L}}(c_\ell, c_{\ell+1}, \mathbf{y}_{\ell+1,\ell}),\end{aligned} \tag{43}$$

with

$$f_{\text{L}}(c_\ell, c_{\ell+1}, \mathbf{y}_{\ell+1,\ell}) = \mathbf{y}_{\ell+1,\ell}^H \bar{\mathbf{c}}_{\ell,\ell+1}\bar{\mathbf{c}}_{\ell,\ell+1}^H \mathbf{y}_{\ell+1,\ell} / \bar{\mathbf{c}}_{\ell,\ell+1}^H \bar{\mathbf{c}}_{\ell,\ell+1}.$$

See (34) for structure of $\bar{\mathbf{c}}_{\ell,\ell+1}$, and (40) for structure of $\mathbf{y}_{m,n}$. Our ability to solve (43) is due to the following property.

Property 1: For the function $f_{\text{L}}(c_\ell, c_{\ell+1}, \mathbf{y}_{\ell+1,\ell})$, the maximum over $c_{\ell+1}$ equals $||\mathbf{y}_{\ell+1,\ell}||^2$, and thus it does not depend on c_ℓ.

[14] These assumptions are only used for the CRLB calculations, and were not used to derive any of the estimators. A possible implication is that the CRLB can be underestimated, but we will see that this is not the case from the simulations' results.

Hence, the ML estimate of $c_{\ell+1}$, i.e. $\hat{c}_{\ell+1}$, can be found for a given c_ℓ. With that, the joint maximization problem (43) can be split into

$$\hat{c}_{\ell+1} = \arg\max_x f_{\mathrm{L}}(\hat{c}_\ell, x, \mathbf{y}_{\ell+1,\ell}).$$

This optimization is a particular case of the Rayleigh quotient problem, and the solution is given in (24) when the reference element (i.e., the starting point) is chosen to be c_1.

We now provide a short proof for Corollary 1. For the case of linear arrays with coupling solely between adjacent antennas, the optimization problem in (10) can be written - ignoring any constraint for now - as

$$\hat{\boldsymbol{c}}_{\mathrm{GMM}} = \arg\min_{\boldsymbol{c}} \sum_{\ell=1}^{M-1} f_{\mathrm{G}}(c_\ell, c_{\ell+1}, \mathbf{y}_{\ell+1,\ell}) \tag{44}$$

where $f_{\mathrm{G}}(c_\ell, c_{\ell+1}, \mathbf{y}_{\ell+1,\ell}) = |y_{\ell+1,\ell}c_{\ell+1} - y_{\ell,\ell+1}c_\ell|^2$. We solve (44) using the following property.

Property 2: Letting $\hat{c}_\ell$ be the ML estimator from (24), it follows that

$$f_{\mathrm{G}}(\hat{c}_\ell, \hat{c}_{\ell+1}, \mathbf{y}_{\ell+1,\ell}) = 0, \ \forall \ell. \tag{45}$$

Thus, the GMM solution (under any of the 2 constraints) coincides with that of the ML up to a common complex scalar. Uniqueness follows since the GMM cost function is quadratic.

REFERENCES

[1] 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 Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[3] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb 2002.

[4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[5] M. Salehi and J. Proakis, *Digital Communications*. McGraw-Hill Education, 2007.

[6] C. A. Balanis, *Antenna Theory: Analysis and Design*. Wiley-Interscience, 2005.

[7] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *2010 Future Network and Mobile Summit*, June 2010.

[8] M. Petermann et al., "Multi-User Pre-Processing in Multi-Antenna OFDM TDD Systems with Non-Reciprocal Transceivers," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3781–3793, September 2013.

[9] K. Nishimori, K. Cho, Y. Takatori, and T. Hori, "Automatic calibration method using transmitting signals of an adaptive array for TDD systems," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 6, pp. 1636–1640, Nov 2001.

[10] K. Nishimori, T. Hiraguri, T. Ogawa, and H. Yamada, "Effectiveness of implicit beamforming using calibration technique in massive MIMO system," in *2014 IEEE International Workshop on Electromagnetics (iWEM)*, Aug 2014, pp. 117–118.

[11] X. Luo, "Robust Large Scale Calibration for Massive MIMO," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015.

[12] H. Wei, D. Wang, H. Zhu, J. Wang, S. Sun, and X. You, "Mutual Coupling Calibration for Multiuser Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 606–619, Jan 2016.

[13] R. Rogalin et al., "Scalable Synchronization and Reciprocity Calibration for Distributed Multiuser MIMO," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, April 2014.

[14] C. Shepard et al., "Argos: Practical many-antenna base stations," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 53–64.

[15] R. Rogalin, O. Y. Bursalioglu, H. C. Papadopoulos, G. Caire, and A. F. Molisch, "Hardware-impairment compensation for enabling distributed large-scale mimo," in *Information Theory and Applications Workshop (ITA), 2013*, Feb 2013.

[16] H. Wei, D. Wang, J. Wang, and X. You, "TDD reciprocity calibration for multi-user massive MIMO systems with iterative coordinate descent," *Science China Information Sciences*, vol. 59, no. 10, p. 102306, 2015.

[17] H. Papadopoulos, O. Bursalioglu, and G. Caire, "Avalanche: Fast RF calibration of massive arrays," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2014, pp. 607–611.

[18] J. Vieira, F. Rusek, and F. Tufvesson, "Reciprocity calibration methods for massive MIMO based on antenna coupling," in *2014 IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 3708–3712.

[19] H. Wei, D. Wang, and X. You, "Reciprocity of mutual coupling for TDD massive MIMO systems," in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, Oct 2015.

[20] J. Xiwen et al., "MIMO-TDD reciprocity under hardware imbalances: Experimental results," in *2015 IEEE International Conference on Communications ICC, 8-12 June 2015, London, United Kingdom*, London, U.K., 2015.

[21] R. Jedlicka, M. Poe, and K. Carver, "Measured mutual coupling between microstrip antennas," *IEEE Transactions on Antennas and Propagation*, vol. 29, no. 1, pp. 147–149, Jan 1981.

[22] J. Vieira et al., "A flexible 100-antenna testbed for Massive MIMO," in *IEEE GLOBECOM 2014 Workshop on Massive MIMO: from theory to practice*, Dec 2014.

[23] A. Hall, *Generalized Method of Moments*, ser. Advanced Texts in Econometrics. OUP Oxford, 2004.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[25] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 42, no. 1, pp. 80–86, Feb. 2000.

[26] L. Scharf and C. Demeure, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, ser. Addison-Wesley series in electrical and computer engineering. Addison-Wesley Publishing Company, 1991.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., 1993.

[28] S. Haykin, *Adaptive Filter Theory (3rd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[29] A. Molisch, *Wireless Communications*, ser. Wiley - IEEE. Wiley, 2010.

[30] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2008.

[31] X. Luo, "Multi-User Massive MIMO Performance with Calibration Errors," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4521–4534, Jul 2016.

[32] W. Zhang et al., "Large-Scale Antenna Systems With UL/DL Hardware Mismatch: Achievable Rates Analysis and Calibration," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1216–1229, April 2015.

[33] T. Schenk, *RF Imperfections in High-rate Wireless Systems: Impact and Digital Compensation*. Springer, 2008.

[34] H. Van Trees, *Detection, Estimation, and Modulation Theory*. Wiley, 2004.

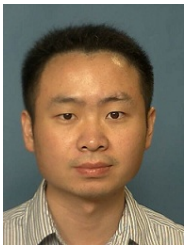
[35] W. Daniel, *Applied nonparametric statistics*, ser. The Duxbury advanced series in statistics and decision sciences. PWS-Kent Publ., 1990.

Joao Vieira received the B.Sc. degree in Electronics and Telecommunications Engineering from University of Madeira in 2011, and the M.Sc. degree in Wireless Communications from Lund University, Sweden in 2013. He is currently working towards a Ph.D. degree at the department of Electrical and Information Technology in Lund University. His main research focus is on parameter estimation and implementation issues in massive MIMO systems.

Fredrik Rusek was born in Lund, Sweden, in 1978. He received the M.S. and Ph.D. degrees in electrical engineering from Lund University, Lund, Sweden, in 2003 and 2007, respectively. Since 2012, he has held an Associate Professorship with the Department of Electrical and Information Technology, Lund Institute of Technology. His research interests include modulation theory, equalization, wireless communications, and applied information theory

Ove Edfors is Professor of Radio Systems at the Department of Electrical and Information Technology, Lund University, Sweden. His research interests include statistical signal processing and low-complexity algorithms with applications in wireless communications. In the context of Massive MIMO, his main research focus is on how realistic propagation characteristics influence system performance and base-band processing complexity.

Steffen Malkowsky received the B.Eng. degree in Electrical Engineering and Information Technology from Pforzheim University, Germany in 2011 and the M.Sc. degree in Electronic Design from Lund University in 2013. He is currently a PhD student in the Digital ASIC group at the department of Electrical and Information Technology, Lund University. His research interests include development of reconfigurable hardware and implementation of algorithms for wireless communication with emphasis on massive MIMO. For the development of a massive MIMO testbed in collaboration with University of Bristol and National Instruments and a set spectral efficiency world record, he received 5 international awards from National Instruments, Xilinx and Hewlett Packard Enterprise.



Liang Liu received his B.S. and Ph.D. degree in the Department of Electronics Engineering (2005) and Micro-electronics (2010) from Fudan University, China. In 2010, he was with Rensselaer Polytechnic Institute (RPI), USA as a visiting researcher. He joined Lund University as a Post-doc in 2010. Since 2016, he is Associate Professor at Lund University. His research interest includes wireless communication system and digital integrated circuits design. He is a board member of the IEEE Swedish SSC/CAS chapter. He is also a member of the technical committees of VLSI systems and applications and CAS for communications of the IEEE circuit and systems society.

Fredrik Tufvesson received his Ph.D. in 2000 from Lund University in Sweden. After two years at a startup company, he joined the department of Electrical and Information Technology at Lund University, where he is now professor of radio systems. His main research interests are channel modelling, measurements and characterization for wireless communication, with applications in various areas such as massive MIMO, UWB, mm wave communication, distributed antenna systems, radio based positioning and vehicular communication. Fredrik is a fellow of the IEEE. He has authored around 60 journal papers and 120 conference papers, recently he got the Neal Shepherd Memorial Award for the best propagation paper in IEEE Transactions on Vehicular Technology.