OPTIMAL ALGORITHMS AND CRB FOR RECIPROCITY CALIBRATION IN MASSIVE MIMO

Kalyana Gopala , Dirk Slock

EURECOM, Sophia-Antipolis, France
Email: gopala,slock@eurecom.fr

ABSTRACT

Gains from Massive MIMO are crucially dependent on the availability of channel state information at the transmitter which is far too costly if it has to estimated directly. Hence, for a time division duplexing system, this is derived from the uplink channel estimates using the concept of channel reciprocity. However, while the propagation channel is reciprocal, the overall digital channel in the downlink also involves the radio frequency chain which is non-reciprocal. This calls for calibration of the uplink channel with reciprocity calibration parameters to derive the downlink channel estimates. Initial approaches towards estimation of the reciprocity calibration parameters [1, 2] were all based on least squares. An ML estimator and a CRB for the estimators was introduced in [3]. This paper presents a more elegant and accurate CRB expression for a general reciprocity calibration framework. An optimal algorithm based on Variational Bayes is presented and it is compared with existing algorithms.

***Index Terms*—** Masive MIMO, Reciprocity Calibration, CRB

1. INTRODUCTION

Massive MIMO (Multiple Input Multiple Output) requires CSIT (Channel state information at Tx) CSIT acquired using channel reciprocity for a TDD (Time Division Duplexing) system. However, Radio Frequency (RF) components are not reciprocal and we need to calibrate to compensate for this. This calibration is typically achieved by a simple complex scalar multiplication at each transmit antenna.Initial approaches to calibration relied on explicit channel feedback from a user equipment (UE) during the calibration phase to estimate the calibration parameters. This is typically referred to as UE aided calibration. However, what is popular today is to perform the calibration across the antennas of the base station (BS) only and is referred to as internal calibration. [1] gave the first experimental evaluation for a Massive MIMO system with a simple algorithm for internal calibration. Though simple, the performance of this algorithm was limited by the requirement for a strategically placed reference antenna at the BS side. Hence, this estimation algorithm was quickly improved upon by making better use of all the available information in a least squares fashion in [2]. This was further generalized to a weighted least squares minimization in [4]. A faster calibration algorithm was introduced in [5]. Recently, in [3], the authors proposed a Cramer-Rao bound (CRB) and a penalized maximum likelihood estimation approach that performs close to the CRB. In [6], the authors propose a generalized approach towards reciprocity calibration of which the existing estimation techniques are special cases. An important innovation in this paper is the ability to transmit from a group of multiple antennas at every channel use. First, an elegant and concise CRB derivation is presented. Then, to compare with the derivation in [3], an alternative and more conventional CRB derivation is provided for the special case of single antenna groups and is observed to concur with the derivation in [6]. More details may be found in [6, 7].

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, ST Microelectronics, Symantec, SAP, Monaco Telecom, iABG, and by the projects HIGHTS (EU H2020), DUPLEX (French ANR) and MASS-START (French FUI).

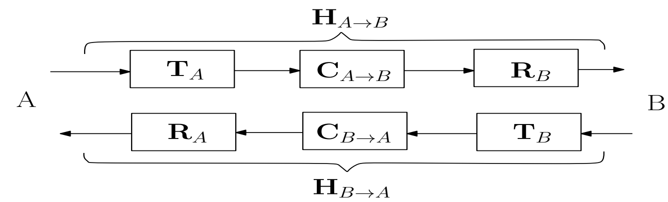

Fig. 1. Reciprocity Model

2. SYSTEM MODEL

Consider a system as in Fig. 1, where A represents a BS and B represents a UE, each containing M_A and M_B antennas, respectively. The channel as observed in the digital domain, $\mathbf{H}_{A\to B}$ and $\mathbf{H}_{B\to A}$ can be represented by,

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

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 RF crosstalk and antenna mutual coupling. It is worth noting that although transmitting and receiving antenna mutual coupling is not generally reciprocal [8], theoretical modeling [9] and experimental results [1, 3, 10] both show that, in practice, RF crosstalk and antenna mutual coupling can 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 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. Thus, 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 $A_1, A_2, \dots, A_G$. 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 in receiving mode. After all G groups have transmitted, the received signal for each resource block of bidirectional transmission between antenna groups i and j is given by

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

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 $1 \le i \neq j \le G$, by eliminating $\mathbf{C}_{i\to j}$ in (3) we have

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

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 $\mathbf{F} = \text{diag}\{\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_G\}$.

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., $\mathbf{F}_i = \text{diag}\{\mathbf{f}_i\}$ and $\mathbf{F} = \text{diag}\{\mathbf{f}\}$. This allows us to vectorize (4) into

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

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

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

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

$$\underbrace{\begin{bmatrix} (\mathbf{Y}_{2\to 1}^T * \mathbf{P}_1^T) & -(\mathbf{P}_2^T * \mathbf{Y}_{1\to 2}^T) & 0 & \dots \\ (\mathbf{Y}_{3\to 1}^T * \mathbf{P}_1^T) & 0 & -(\mathbf{P}_3^T * \mathbf{Y}_{1\to 3}^T) & \dots \\ 0 & (\mathbf{Y}_{3\to 2}^T * \mathbf{P}_2^T) & -(\mathbf{P}_3^T * \mathbf{Y}_{2\to 3}^T) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{(\sum_{j=2}^{G} \sum_{i=1}^{j-1} L_i L_j) \times M}. \tag{7}$$

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

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

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

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

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):

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

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

2) Linear constraint:
$$\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f}) = \hat{\mathbf{f}}^H \mathbf{g} - c = 0\ . \tag{12}$$

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

3. OPTIMAL ESTIMATION AND PERFORMANCE LIMITS

From (3), we have

$$\mathbf{Y}_{i\to j} = \underbrace{\mathbf{R}_j \mathbf{C}_{i\to j} \mathbf{R}_i^T}_{\mathcal{H}_{i\to j}} \mathbf{F}_i \mathbf{P}_i + \mathbf{N}_{i\to j}. \tag{13}$$

We define $\mathcal{H}_{i\to j} = \mathbf{R}_j \mathbf{C}_{i\to j} \mathbf{R}_i^T$ 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 $\mathcal{H}_{i\to j}$ inherits the reciprocity from the channel $\mathbf{C}_{i\to j}$: $\mathcal{H}_{i\to j} = \mathcal{H}_{j\to i}^T$. Upon applying the vectorization operator for each bidirectional transmission between groups i and j, we have, similarly to (6)

$$\text{vec}(\mathbf{Y}_{i\to j}) = (\mathbf{P}_i^T * \mathcal{H}_{i\to j})\, \mathbf{f}_i + \text{vec}(\mathbf{N}_{i\to j}). \tag{14}$$

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

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

Alternatively, (14) and (15) may also be written as

$$\begin{cases} \text{vec}(\mathbf{Y}_{i\to j}) = \left[(\mathbf{F}_i \mathbf{P}_i)^T \otimes \mathbf{I}\right] \text{vec}(\mathcal{H}_{i\to j}) + \text{vec}(\mathbf{N}_{i\to j}) \\ \text{vec}(\mathbf{Y}_{j\to i}^T) = \left[\mathbf{I} \otimes (\mathbf{P}_j^T \mathbf{F}_j)\right] \text{vec}(\mathcal{H}_{i\to j}) + \text{vec}(\mathbf{N}_{j\to i}). \end{cases} \tag{16}$$

Stacking these observations into a vector $\mathbf{y} = \left[\text{vec}(\mathbf{Y}_{1\to 2})^T\, \text{vec}(\mathbf{Y}_{2\to 1}^T)^T\, \text{vec}(\mathbf{Y}_{1\to 3})^T \dots\right]^T$, the above two alternative formulations can be summarized into

$$\mathbf{y} = \mathcal{H}(\mathbf{h}, \mathbf{P})\mathbf{f} + \mathbf{n} = \mathcal{F}(\mathbf{f}, \mathbf{P})\mathbf{h} + \mathbf{n}, \tag{17}$$

where $\mathbf{h} = \left[\text{vec}(\mathcal{H}_{1\to 2})^T\, \text{vec}(\mathcal{H}_{1\to 3})^T\, \text{vec}(\mathcal{H}_{2\to 3})^T \dots\right]^T$, and $\mathbf{n}$ is the corresponding noise vector. The composite matrices $\mathcal{H}$ and $\mathcal{F}$ are given by,

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

$$\mathcal{F}(\mathbf{f}, \mathbf{P}) = \begin{bmatrix} \mathbf{P}_1^T \mathbf{F}_1 \otimes \mathbf{I} & 0 & 0 & 0 & \dots \\ \mathbf{I} \otimes \mathbf{P}_2^T \mathbf{F}_2 & 0 & 0 & 0 & \dots \\ 0 & \mathbf{P}_1^T \mathbf{F}_1 \otimes \mathbf{I} & 0 & 0 & \dots \\ 0 & \mathbf{I} \otimes \mathbf{P}_3^T \mathbf{F}_3 & 0 & 0 & \dots \\ 0 & 0 & \mathbf{P}_2^T \mathbf{F}_2 \otimes \mathbf{I} & 0 & \dots \\ 0 & 0 & \mathbf{I} \otimes \mathbf{P}_3^T \mathbf{F}_3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{18}$$

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.

[1] With matrices $\mathbf{A}$ and $\mathbf{B}$ partitioned into columns, $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_M\end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_M\end{bmatrix}$ where $\mathbf{a}_i$ and $\mathbf{b}_i$ are column vectors for $i \in 1 \dots M$, then, $\mathbf{A} * \mathbf{B} = \begin{bmatrix}\mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \dots & \mathbf{a}_M \otimes \mathbf{b}_M\end{bmatrix}$ [13].

3.1. Cramér-Rao bound

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 (17) as

$$\mathbf{J} = \frac{1}{\sigma^2}\begin{bmatrix}\mathcal{H}^H \\ \mathcal{F}^H\end{bmatrix}\begin{bmatrix}\mathcal{H} & \mathcal{F}\end{bmatrix}. \tag{19}$$

The computation of the CRB requires $\mathbf{J}$ to be non-singular. However, for the problem at hand, $\mathbf{J}$ is inherently singular. In fact, the calibration factors (and the auxiliary channel) can only be estimated up to a complex scale factor since the received data (17) involves the product of the channel and the calibration factors, $\mathcal{H}\mathbf{f} = \mathcal{F}\mathbf{h}$. As a result the FIM has the following null space [16], [17]

$$\mathbf{J}\begin{bmatrix}\mathbf{f} \\ -\mathbf{h}\end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix}\mathcal{H} & \mathcal{F}\end{bmatrix}^H(\mathcal{H}\mathbf{f} - \mathcal{F}\mathbf{h}) = \mathbf{0}. \tag{20}$$

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. Another issue is that we are mainly interested in the CRB for $\mathbf{f}$, the parameters of interest, 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 (19). Incorporating the effect of the constraint (9) on $\mathbf{f}$, we can derive from [17] the following constrained CRB for $\mathbf{f}$

$$\text{CRB}_{\mathbf{f}} = \sigma^2\mathcal{V}_{\mathbf{f}}\left(\mathcal{V}_{\mathbf{f}}^H\mathcal{H}^H\mathcal{P}_{\mathcal{F}}^{\perp}\mathcal{H}\mathcal{V}_{\mathbf{f}}\right)^{-1}\mathcal{V}_{\mathbf{f}}^H \tag{21}$$

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. Note that in some group calibration scenarios, $\mathcal{F}^H\mathcal{F}$ can be singular (i.e, $\mathbf{h}$ could be not identifiable even if $\mathbf{f}$ is identifiable or even known). 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., $\mathcal{P}_{\mathcal{V}_{\mathbf{f}}} = \mathcal{P}^{\perp}_{\frac{\partial\mathcal{C}}{\partial\mathbf{f}^*}}$.

It is shown in [16], [17], [18] that a choice of constraints such that their linearized version $\frac{\partial\mathcal{C}}{\partial\mathbf{f}^*}$ fills up the null space of the FIM results in the lowest CRB, while not adding information in subspaces where the data provides information. One such choice is the set (10), (11) (NPC). Another choice is (12) with $\mathbf{g} = \mathbf{f}$. With such constraints, $\frac{\partial\mathcal{C}}{\partial\mathbf{f}^*} \sim \mathbf{f}$ which spans the null space of $\mathcal{H}^H\mathcal{P}_{\mathcal{F}}^{\perp}\mathcal{H}$. The CRB then corresponds to the pseudo inverse of the FIM and (21) becomes $\text{CRB}_{\mathbf{f}} = \sigma^2\left(\mathcal{H}^H\mathcal{P}_{\mathcal{F}}^{\perp}\mathcal{H}\right)^{\dagger}$. If the FCC constraint is used instead (i.e., (12) with $\mathbf{g} = \mathbf{e}_1$, $c = 1$), the corresponding CRB is (21) where $\mathcal{V}_{\mathbf{f}}$ corresponds now to an identity matrix without the first column (and hence its column space is the orthogonal complement of that of $\mathbf{e}_1$).

Note that [3] also addresses the CRB for a scenario where transmission happens one antenna at a time. The relative calibration factors are derived from the absolute Tx and Rx side calibration parameters, which become identifiable because a model is introduced for the internal propagation channel. In this Gaussian prior the mean is taken as the line of sight (LoS) component (distance induced delay and attenuation) and complex Gaussian non-LoS (NLOS) components are contributing to the covariance of this channel as a scaled identity matrix. The scale factor is taken 60dB below the mean channel power. This implies an almost deterministic prior for the (almost known) channel and would result in underestimation of the CRB, as noted in [3, Sec. III-E-2]. Hence, we rederive the CRB for this case in a more conventional fashion assuming an unknown channel $\mathbf{h}$. Assuming that the first calibration coefficient is $f_1 = t_1/r_1 = 1$ (FCC), we choose $t_1 = r_1 = 1$, where t_i, r_i correspond to the transmit and receive calibration parameters for antenna i. Note that as there are three unknown parameter sets that appear in product form in $\mathbf{Y}$, we can choose the scale factor for two. Then, $\theta^T = [\underbrace{t_2\, r_2 \ldots t_M\, r_M}_{\theta_1}\ \underbrace{c_{1\to 2}\, c_{1\to 3} \ldots c_{M-1\to M}}_{\theta_2}]$, where $c_{i\to j}$ corresponds to the prop. channel from antenna i to antenna j. Then,

$$\text{CRB}_{\mathbf{f}'} = \frac{\partial\mathbf{f}'}{\partial\theta^T}\mathbf{J}_\theta^{-1}\left(\frac{\partial\mathbf{f}'}{\partial\theta^T}\right)^H, \quad \mathbf{J}_\theta = \frac{1}{\sigma^2}\,\mathrm{E}\,\frac{\partial\mu_{\mathbf{y}}^H}{\partial\theta^*}\left(\frac{\partial\mu_{\mathbf{y}}^H}{\partial\theta^*}\right)^H$$

$$\mu_{\mathbf{y}} = [r_2c_{1\to 2}t_1\ \ r_1c_{1\to 2}t_2\ \ r_3c_{1\to 2}t_1\ \ \ldots]^T$$

$$\frac{\partial\mu_{\mathbf{y}}^H}{\partial\theta_2^*} = \begin{bmatrix} r_2^*t_1^* & r_1^*t_2^* & 0 & 0 & \cdots \\ 0 & 0 & r_2^*t_1^* & r_1^*t_2^* & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\frac{\partial\mathbf{f}'}{\partial\theta_1} = \begin{bmatrix} \frac{1}{r_2} & 0 & 0 & 0 & \cdots \\ 0 & \frac{-t_2}{r_2^2} & 0 & 0 & \cdots \\ 0 & 0 & \frac{1}{r_3} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad \frac{\partial\mathbf{f}'}{\partial\theta_2} = \mathbf{0}. \tag{22}$$

Here, σ^2 is the noise variance, $\mathbf{f}' = f_{2:M}$ and $\frac{\partial\mu_{\mathbf{y}}^H}{\partial\theta_1^*} =$

$$\begin{bmatrix} 0 & r_1^*c_{1\to 2}^* & 0 & 0 & r_3^*c_{2\to 3}^* & 0 & \cdots \\ c_{1\to 2}^*t_1^* & 0 & 0 & 0 & 0 & c_{2\to 3}^*t_3^* & \cdots \\ 0 & 0 & 0 & r_1^*c_{1\to 3}^* & 0 & r_2^*c_{2\to 3}^* & \cdots \\ 0 & 0 & c_{1\to 3}^*t_1^* & 0 & c_{2\to 3}^*t_2^* & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Now, for the FCC, (21) becomes

$$\text{CRB}_{\mathbf{f}'} = \sigma^2\left([\mathcal{H}^H\mathcal{P}_{\mathcal{F}}^{\perp}\mathcal{H}]_{2:M,2:M}\right)^{-1} = \frac{\partial\mathbf{f}'}{\partial\phi^T}\mathbf{J}_\phi^{-1}\left(\frac{\partial\mathbf{f}'}{\partial\phi^T}\right)^H \tag{23}$$

where $\phi = [\mathbf{f}^T\ \mathbf{h}^T]^T$ and $\mathbf{J}_\phi$ is defined similarly to $\mathbf{J}_\theta$. Now, since ϕ and θ are both valid parameterizations of the estimated channels, the CRBs in (22) and (23) and hence (21) are identical.

3.2. Variational Bayes (VB) Estimation

In VB, a Bayesian estimate is obtained by computing an approximation to the posterior distribution of the parameters $\mathbf{h}, \mathbf{f}$ with priors $\mathbf{f} \sim \mathcal{CN}(0, \alpha^{-1}\mathbf{I_M})$, $\mathbf{h} \sim \mathcal{CN}(0, \beta^{-1}\mathbf{I_{N_h}})$ and α, β are assumed to have themselves a uniform prior. N_h is the number of elements in $\mathbf{h}$. This approximation, called the variational distribution, is chosen to minimize the Kullback-Leibler distance between the true posterior distribution $p(\mathbf{h}, \mathbf{f}, \alpha, \beta|\mathbf{y})$ and a factored variational distribution

$$q_{\mathbf{h}}(\mathbf{h})\, q_{\mathbf{f}}(\mathbf{f})\, q_\alpha(\alpha)\, q_\beta(\beta). \tag{24}$$

The factors can be obtained in an alternating fashion as,

$$\ln(q_{\psi_i}(\psi_i)) = <\ln p(\mathbf{y}, \mathbf{h}, \mathbf{f}, \alpha, \beta)>_{k\neq i} + c_i, \tag{25}$$

where ψ_i refers to the i^{th} block of $\psi = [\mathbf{h}, \mathbf{f}, \alpha, \beta]$ and $<>_{k\neq i}$ represents the expectation operator over the distributions q_{ψ_k} for all $k \neq i$. c_i is a normalizing constant. The log likelihood,

$$\begin{aligned} &\ln p(\mathbf{y}, \mathbf{h}, \mathbf{f}, \alpha, \beta) = \ln p(\mathbf{y}|\mathbf{h}, \mathbf{f}, \alpha, \beta) + \ln p(\mathbf{f}|\alpha) + \ln p(\mathbf{h}|\beta) \\ &= -N_y\ln\sigma^2 - \frac{1}{\sigma^2}\|\mathbf{y} - \mathcal{H}\mathbf{f}\|^2 + M\ln\alpha - \alpha\|\mathbf{f}\|^2 \\ &+ N_h\ln\beta - \beta\|\mathbf{h}\|^2 + c. \end{aligned} \tag{26}$$

Here, N_y refers to the number of elements in $\mathbf{y}$ and c is a constant. We shall assume $\sigma^2 = 1$, which is equivalent to considering α, β as relative to σ^2. It is straightforward to see that proceeding as in (25),

α, β would have a Gamma distribution with mean $<\alpha>= \frac{M}{<\|\mathbf{f}\|^2>}$ and $<\beta>= \frac{N_h}{<\|\mathbf{h}\|^2>}$. On the other hand (taking only relevant terms),

$$\ln q_{\mathbf{f}}(\mathbf{f}) = \mathbf{f}^H <\mathcal{H}^H> \mathbf{y} - \mathbf{f}^H <\mathcal{H}^H\mathcal{H}> \mathbf{f} - <\alpha> \mathbf{f}^H\mathbf{f}$$
$$\ln q_{\mathbf{h}}(\mathbf{h}) = \mathbf{h}^H <\mathcal{F}^H> \mathbf{y} - \mathbf{h}^H <\mathcal{F}^H\mathcal{F}> \mathbf{h} - <\beta> \mathbf{h}^H\mathbf{h} \tag{27}$$

This implies that $\mathbf{f} \sim \mathcal{CN}(\hat{\mathbf{f}}, \mathbf{C}_{\tilde{f}\tilde{f}})$ and $\mathbf{h} \sim \mathcal{CN}(\hat{\mathbf{h}}, \mathbf{C}_{\tilde{h}\tilde{h}})$. The overall algorithm may now be summarized as in Algorithm 1.

Algorithm 1 Variational Bayes Estimation of calibration parameters

Initialization: Initialize $\hat{\mathbf{f}}$ using existing calibration methods. Use this estimate to determine $\hat{\mathbf{h}}, <\alpha>, <\beta>$.
repeat
  $<\mathcal{H}^H\mathcal{H}>= \mathcal{H}^H(\hat{\mathbf{h}})\mathcal{H})(\hat{\mathbf{h}})+ <\mathcal{H}^H(\tilde{\mathbf{h}})\mathcal{H}(\tilde{\mathbf{h}})>$
  $\hat{\mathbf{f}} = (<\mathcal{H}^H\mathcal{H}> + <\alpha> \mathbf{I})^{-1}\mathcal{H}^{\mathbf{H}}\mathbf{y}$
  $\mathbf{C}_{\tilde{f}\tilde{f}} = (<\mathcal{H}^H\mathcal{H}> + <\alpha> \mathbf{I})^{-1}$
  $<\mathcal{F}^H\mathcal{F}>= \mathcal{F}^H(\hat{\mathbf{f}})\mathcal{F}(\hat{\mathbf{f}})+ <\mathcal{F}^H(\tilde{\mathbf{f}})\mathcal{F}(\tilde{\mathbf{f}})>$
  $\hat{\mathbf{h}} = (<\mathcal{F}^H\mathcal{F}> + <\beta> \mathbf{I})^{-1}\mathcal{F}^{\mathbf{H}}\mathbf{y}$
  $\mathbf{C}_{\tilde{h}\tilde{h}} = (<\mathcal{F}^H\mathcal{F}> + <\beta> \mathbf{I})^{-1}$
  $<\alpha>= \frac{M}{<\|\mathbf{f}\|^2>}, <\|\mathbf{f}\|^2>= \hat{\mathbf{f}}^H\hat{\mathbf{f}} + tr\{\mathbf{C}_{\tilde{f}\tilde{f}}\}$.
  $<\beta>= \frac{N_h}{<\|\mathbf{h}\|^2>}, <\|\mathbf{h}\|^2>= \hat{\mathbf{h}}^H\hat{\mathbf{h}} + tr\{\mathbf{C}_{\tilde{h}\tilde{h}}\}$.
until convergence.

When $G = M$, $\mathbf{C}_{\tilde{f}\tilde{f}}$ and $\mathbf{C}_{\tilde{h}\tilde{h}}$ are diagonal and $<\mathcal{F}^H(\tilde{\mathbf{f}})\mathcal{F}(\tilde{\mathbf{f}})>$, $<\mathcal{H}^H(\tilde{\mathbf{h}})\mathcal{H}(\tilde{\mathbf{h}})>$ can be computed easily (diagonal). However, when $G < M$, these matrices are block diagonal. To simplify the computation, we propose the following,

$$\mathbf{C}_{\tilde{f}\tilde{f}} \approx \frac{tr\{(<\mathcal{H}^H\mathcal{H}> + <\alpha> \mathbf{I})^{-1}\}}{M}\mathbf{I}_M$$
$$\mathbf{C}_{\tilde{h}\tilde{h}} \approx \frac{tr\{(<\mathcal{F}^H\mathcal{F}> + <\beta> \mathbf{I})^{-1}\}}{N_h}\mathbf{I}_{N_h}. \tag{28}$$

We call this approach EC-VB (Expectation consistent [19] VB). Note here that by forcing the matrices $\mathbf{C}_{\tilde{f}\tilde{f}}$, $\mathbf{C}_{\tilde{h}\tilde{h}}$ to zero and α, β to zero, this algorithm reduces to the Alternating Maximum Likelihood (AML) algorithm [14, 15] which iteratively maximizes the likelihood by alternating between the desired parameters $\mathbf{f}$ and the nuisance parameters $\mathbf{h}$ for the formulation (17). The penalized ML method used in [3] uses quadratic regularization terms for both $\mathbf{f}$ and $\mathbf{h}$ which can be interpreted as Gaussian priors and which may improve estimation in ill-conditioned cases. In our case, we arrive at a similar solution from the VB perspective and more importantly, the regularization terms are optimally tuned.

4. SIMULATIONS

In this section, we assess numerically the performance of various calibration algorithms and also compare them against their CRBs. 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. In Fig. 2, $\delta = 0.5$. We consider transmit schemes that transmit from one antenna at a time ($G = M$) and compare their MSE performance with the CRB. The MSE with FCC for Argos, Rogalin [2] and the AML method in Algorithm 1 is plotted. The curves are generated over one realization of an i.i.d. Rayleigh channel and known first coefficient constraint is used. As expected, the Rogalin method improves over Argos by using all the bi-directional received data. AML outperforms the Rogalin performance at low SNR. These curves are compared with the CRB derived in 3.1 for the FCC case and it can

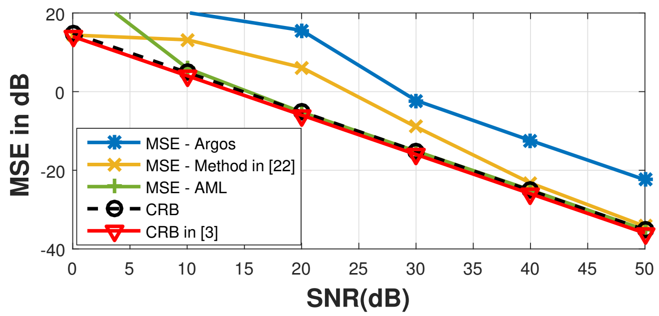

Fig. 2. Comparison of single antenna transmit schemes with the CRB ($G = M = 16$).

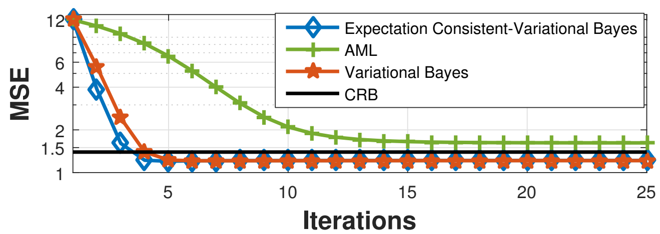

Fig. 3. Convergence of the various iterative schemes for $M = 16$ and $G = M$.

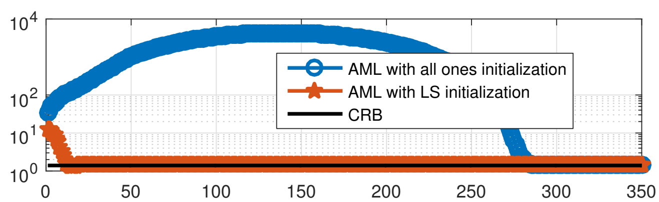

Fig. 4. Slow convergence of the iterative AML scheme for $M = 16$ and $G = M$ when initialized with all ones for the calibration factors.

be seen that the AML curve overlaps with the CRB at higher SNRs. Also plotted is the CRB as given in [3] assuming the internal propagation channel is fully known (the mean is known and the variance is negligible) and a (small) underestimation of the MSE can be observed as expected. Though not explicitly shown to reduce cluttering the figure, the two different CRB derivations in 3.1 lead to exactly same results, further validating our CRB derivations.

Next, we compare the convergence of the proposed iterative algorithms when the calibration parameters are generated with $\delta = 0.25$ in Figures 3 and 4. The curves in Figure 3 are generated for a single channel and calibration parameter realization and averaged over 200 noise realizations. We clearly see that the VB methods (initialized by LS) are far superior to the AML in terms of both MSE achievable and speed of convergence. Figure 4 shows the importance of a good initialization for the iterative algorithms. In this case, when the AML is initialized as a vector of all 1's, a very large number of iterations is necessary for convergence.

5. CONCLUSION

In this paper, we came up with a simple and elegant derivation of the CRB for a general calibration framework that includes as subsets all existing calibration techniques. For the case of groups involving a single antenna, the conventional CRB derivation assuming first coefficient known has also been provided. An optimal estimation algorithm based on VB is also introduced along with it's variants. All these techniques have been compared via simulations in terms of both MSE performance and speed of convergence.

6. REFERENCES

[1] 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 Intern. Conf. Mobile Computing and Netw. (Mobicom)*, Istanbul, Turkey, Aug. 2012.

[2] 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.

[3] Joao Vieira, Fredrik Rusek, Ove Edfors, Steffen Malkowsky, Liang Liu, and Fredrik Tufvesson, "Reciprocity Calibration for Massive MIMO: Proposal, Modeling and Validation," *IEEE Trans. Wireless Commun.*, May 2017.

[4] 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.

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

[6] 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," *Accepted, IEEE Trans. Wireless Commun.*, 2018.

[7] 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 e-prints*, Oct. 2017.

[8] H. Wei, Wang D., and X. You, "Reciprocity of mutual coupling for TDD massive MIMO systems," in *Proc. Intern. Conf. on Wireless Commun. and Signal Process. (WCSP)*, Nanjing, China, Oct. 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 Trans. Commun.*, Sep. 2013.

[10] X. Jiang, M. Čirkić, F. Kaltenberger, Erik G Larsson, L. Deneire, and R. Knopp, "MIMO-TDD reciprocity and hardware imbalances: experimental results," in *Proc. IEEE Intern. Conf. on Commun. (ICC)*, London, United Kingdom, Jun. 2015.

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

[13] CG Khatri and C Radhakrishna Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhyā: The Indian Journal of Statistics, Series A*, 1968.

[14] E. de Carvalho and D. Slock, "Semi–Blind Methods for FIR Multichannel Estimation," in *Signal processing advances in wireless communications*, Y. Hua G. Giannakis, P. Stoica and L. Tong, Eds., chapter 7. Prentice Hall, 2000.

[15] E. de Carvalho, S.M. Omar, and D. Slock, "Performance and complexity analysis of blind FIR channel identification algorithms based on deterministic maximum likelihood in SIMO systems," *Circuits, Systems, and Signal Processing*, vol. 34, no. 4, Aug. 2012.

[16] E. de Carvalho and D. Slock, "Blind and semi-blind FIR multichannel estimation: (Global) identifiability conditions," *IEEE Trans. Sig. Proc.*, Apr. 2004.

[17] E. de Carvalho and D. Slock, "Cramér-Rao bounds for blind multichannel estimation," arXiv:1710.01605 [cs.IT], 2017.

[18] E. de Carvalho, J.M. Cioffi, and D. Slock, "Cramér-Rao bounds for blind multichannel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Francisco, CA, USA, Nov. 2000.

[19] M. Opper and O. Winther, "Expectation Consistent Approximate Inference," *J. Mach. Learn. Res.*, vol. 6, Dec. 2005.