Performance Analysis of Reciprocity Calibration in Massive MIMO

Kalyana Gopala
Sequans Communications, Sophia Antipolis, France
kalyan.krishnan@gmail.com

Dirk Slock
EURECOM, Sophia Antipolis, France
slock@eurecom.fr

***Abstract*—Time Division Duplexing (TDD) Massive MIMO (MaMIMO) relies on channel reciprocity to derive the channel state information at the transmit side (CSIT) from the uplink (UL) channel estimates. This reciprocity is impacted by the transmit and receive front ends which are non-reciprocal and hence a calibration is required to obtain the downlink (DL) channels from the UL channel estimates. The Cramer-Rao Bound (CRB) for reciprocity calibration and an optimal algorithm were derived in [1], [2] recently. In this work, we extend this and derive a generalized CRB that applies to both coherent and non-coherent calibration schemes (introduced in [1]). More importantly, we provide original mean squared error analysis for the commonly used Least-Squares (LS) estimator under different parameter constraints, and relate it to the CRB. Finally, we compare some antenna grouping strategies for calibration based on their CRB.**

I. Introduction

In this paper Tx and Rx may denote transmit/transmitter/transmission and receive/receiver/reception. In Time Division Duplexing (TDD) Massive MIMO(MaMIMO), reciprocity of the propagation channel is exploited to derive the downlink (DL) channel from the uplink channel (UL) that is relatively easier to obtain. This becomes clear once we see that in the UL with one single pilot transmission from a user equipment (UE) the channel for all the Base Station (BS) antennas may be derived, whereas, in the DL, separate pilots are needed for every BS antenna which becomes prohibitive in the MaMIMO scenario. However, as the Radio Frequency (RF) components are not reciprocal, the DL estimation requires additional scale factor to compensate for the non-reciprocity of the RF front ends.

Many approaches to reciprocity calibration have been proposed in the literature. In UE aided calibration, explicit channel feedback from a UE during the calibration phase is used to estimate the calibration parameters.However, a popular approach is to perform the calibration across the antennas of the base station (BS) only and is referred to as internal calibration. The authors of [3] gave the first experimental evaluation for a Massive MIMO system with a simple algorithm for internal calibration. [4] improved up on the performance of Argos by making better use of all the available information in the least squares fashion. A weighted least squares minimization was presented in [5]. A faster calibration algorithm called Avalanche was introduced in [6]. In [7], the authors proposed a Cramer-Rao bound (CRB) and a penalized maximum likelihood estimation approach that performs close to the CRB. A generalized approach towards reciprocity calibration was proposed in [1] of which the existing estimation techniques are special cases. [1] proposes a model where transmission can happen from a group of multiple antennas at every channel use. The paper also shows an elegant way to derive the CRB for this generalized approach. For the first time, it also introduced a non-coherent approach to reciprocity calibration. In [2], a variational Bayes based optimal algorithm was presented. In this paper, we focus on the performance analysis of the reciprocity calibration algorithms for Massive MIMO. The main contributions of our current work are as follows.

- Generalized CRB derivation that applies to both coherent and non-coherent reciprocity calibration techniques.
- We provide an original analysis of the mean squared error (MSE) for the commonly used Least Squares (LS) estimator and relate it to the CRB in a manner that makes the results intuitively appealing.
- Comparison and differentiation of different antenna grouping strategies based on their CRB.

In the following, bold-face capital letters refer to matrices and bold-face small letters correspond to column vectors. The notation $\otimes$ refers to the Kronecker product.

II. System Model

The system model and notations are kept similar to that in [2] for better readability. Consider a system where A represents a BS and B represents a UE, each containing M_A and M_B antennas resp. The channel, as observed in the digital domain, $\mathbf{H}_{A\to B}$ and $\mathbf{H}_{B\to A}$ can be represented by,

$$\mathbf{H}_{A\to B} = \mathbf{R}_B \mathbf{C}_{A\to B} \mathbf{T}_A,\ \mathbf{H}_{B\to A} = \mathbf{R}_A \mathbf{C}_{B\to A} \mathbf{T}_B, \quad (1)$$

where matrices $\mathbf{T}_A$, $\mathbf{R}_A$, $\mathbf{T}_B$, $\mathbf{R}_B$ model the response of the transmit and receive RF front-ends, while $\mathbf{C}_{A\to B}$ and $\mathbf{C}_{B\to A}$ model the propagation channels, respectively from A to B and from B to A. The dimension of $\mathbf{T}_A$ and $\mathbf{R}_A$ are $M_A \times M_A$, whereas that of $\mathbf{T}_B$ and $\mathbf{R}_B$ are $M_B \times M_B$. The diagonal elements in these matrices represent the linear effects attributable to the impairments in the transmitter and receiver parts of the RF front-ends respectively, whereas the off-diagonal elements correspond to non-reciprocity in RF crosstalk and antenna mutual coupling. It is worth noting that although transmitting and receiving antenna mutual coupling are not generally reciprocal [8], theoretical modeling [9] and experimental results [10], [3], [7] both show that, in practice, RF crosstalk and antenna mutual coupling are sufficiently reciprocal to be ignored for the purpose of reciprocity calibration, which implies that $\mathbf{T}_A$, $\mathbf{R}_A$, $\mathbf{T}_B$, $\mathbf{R}_B$ can safely be assumed to be diagonal.

Assuming that the system is operating in TDD mode, the channel responses enjoy reciprocity within the channel coherence time, i.e., $\mathbf{C}_{A\to B} = \mathbf{C}_{B\to A}^T$. Therefore, we obtain the following relationship between the channels measured in both directions:

$$\mathbf{H}_{A\to B} = \underbrace{\mathbf{R}_B\mathbf{T}_B^{-T}}_{\mathbf{F}_B^{-T}}\mathbf{H}_{B\to A}^T\underbrace{\mathbf{R}_A^{-T}\mathbf{T}_A}_{\mathbf{F}_A} = \mathbf{F}_B^{-T}\mathbf{H}_{B\to A}^T\mathbf{F}_A. \quad (2)$$

Note that the studies in [11], [12] pointed out that in a practical multi-user MIMO system, it is mainly the calibration at the BS side which restores the hardware asymmetry and helps to achieve the multi-user MIMO performance. Hence, in the sequel, the focus is on the estimation of $\mathbf{F}_A$.

Let us consider an antenna array of M elements partitioned into G groups denoted by $A_1, A_2, \ldots, A_G$. Group A_i contains M_i antennas such that $\sum_{i=1}^G M_i = M$. Each group A_i transmits a sequence of L_i pilot symbols, defined by matrix $\mathbf{P}_i \in \mathbb{C}^{M_i\times L_i}$ where the rows correspond to antennas and the columns to successive channel uses. Note that a channel use can be understood as a time slot or a subcarrier in an OFDM-based system, as long as the calibration parameter can be assumed constant over all channel uses. When an antenna group i transmits, all other groups are considered to be in receiving mode. After all G groups have transmitted, the received signal for each resource block of bi-directional transmission between antenna groups i and j is given by

$$\begin{cases} \mathbf{Y}_{i\to j} = \mathbf{R}_j\mathbf{C}_{i\to j}\mathbf{T}_i\mathbf{P}_i + \mathbf{N}_{i\to j}, \\ \mathbf{Y}_{j\to i} = \mathbf{R}_i\mathbf{C}_{j\to i}\mathbf{T}_j\mathbf{P}_j + \mathbf{N}_{j\to i}, \end{cases} \quad (3)$$

where $\mathbf{Y}_{i\to j} \in \mathbb{C}^{M_j\times L_i}$ and $\mathbf{Y}_{j\to i} \in \mathbb{C}^{M_i\times L_j}$ are received signal matrices at antenna groups j and i respectively when the other group is transmitting. $\mathbf{N}_{i\to j}$ and $\mathbf{N}_{j\to i}$ represent the corresponding received noise matrix. $\mathbf{T}_i,\ \mathbf{R}_i \in \mathbb{C}^{M_i\times M_i}$ and $\mathbf{T}_j,\ \mathbf{R}_j \in \mathbb{C}^{M_j\times M_j}$ represent the effect of the transmit and receive RF front-ends of antenna elements in groups i and j respectively.

The reciprocity property implies that $\mathbf{C}_{i\to j} = \mathbf{C}_{j\to i}^T$, thus for two different groups $1 \le i \ne j \le G$, by eliminating $\mathbf{C}_{i\to j}$ in (3) we have

$$\mathbf{P}_i^T\mathbf{F}_i^T\mathbf{Y}_{j\to i} - \mathbf{Y}_{i\to j}^T\mathbf{F}_j\mathbf{P}_j = \widetilde{\mathbf{N}}_{ij}, \quad (4)$$

where the noise component $\widetilde{\mathbf{N}}_{ij} = \mathbf{P}_i^T\mathbf{F}_i^T\mathbf{N}_{j\to i} - \mathbf{N}_{i\to j}^T\mathbf{F}_j\mathbf{P}_j$, while $\mathbf{F}_i = \mathbf{R}_i^{-T}\mathbf{T}_i$ and $\mathbf{F}_j = \mathbf{R}_j^{-T}\mathbf{T}_j$ are the calibration matrices for groups i and j. The calibration matrix $\mathbf{F}$ is diagonal, and thus takes the form of $\mathbf{F} = \mathrm{diag}\{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_G\}$. Let us use $\mathbf{f}_i$ and $\mathbf{f}$ to denote the vectors of the diagonal coefficients of $\mathbf{F}_i$ and $\mathbf{F}$ respectively, i.e., $\mathbf{F}_i = \mathrm{diag}\{\mathbf{f}_i\}$ and $\mathbf{F} = \mathrm{diag}\{\mathbf{f}\}$. This allows us to vectorize (4) into

$$(\mathbf{Y}_{j\to i}^T * \mathbf{P}_i^T)\mathbf{f}_i - (\mathbf{P}_j^T * \mathbf{Y}_{i\to j}^T)\mathbf{f}_j = \widetilde{\mathbf{n}}_{ij}, \quad (5)$$

where $*$ denotes the Khatri–Rao product (or column-wise Kronecker product), where we have used the equality, $\mathrm{vec}(\mathbf{A}\,\mathrm{diag}(\mathbf{x})\,\mathbf{B}) = (\mathbf{B}^T * \mathbf{A})\,\mathbf{x}$. Finally, stacking equations (5) for all $1 \le i < j \le G$ yields

$$\mathcal{Y}(\mathbf{P})\mathbf{f} = \widetilde{\mathbf{n}}, \quad (6)$$

with $\mathcal{Y}(\mathbf{P})$ defined as

$$\underbrace{\begin{bmatrix} (\mathbf{Y}_{2\to1}^T * \mathbf{P}_1^T) & -(\mathbf{P}_2^T * \mathbf{Y}_{1\to2}^T) & 0 & \ldots \\ (\mathbf{Y}_{3\to1}^T * \mathbf{P}_1^T) & 0 & -(\mathbf{P}_3^T * \mathbf{Y}_{1\to3}^T) & \ldots \\ 0 & (\mathbf{Y}_{3\to2}^T * \mathbf{P}_2^T) & -(\mathbf{P}_3^T * \mathbf{Y}_{2\to3}^T) & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{(\sum_{j=2}^G\sum_{i=1}^{j-1} L_iL_j)\times M}. \quad (7)$$

A typical way to estimate $\mathbf{f}$ consists in solving an LS problem such as

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\mathcal{Y}(\mathbf{P})\,\mathbf{f}\|^2 = \arg\min_{\mathbf{f}} \sum_{i<j} \|(\mathbf{Y}_{j\to i}^T * \mathbf{P}_i^T)\mathbf{f}_i - (\mathbf{P}_j^T * \mathbf{Y}_{i\to j}^T)\mathbf{f}_j\|^2\ , \quad (8)$$

where $\mathcal{Y}(\mathbf{P})$ is defined in (7). This needs to be augmented with a constraint

$$\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f}) = 0, \quad (9)$$

in order to exclude the trivial solution $\hat{\mathbf{f}} = \mathbf{0}$ in (8). The constraint on $\hat{\mathbf{f}}$ may depend on the true parameters $\mathbf{f}$. As we shall see further, this constraint needs to be complex valued (which represents two real constraints). Typical choices for the constraint are

1) Norm plus phase constraint (NPC):

$$\text{norm: } \mathrm{Re}\{\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f})\} = ||\hat{\mathbf{f}}||^2 - c\,,\ c = ||\mathbf{f}||^2, \quad (10)$$

$$\text{phase: } \mathrm{Im}\{\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f})\} = \mathrm{Im}\{\hat{\mathbf{f}}^H\mathbf{f}\} = 0. \quad (11)$$

Here Re, Im refer to real and imaginary part, respectively.

2) Linear constraint:

$$\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f}) = \hat{\mathbf{f}}^H\mathbf{g} - c = 0\ . \quad (12)$$

If we choose the vector $\mathbf{g} = \mathbf{f}$ and $c = ||\mathbf{f}||^2$, then the $\mathrm{Im}\{.\}$ part of (12) corresponds to (11). The most popular linear constraint is the First Coefficient Constraint (FCC), which is (12) with $\mathbf{g} = \mathbf{e}_1$, $c = 1$. $\mathbf{e}_1$ represents the first column of the Identity matrix.

The discussion so far has revolved around a coherent reciprocity calibration scheme where it is assumed that the transmission from all the antenna groups may be completed while the channel remains a constant. However, as shown in [13],[1], this can be relaxed to a scenario where the overall calibration can occur across T independent coherent time slots. Within each of these coherent time slots that comprise of several channel uses, the channel stays constant, but is independent across different coherent time slots. During these coherent time slots, the calibration factors are assumed to remain constant. For completeness, we reproduce the metric as was introduced in [13],[1].

$$\hat{\mathbf{f}} = \min_{\mathbf{f}} \|\mathcal{Y}(\mathbf{P})\mathbf{f}\|^2 = \min_{\mathbf{f}} \sum_{t=1}^{T} \sum_{\substack{i,j\in\mathcal{G}(t) \\ i\ne j}} \left\|(\mathbf{Y}_{j\to i,t}^T * \mathbf{P}_i^T(t))\mathbf{f}_i - (\mathbf{P}_j^T(t) * \mathbf{Y}_{i\to j,t}^T)\mathbf{f}_j\right\|^2 \quad (13)$$

where $\mathcal{Y}(\mathbf{P}) = [\mathcal{Y}_1(\mathbf{P}(1))^T, \ldots, \mathcal{Y}_T(\mathbf{P}(T))^T]^T$.

III. Cramer Rao Bound

The CRB for the coherent calibration scenario was presented by the same authors in [2]. While the approach to the computation of the CRB remains the same, we bring in an important generalization here.

- We discuss the CRB computation for the non-coherent reciprocity calibration with T coherent time slots as was introduced in [1], [13]. The non-coherent case reduces to the coherent case when there is just one coherent time slot being considered and bi-directional transmissions between all the antennas can be completed within this time slot.

Note that while the channel can vary across the different coherent time slots, the calibration factors vary more slowly and remain constant across all the coherent slots. Consider a coherent time slot, t where a set of bi-directional transmissions are performed. From (3), we have,

$$\mathbf{Y}_{i\to j}(t) = \underbrace{\mathbf{R}_j\mathbf{C}_{i\to j}(t)\mathbf{R}_i^T}_{\mathcal{H}_{i\to j}(t)}\mathbf{F}_i\mathbf{P}_i(t) + \mathbf{N}_{i\to j}(t). \tag{14}$$

We define $\mathcal{H}_{i\to j}(t) = \mathbf{R}_j\mathbf{C}_{i\to j}(t)\mathbf{R}_i^T$ to be an auxiliary internal channel (not corresponding to any physically measurable quantity) that appears as a nuisance parameter in the estimation of the calibration parameters. Note that the auxiliary channel $\mathcal{H}_{i\to j}(t)$ inherits the reciprocity from the channel $\mathbf{C}_{i\to j}(t)$: $\mathcal{H}_{i\to j}(t) = \mathcal{H}_{j\to i}^T(t)$. Upon applying the vectorization operator for each bidirectional transmission between groups i and j, we have, similarly to (6),

$$\text{vec}(\mathbf{Y}_{i\to j}(t)) = (\mathbf{P}_i^T(t) * \mathcal{H}_{i\to j}(t))\,\mathbf{f}_i + \text{vec}(\mathbf{N}_{i\to j}(t)). \tag{15}$$

In the reverse direction, using $\mathcal{H}_{i\to j}(t) = \mathcal{H}_{j\to i}^T(t)$, we have

$$\text{vec}(\mathbf{Y}_{j\to i}^T(t)) = (\mathcal{H}_{i\to j}^T(t) * \mathbf{P}_j^T)(t)\mathbf{f}_j + \text{vec}(\mathbf{N}_{j\to i}^T(t)). \tag{16}$$

Alternatively, (15) and (16) may also be written as

$$\begin{cases} \text{vec}(\mathbf{Y}_{i\to j}(t)) = \left[(\mathbf{F}_i\mathbf{P}_i(t))^T \otimes \mathbf{I}\right]\text{vec}(\mathcal{H}_{i\to j}(t)) \\ \qquad\qquad + \text{vec}(\mathbf{N}_{i\to j}(t)) \\ \text{vec}(\mathbf{Y}_{j\to i}^T(t)) = \left[\mathbf{I} \otimes (\mathbf{P}_j^T(t)\mathbf{F}_j)\right]\text{vec}(\mathcal{H}_{i\to j}(t)) \\ \qquad\qquad + \text{vec}(\mathbf{N}_{j\to i}(t)). \end{cases} \tag{17}$$

In the case of non-coherent calibration, the key point is to pick only bi-directional transmissions that happen every coherent time slot. Stack all the bi-directional observations into a vector,

$$\mathbf{y}(t) = \left[\text{vec}(\mathbf{Y}_{1\to 2}(t))^T\ \text{vec}(\mathbf{Y}_{2\to 1}^T(t))^T\ \text{vec}(\mathbf{Y}_{1\to 3}(t))^T\ \ldots\right]^T \tag{18}$$

The two alternative formulations in (17) can be summarized into,

$$\mathbf{y}(t) = \mathcal{H}(\mathbf{h}(t), \mathbf{P}(t), t)\mathbf{f} + \mathbf{n} = \mathcal{F}(\mathbf{f}, \mathbf{P}(t), t)\mathbf{h}(t) + \mathbf{n}, \tag{19}$$

where $\mathbf{h}(t) = \left[\text{vec}(\mathcal{H}_{1\to 2})^T\ \text{vec}(\mathcal{H}_{1\to 3})^T\ \text{vec}(\mathcal{H}_{2\to 3})^T\ \ldots\right]^T$, and $\mathbf{n}$ is the corresponding noise vector. The pilot matrix $\mathbf{P}$ is different across different coherent time slots, but to simplify the notation, we omit the explicit dependence on t where there is no room for confusion. The composite matrices $\mathcal{H}(\mathbf{h}(t), \mathbf{P}, t)$ and $\mathcal{F}(\mathbf{f}, \mathbf{P}, t)$ are given below for an example scenario where in t^{th} time slot, the bi-directional transmissions happened between the antenna groups $1, 2$ and 3.

$$\mathcal{H}(\mathbf{h}(t), \mathbf{P}, t) = \begin{bmatrix} \mathbf{P}_1^T * \mathcal{H}_{1\to 2} & 0 & 0 \\ 0 & \mathcal{H}_{1\to 2}^T * \mathbf{P}_2^T & 0 \\ \mathbf{P}_1^T * \mathcal{H}_{1\to 3} & 0 & 0 \\ 0 & 0 & \mathcal{H}_{1\to 3}^T * \mathbf{P}_3^T \\ 0 & \mathcal{H}_{3\to 2}^T * \mathbf{P}_2^T & 0 \\ 0 & 0 & \mathcal{H}_{2\to 3}^T * \mathbf{P}_3^T \end{bmatrix}$$

$$\mathcal{F}(\mathbf{f}, \mathbf{P}, t) = \begin{bmatrix} \mathbf{P}_1^T\mathbf{F}_1 \otimes \mathbf{I} & 0 & 0 \\ \mathbf{I} \otimes \mathbf{P}_2^T\mathbf{F}_2 & 0 & 0 \\ 0 & \mathbf{P}_1^T\mathbf{F}_1 \otimes \mathbf{I} & 0 \\ 0 & \mathbf{I} \otimes \mathbf{P}_3^T\mathbf{F}_3 & 0 \\ 0 & 0 & \mathbf{P}_2^T\mathbf{F}_2 \otimes \mathbf{I} \\ 0 & 0 & \mathbf{I} \otimes \mathbf{P}_3^T\mathbf{F}_3 \end{bmatrix}. \tag{20}$$

Stacking these equations over all T coherent time slots, we get,

$$\begin{aligned} \mathbf{y} &= \underbrace{\begin{bmatrix} \mathcal{H}(\mathbf{h}(1), \mathbf{P}, 1) \\ \vdots \\ \mathcal{H}(\mathbf{h}(T), \mathbf{P}, T) \end{bmatrix}}_{\mathcal{H}(\mathbf{h}, \mathbf{P})}\mathbf{f} + \mathbf{n} \\ \mathbf{y} &= \begin{bmatrix} \mathcal{F}(\mathbf{f}, \mathbf{P}, 1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathcal{F}(\mathbf{f}, \mathbf{P}, T) \end{bmatrix}\underbrace{\begin{bmatrix} \mathbf{h}(1) \\ \vdots \\ \mathbf{h}(t) \end{bmatrix}}_{\mathbf{h}} + \mathbf{n} \\ &= \mathcal{H}(\mathbf{h}, \mathbf{P})\mathbf{f} + \mathbf{n} = \mathcal{F}(\mathbf{f}, \mathbf{P})\mathbf{h} + \mathbf{n}. \end{aligned} \tag{21}$$

Here, $\mathcal{F}(\mathbf{f}, \mathbf{P})$ is a block diagonal matrix whose diagonal block t is $\mathcal{F}(\mathbf{f}, \mathbf{P}, t)$.The scenario is now identical to that encountered in some blind channel estimation scenarios and hence we can take advantage of some existing tools [14],[15], which we exploit next. Treating $\mathbf{h}$ and $\mathbf{f}$ as deterministic unknown parameters, and assuming that the receiver noise $\mathbf{n}$ is distributed as $\mathcal{CN}(0, \sigma^2\mathbf{I})$, the Fisher Information Matrix (FIM) $\mathbf{J}$ for jointly estimating $\mathbf{f}$ and $\mathbf{h}$ can immediately be obtained from (19) as

$$\mathbf{J} = \frac{1}{\sigma^2}\begin{bmatrix} \mathcal{H}^H \\ \mathcal{F}^H \end{bmatrix}\begin{bmatrix} \mathcal{H} & \mathcal{F} \end{bmatrix}. \tag{22}$$

The computation of the CRB requires $\mathbf{J}$ to be non-singular. However, for the problem at hand, $\mathbf{J}$ is inherently singular. To determine the CRB when the FIM is singular, constraints have to be added to regularize the estimation problem. As the calibration parameters are complex, one complex constraint corresponds to two real constraints. Further, we are only interested in the CRB for $\mathbf{f}$ in the presence of the nuisance parameters $\mathbf{h}$. Hence we are only interested in the $(1,1)$ block of the inverse of the 2×2 block matrix $\mathbf{J}$ in (22). Incorporating the effect of the constraint (9) on $\mathbf{f}$, we can derive from [16] the following constrained CRB for $\mathbf{f}$

$$\text{CRB}_{\mathbf{f}} = \sigma^2\mathcal{V}_{\mathbf{f}}\left(\mathcal{V}_{\mathbf{f}}^H\mathcal{H}^H\mathcal{P}_{\mathcal{F}}^{\perp}\mathcal{H}\mathcal{V}_{\mathbf{f}}\right)^{-1}\mathcal{V}_{\mathbf{f}}^H \tag{23}$$

where $\mathcal{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^H\mathbf{X})^{\dagger}\mathbf{X}^H$ and $\mathcal{P}_{\mathbf{X}}^{\perp} = \mathbf{I} - \mathcal{P}_{\mathbf{X}}$ are the projection operators on resp. the column space of matrix $\mathbf{X}$ and its orthogonal complement, and $\dagger$ corresponds to the Moore-Penrose pseudo inverse. The $M \times (M-1)$ matrix $\mathcal{V}_{\mathbf{f}}$ is such that its column space spans the orthogonal complement of that

of $\frac{\partial \mathcal{C}(f)}{\partial \mathbf{f}^*}$, i.e., $\mathcal{P}_{\mathcal{V}_\mathbf{f}} = \mathcal{P}^{\perp}_{\frac{\partial \mathcal{C}}{\partial \mathbf{f}^*}}$. For example, in the FCC case, $\frac{\partial \mathcal{C}(f)}{\partial \mathbf{f}^*} = \mathbf{g} = \mathbf{e}_1$, where $\mathbf{e}_1$ is the first column of the Identity matrix. Hence, $\mathcal{V}_\mathbf{f}$ would be the orthogonal complement of this matrix which would be the remaining columns of the Identity matrix.

IV. MSE ANALYSIS FOR THE LEAST SQUARES ESTIMATOR

In this section, we analyze the mean square error for the LS estimator and compare it with the CRB. From (6), the metric to be optimized is $||\mathcal{Y}\mathbf{f}||^2$ which was derived as a result of eliminating the propagation channel $\mathbf{C}$. The same objective may also be obtained by eliminating the factor $\mathbf{h}$ in (21), specifically by minimizing $||\mathcal{F}^{\perp H}\mathbf{y}||^2$. Here, $\mathcal{F}^{\perp}$ is the orthogonal matrix to $\mathcal{F}$ such that $\mathcal{F}^{\perp H}\mathcal{F}{=}\mathbf{0}$. $\mathcal{F}^{\perp}$ corresponding to (20) is shown in (24).

$$\mathcal{F}^{\perp} = \begin{bmatrix} \mathbf{I}\otimes(\mathbf{F}_2\mathbf{P}_2)^* & 0 & 0 \\ -(\mathbf{F}_1\mathbf{P}_1)^*\otimes\mathbf{I} & 0 & 0 \\ 0 & \mathbf{I}\otimes(\mathbf{F}_3\mathbf{P}_3)^* & 0 \\ 0 & -(\mathbf{F}_1\mathbf{P}_1)^*\otimes\mathbf{I} & 0 \\ 0 & 0 & \mathbf{I}\otimes(\mathbf{F}_3\mathbf{P}_3)^* \\ 0 & 0 & -(\mathbf{F}_2\mathbf{P}_2)^*\otimes\mathbf{I} \end{bmatrix} \tag{24}$$

Thus, we can write,

$$\min_{\hat{\mathbf{f}}} ||\mathcal{F}^{\perp H}(\hat{\mathbf{f}})\mathbf{y}||^2 = \min_{\hat{\mathbf{f}}} ||\mathcal{F}^{\perp H}(\hat{\mathbf{f}})\mathcal{F}\mathbf{h} + \mathcal{F}^{\perp H}(\hat{\mathbf{f}})\mathbf{n}||^2, \tag{25}$$

where the additional term in braces highlights that in this case, the matrix $\mathcal{F}^{\perp H}$ is generated with $\hat{\mathbf{f}}$. When this is not explicitly indicated, the matrix is constructed out of the true $\mathbf{f}$. Next, to get (25) into a more convenient form, we make the following observation.

$$\begin{aligned} &\mathcal{F}^{\perp H}(\mathbf{f}-\hat{\mathbf{f}})\mathcal{F}(\mathbf{f}-\hat{\mathbf{f}}) = 0 \\ \implies &(\mathcal{F}^{\perp H}(\mathbf{f}) - \mathcal{F}^{\perp H}(\hat{\mathbf{f}}))(\mathcal{F}(\mathbf{f}) - \mathcal{F}(\hat{\mathbf{f}})) = 0 \\ \implies &\mathcal{F}^{\perp H}(\mathbf{f})\mathcal{F}(\hat{\mathbf{f}}) = -\mathcal{F}^{\perp H}(\hat{\mathbf{f}})\mathcal{F}(\mathbf{f}) \end{aligned} \tag{26}$$

Further, under the assumption of small noise, $\mathcal{F}^{\perp H}(\hat{\mathbf{f}})\mathbf{n} \approx \mathcal{F}^{\perp H}(\mathbf{f})\mathbf{n}$. From (25), (26) and the small noise assumption, we can write the LS criterion as,

$$\min_{\hat{\mathbf{f}}} ||\mathcal{F}^{\perp H}\mathcal{F}(\hat{\mathbf{f}})\mathbf{h} - \mathcal{F}^{\perp H}\mathbf{n}||^2 = \min_{\hat{\mathbf{f}}} ||\mathcal{F}^{\perp H}\mathcal{H}\hat{\mathbf{f}} - \mathcal{F}^{\perp H}\mathbf{n}||^2. \tag{27}$$

At this point, note that $\mathcal{F}^{\perp H}\mathcal{H}$ is not full rank. This can be seen as follows.

$$\mathcal{F}^{\perp H}\mathcal{H}\mathbf{f} = \mathcal{F}^{\perp H}\mathcal{F}\mathbf{h} = \mathbf{0}. \tag{28}$$

Hence, to regularize the problem, we need to add constraints. Consider a linear constraint of the form

$$\hat{\mathbf{f}}^H\mathbf{g} = c. \tag{29}$$

Let $\hat{\mathbf{f}} = \begin{bmatrix}\mathbf{g} & \mathcal{V}\end{bmatrix}\begin{bmatrix}\alpha_1 \\ \alpha_{2:M}\end{bmatrix}$, where $\mathcal{V}$ is the orthonormal complement to the vector $\mathbf{g}$. The notation $\alpha_{2:M}$ indicates the elements 2 till M of the vector α. α_1 denotes the first element of the same vector. Now, using the constraint $\hat{\mathbf{f}}^H\mathbf{g} = \begin{bmatrix}\alpha_1^* & \alpha_{2:M}^*\end{bmatrix}\begin{bmatrix}\mathbf{g}^H\mathbf{g} \\ \mathcal{V}^H\mathbf{g}\end{bmatrix} = \alpha_1^*\mathbf{g}^H\mathbf{g} = c$. This implies that α_1 is real and is equal to a fixed value $\frac{c}{\mathbf{g}^H\mathbf{g}}$. As a result, (27) may be written in terms of the new parameters $\alpha_{2:M}$ as follows,

$$\begin{aligned} &\min_{\hat{\mathbf{f}}} ||\mathcal{F}^{\perp H}\mathcal{H}\hat{\mathbf{f}} - \mathcal{F}^{\perp H}\mathbf{n}||^2 = \\ &\min_{\alpha_{2:M}} ||\mathcal{F}^{\perp H}\mathcal{H}\mathcal{V}\alpha_{2:M} - \mathcal{F}^{\perp H}\mathbf{n} + \mathcal{F}^{\perp H}\mathcal{H}\mathbf{g}\frac{c}{\mathbf{g}^H\mathbf{g}}||^2 \end{aligned} \tag{30}$$

The MSE for this minimization is straightforward and may be obtained as

$$\begin{aligned} &\text{MSE}_{\text{LS}} = \\ &\sigma^2\mathcal{V}(\mathcal{V}^H\mathbf{A}^H\mathbf{A}\mathcal{V})^{-1}\mathcal{V}^H\mathbf{A}^H\mathbf{B}\mathbf{A}\mathcal{V}(\mathcal{V}^H\mathbf{A}^H\mathbf{A}\mathcal{V})^{-1}\mathcal{V}^H. \end{aligned} \tag{31}$$

Here, $\mathbf{A} = \mathcal{F}^{\perp H}\mathcal{H}$ and $\mathbf{B} = \mathcal{F}^{\perp H}\mathcal{F}^{\perp}$. In comparison, the CRB from (23) is given by

$$\begin{aligned} \text{CRB} &= \sigma^2\mathcal{V}_f(\mathcal{V}_f^H\mathcal{H}^H\mathcal{F}^{\perp}(\mathcal{F}^{\perp H}\mathcal{F}^{\perp})^{-1}\mathcal{F}^{\perp H}\mathcal{H}\mathcal{V}_f)^{-1}\mathcal{V}_f^H \\ &= \sigma^2\mathcal{V}_f(\mathcal{V}_f^H\mathbf{A}^H\mathbf{B}^{-1}\mathbf{A}\mathcal{V}_f)^{-1}\mathcal{V}_f^H, \end{aligned} \tag{32}$$

where we used $\mathcal{P}^{\perp}_{\mathcal{F}} = \mathcal{P}_{\mathcal{F}^{\perp}} = \mathcal{F}^{\perp}(\mathcal{F}^{\perp H}\mathcal{F}^{\perp})^{-1}\mathcal{F}^{\perp H}$. Further, note that the matrix $\mathcal{V}_f$ is a matrix that has to span the orthogonal complement of the derivative of the constraint. In our linear constraint (29), $\frac{\partial \mathcal{C}(f)}{\partial \mathbf{f}^*} = \mathbf{g}$. Hence, the matrix $\mathcal{V}$ is the same as the desired matrix $\mathcal{V}_f$.

To understand the relation between (31) and (32), let us write down the expression for a weighted least squares with weight matrix $\mathbf{W}$.

$$\begin{aligned} &\text{MSE}_{\text{WLS}} = \sigma^2\mathcal{V}\mathbf{U}\mathcal{V}^H, \\ &\mathbf{U} = (\mathcal{V}^H\mathbf{A}^H\mathbf{W}\mathbf{A}\mathcal{V})^{-1}\mathcal{V}^H\mathbf{A}^H\mathbf{W}\mathbf{B}\mathbf{W}\mathbf{A}\mathcal{V}(\mathcal{V}^H\mathbf{A}^H\mathbf{W}\mathbf{A}\mathcal{V})^{-1}. \end{aligned} \tag{33}$$

Now, if we choose the weighting factor $\mathbf{W}$ as $\mathbf{B}^{-1}$, we get back the same expression as that of the CRB. This implies, that to bring down the error variance to that of the CRB, a weighted least squares is called for. Thus, we establish that the key difference between the CRB and that of the MSE of the LS is in the lack of weighting factor $(\mathcal{F}^{\perp H}\mathcal{F}^{\perp})^{-1}$, which is intuitively appealing as this weighting factor is nothing but the inverse covariance of the coloured noise in $\mathcal{F}^{\perp H}\mathbf{y}$.

The expression here covers the linear constraint as was given in II. For the FCC, we need to choose $\mathbf{g} = \mathbf{e}_1$. In this case $\hat{\mathbf{f}}^H\mathbf{g} = c$ chooses the first element of $\hat{\mathbf{f}}$ to be known and the matrix $\mathcal{V}$ is composed of the rest of the columns of the Identity matrix. Now, the NPC constraint may also be approximated by a linear constraint in the vicinity of the true $\mathbf{f}$. To see this, first note that the phase part of the NPC constraint, as given in (11), is already a linear constraint where in (29), we need to choose $\mathbf{g} = \mathbf{f}$. In the vicinity of the true $\mathbf{f}$, $\hat{\mathbf{f}}^H\hat{\mathbf{f}}$ may also be approximated by $\hat{\mathbf{f}}^H\mathbf{f}$, and hence the choice of $\mathbf{g} = \mathbf{f}$ serves as a linear approximation for the norm and phase part of the NPC constraint. With this choice, the matrix $\mathcal{V}$ spans the orthogonal complement of the space of $\mathbf{f}$.

V. SIMULATIONS

We perform simulations assuming an i.i.d. Gaussian channel across the antennas. The Tx and Rx calibration parameters for the BS antennas are assumed to have random phases uniformly distributed over $[-\pi, \pi]$ and amplitudes uniformly distributed in the range $[1-\delta, 1+\delta]$. SNR is defined as the ratio of the average received signal power across channel realizations at an antenna and the noise power at that antenna. First, we make a comparison between different grouping schemes on the basis of their CRB. Consider a system with $M = 64$ antennas. With single antenna grouping, the minimum number of channel uses required for calibration is M. Hence, we assume that

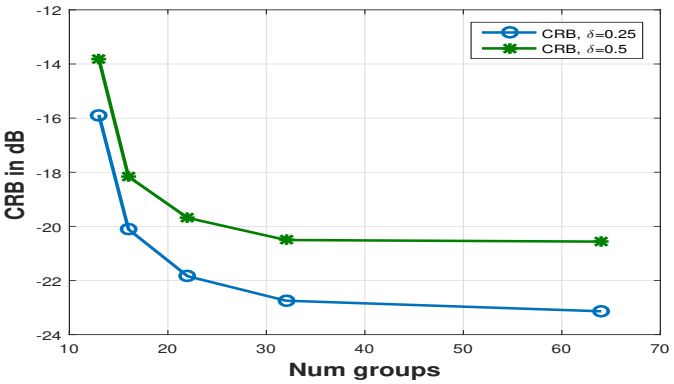

Fig. 1. Comparison of CRB with different antenna group sizes for $M = 64$ antenna scenario.

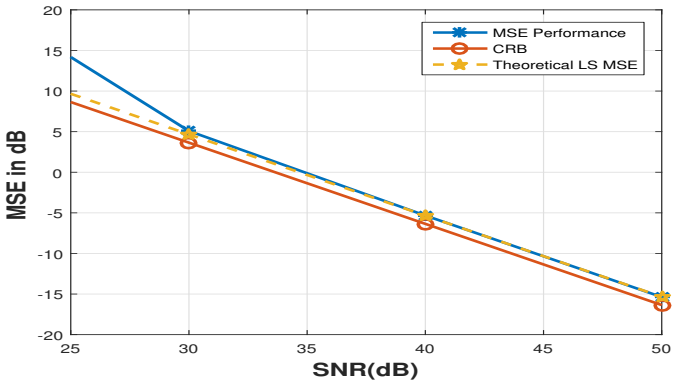

Fig. 2. Comparison of simulation MSE for an $M = 32$ antenna scenario with the theoretical expression for MSE for the FCC constraint.

there are M channel uses available for reciprocity calibration irrespective of the antenna grouping strategy used. At the same time, the minimum number of channel uses required for calibration is ([1]) given by $G(G-1) \geq (2M-1)$, giving $G = 12$. In this case, the number of antennas in each group would be 5 or 6 if we distribute the antennas as equitably as possible across the groups. In this scheme every antenna group can transmit pilots for multiple channel uses during the overall M channel uses required for the single antenna grouping. In Figure 1, we plot the CRB for different group sizes between the slowest single antenna grouping (Num groups = $M = 64$) and the fastest multiple antenna grouping. Whenever a faster grouping is used, antenna groups can transmit pilots during multiple channel uses which further scales down the CRB by that pilot use factor. The CRB with NPC constraint is averaged over multiple channel realizations and SNR = 30dB. Two separate curves are displayed for $\delta = 0.25$ and $\delta = 0.5$. From this plot, we can infer that given a fixed number of available channel uses, it is more beneficial to use the smallest possible size for antenna groups. In the case where there are a sufficient number of channel uses available, this would imply a preference for the single antenna grouping over multiple antenna grouping. However, it is also noteworthy that a grouping with two antennas per group also has a performance comparable to that of single antenna grouping while requiring only half the time for calibration. Fig. 2 compares the MSE that was derived for LS with the actual performance obtained in simulations using FCC. The results have been averaged across multiple channel and noise realizations. A large variation in calibration values ($\delta = 0.95$) is chosen for this simulation so as to enhance the gap between the LS performance and the CRB. The simulated MSE is seen to closely match the theoretical MSE. The plot also shows the CRB for this case which forms a lower bound, as expected.

VI. Conclusion

We first derived a generalized CRB that is applicable to both the coherent and non-coherent scenarios. Next, we derived a simple and intuitive relationship between the MSE of the LS algorithms to the CRB for both NPC and FCC. We also compared the performance of different grouping strategies for an i.i.d Rayleigh channel which indicates that given sufficient number of channel uses, a single antenna per group shows superior MSE performance.

Acknowledgment

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, ST Microelectronics, Symantec, SAP, Monaco Telecom, iABG, by the projects DUPLEX (French ANR) and MASS-START (French FUI).

References

[1] X. Jiang, A. Decurninge, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock, and L. Deneire, "A framework for over-the-air reciprocity calibration for tdd massive mimo systems," *IEEE TWC*, Sept 2018.

[2] K. Gopala and D. Slock, "Optimal algorithms and CRB for reciprocity calibration in Massive MIMO," in *IEEE ICASSP*, Apr. 2018.

[3] C. Shepard, N. Yu, H.and Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. ACM Mobicom*, Istanbul, Turkey, 2012.

[4] R. Rogalin, O. Bursalioglu, H. Papadopoulos, G. Caire, A. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1815–1831, Apr. 2014.

[5] J. Vieira, F. Rusek, and F. Tufvesson, "Reciprocity calibration methods for massive MIMO based on antenna coupling," in *2014 IEEE Global Communications Conference*, Dec 2014.

[6] H. Papadopoulos, O. Y. Bursalioglu, and G. Caire, "Avalanche: Fast RF calibration of massive arrays," in *Proc. IEEE Global Conf. on Signal and Information Process. (GlobalSIP)*, Washington, DC, USA, Dec. 2014.

[7] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu, and F. Tufvesson, "Reciprocity Calibration for Massive MIMO: Proposal, Modeling and Validation," *IEEE Trans. Wireless Commun.*, May 2017.

[8] H. Wei, W. D., and X. You, "Reciprocity of mutual coupling for TDD massive MIMO systems," in *Proc. WCSP*, Nanjing, China, 2015.

[9] M. Petermann, M. Stefer, F. Ludwig, D. Wübben, M. Schneider, S. Paul, and K. Kammeyer, "Multi-user pre-processing in multi-antenna OFDM TDD systems with non-reciprocal transceivers," *IEEE T. Comm.*, 2013.

[10] X. Jiang, M. Čirkić, F. Kaltenberger, E. G. Larsson, L. Deneire, and R. Knopp, "MIMO-TDD reciprocity and hardware imbalances: experimental results," in *Proc. ICC*, UK, 2015.

[11] R1-091794, "Hardware calibration requirement for dual layer beamforming," Huawei, 3GPP RAN1 #57, San Francisco, USA, May 2009.

[12] R1-091752, "Performance study on Tx/Rx mismatch in LTE TDD dual-layer beamforming," Nokia, Nokia Siemens Networks, CATT, ZTE, 3GPP RAN1 #57, San Francisco, USA, May 2009.

[13] X. Jiang, A. Decurninge, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock, and L. Deneire, "A Framework for Over-the-air Reciprocity Calibration for TDD Massive MIMO Systems," *ArXiv 1710.10830*.

[14] E. de Carvalho and D. Slock, "Semi–Blind Methods for FIR Multichannel Estimation," in *Signal processing advances in wireless communications*, Y. H. G. Giannakis, P. Stoica and L. Tong, Eds. Prentice Hall, 2000, ch. 7. [Online]. Available: http://www.eurecom.fr/publication/469

[15] E. de Carvalho, S. Omar, and D. Slock, "Performance and complexity analysis of blind FIR channel identification algorithms based on deterministic maximum likelihood in SIMO systems," *CSSP*, Aug. 2012. [Online]. Available: www.eurecom.fr/fr/publication/3790

[16] E. de Carvalho and D. Slock, "Cramér-Rao bounds for blind multichannel estimation," arXiv:1710.01605.