This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619086.

D2.4

Analysis of non-reciprocity impact and possible solutions

Project number:	619086
Project acronym:	MAMMOET
Project title:	MAMMOET: Massive MIMO for Efficient Transmission
Start date of the project:	1st January, 2014
Duration:	36 months
Programme:	FP7/2007-2013

Deliverable type:	Report
Deliverable reference number:	ICT-619086 / D2.4
Work package contributing to the deliverable:	WP2
Due date:	31 August, 2015
Actual submission date:	1st September, 2015

Responsible organisation:	imec
Editor:	André Bourdoux, Liesbet Van der Perre
Dissemination level:	PU
Revision:	1.0

Abstract:	This deliverable reports on the impact study of non-reciprocal transceivers on the channel state acquisition, and consequently the performance of Massive MIMO systems. Two possible mitigation approaches are introduced, which may both be adequate depending on the system set-up.
Keywords:	Non-reciprocity, channel state information, RF front-ends, antennas, calibration

Editor

André Bourdoux, Liesbet Van der Perre (imec)

Contributors (ordered according to beneficiary numbers)

André Bourdoux, Liesbet Van der Perre, Claude Desset (imec)

Disclaimer

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° ICT-619086 (MAMMOET).

Executive Summary

Channel state information (CSI) is required to perform effective Massive MIMO (MaMi) communication. It is highly preferred to acquire the CSI based on uplink estimation to reduce the overhead. Indeed, typically the number of antennas at the base-station is significantly larger than the number of user terminals. Consequently, a much larger number of channels can be measured simultaneously in the uplink. When using this uplink-CSI in the pre-coder for the downlink transmission, the non-reciprocity of transceiver front-ends can impact system performance and ideally should be compensated for.

In this deliverable, first the problem is analysed in more details as the structure of the problem reveals potential calibration approaches to resolve the problem. Especially, the fact that only the transceivers at the base-station should be taken care of is an important feature of the problem.

Next, the impact of the non-reciprocity is studied for Massive MIMO transmission specifically. Different precoders and systems loads are considered. The results show that Massive MIMO communication is less sensitive to non-reciprocity of the transceiver circuits. This can be understood from the averaging effects resulting from the large number of antennas. Still, performance degradation is observed and in order to realize high-capacity MaMi, a solution to compensate (calibrate) for the mismatches should be designed and included in the system.

Several calibration solutions have been proposed in the state of the art, whereby the transfer function of the transceiver circuits at the base-station is measured mostly relying on the usage of a reference transceiver, over-the-air signals, or antenna coupling. These approaches are extended to the case with a large number of antennas and the performance is assessed.

Finally, it is concluded that the transceivers non-reciprocity has to be taken care of in order to enable high performance MaMi. Yet, calibration solutions are known and can be applied also for the large number of antennas. It is therefore not expected that this effect would hamper the eventual large-scale deployment of MaMi systems. This is to be further substantiated in real-life validations.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

Massive MIMO relies on spatial multiplexing, which in turn relies on the base station (BS) having relatively accurate channel knowledge, both in the uplink and downlink. Uplink channel knowledge can be acquired in a conventional way with e.g. a known preamble preceding the uplink payload data. When the number of BS antennas is large, downlink channel knowledge is preferably extracted from the uplink channel estimation. TDD operation is then preferred. The key assumption for this method is the so-called channel reciprocity assumption whereby the uplink and downlink channel matrices are supposed to be the transpose of each other. Indeed, the electromagnetic radio propagation can safely be assumed not to exhibit significant non-reciprocal effects in typical environment. However, this reciprocity assumption is not valid for the radio-frequency (RF) transceivers transfer functions that are part of the estimated channel responses: the RF transceivers may exhibit amplitude and phase differences between the up- and downlink. These differences between TX and RX responses cause multi-user interference (MUI) in the downlink. It is therefore crucial to know how much MUI due to non-reciprocity is present in the system and, if this MUI is too high, to minimize the non-reciprocity so as to reduce the level of MUI to acceptable values.

This deliverable is further organized as follows. First, the problem statement and he mathematical models of non-reciprocal transceivers in the channel state acquisition process is explained in Chapter 2. Chapter 3 details the results of the impact study of non-reciprocity on the performance of Massive MIMO systems. In Chapter 4, possible approaches to estimate and compensate the non-reciprocity of the transceivers are introduced and discussed. Finally the overall conclusions are formulated in Chapter 5 and the next steps are mentioned.

Chapter 2 Problem statement and system model

2.1 Introduction

Massive MIMO is an attractive technique to enhance the capacity of future 5G and beyond 5G cellular networks as described in MAMMOET deliverable [1]. Combined with e.g. OFDM, Massive MIMO allows to mitigate the frequency selective channel fading (thanks to OFDM) and increase the spectral efficiency by accommodating many users in the same time-frequency slots (MU-MISO/MU-MIMO). In the downlink, precoding at the base station (BS) side allows to pre-compensate the phase and amplitude of the channel responses in such a manner that all simultaneous users receive their own signal free of multi-user interference (MUI) [2]. It is in fact this interference cancellation property that makes MU-MISO possible in the downlink. Since the users' mobile stations (MS) have only one antenna (or since they cannot estimate the information symbols sent to the other users), they have no means to mitigate the spatial MUI: the MUI must be eliminated by means of the precoding. This requires a relatively accurate channel knowledge at the BS (transmit side in the downlink): any inaccuracy in the TX-CSI will result in MUI that cannot be compensated.

There are basically two means to estimate the downlink channel:

- *Explicit* estimation: the BS transmits known pilots or preambles that each MS receives and processes to estimate its own Mx1 channel. Each MS feeds back the measured channel response (with or without compression) to the BS. The number of BS antennas (M) being very high, explicit estimation is not practical for two reasons:
    - each MS must perform M channel estimations
    - the overhead of feeding back KxM channel responses would severely reduce the spectral efficiency of the system

  Note that another reason for discarding explicit estimation is that the time elapsed between channel estimation and channel use is longer, which results in higher "channel aging" problems in time-varying channel.

- *Implicit* estimation: the BS estimates the channel in the uplink, based on known pilots or preambles sent by the MS's. The BS makes the reciprocity assumption and takes the transpose of the uplink channel matrix. This method has minimal overhead (uplink channel estimation must be done for uplink traffic) and also does not put any signal processing burden on the MS. The channel non-reciprocity, however, must be taken care of.

When the channel is estimated in the uplink, the downlink channel matrix is just the transpose of the uplink matrix, assuming that the channel is reciprocal. However, the "channel" that is measured in the channel estimation phase is actually made up of the propagation channel (the medium between the antennas), the antennas and the transceiver RF, IF and baseband circuits at both sides of the link. The transceiver circuits are usually not reciprocal (the TX and RX frequency responses are different) and this can jeopardise the performance of the MU-MISO system ([3] Section 4.7). In short, the channel information that is fed to the pre-coder is not exactly equal to the downlink channel response and some degradation can be expected.

2.2 System model of the transmission including non-reciprocity

We will consider a frequency flat environment. The reasoning that we will follow is also applicable to frequency selective environments if OFDM is used: in this case, the channel is flat on every sub-carrier and the model can thus be applied per sub-carrier.

2.2.1 UPLINK

In the uplink, K MS's transmit simultaneously to a BS using M antennas. Each MS employs conventional OFDM modulation [2]. The following linear frequency domain model results on each sub-carrier:

$$\boldsymbol{y}^U = \boldsymbol{G}^U \cdot \boldsymbol{d}^U + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{d}^U$ is the column vector of the K frequency domain symbols transmitted by the MS's in the uplink, $\boldsymbol{y}^U$ is the column vector of the M signals received by the BS antenna branches, and $\boldsymbol{G}^U$ is the composite uplink channel.

Including the transmitter and receiver responses, $\boldsymbol{G}^U$ can be expressed as:

$$\boldsymbol{G}^U = \boldsymbol{R}^B \widetilde{\boldsymbol{G}}^U \boldsymbol{T}^M \tag{2}$$

where $\boldsymbol{R}^B$ and $\boldsymbol{T}^M$ are complex diagonal matrices containing respectively the BS receivers and MS transmitters frequency responses. The matrix $\widetilde{\boldsymbol{G}}^U$ contains the propagation channel itself, which is reciprocal. The channel estimated by the BS is $\boldsymbol{G}^U$ as in (2).

2.2.2 DOWNLINK

For the downlink, user separation is achieved by applying a per sub-carrier precoder $\boldsymbol{W}$ that pre-equalises the channel. This pre-filtering is contained in the matrix $\boldsymbol{W}$ of the linear model:

$$\boldsymbol{y}^D = \boldsymbol{G}^D \boldsymbol{W} \boldsymbol{P} \boldsymbol{d}^D + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{d}^D$ is the column vector of the K symbols transmitted by the BS in the downlink, $\boldsymbol{y}^D$ is the column vector of the K signals received by the MS's, $\boldsymbol{P}$ is a power scaling real diagonal matrix (that accounts for power allocation if needed, otherwise $\boldsymbol{P} = \boldsymbol{I_K}$) and $\boldsymbol{G}^D$ is the composite downlink channel. $\boldsymbol{G}^D$ is also affected by the BS transmitters ($\boldsymbol{T}^B$) and MS receivers ($\boldsymbol{R}^M$):

$$\boldsymbol{G}^D = \boldsymbol{R}^M \widetilde{\boldsymbol{G}}^D \boldsymbol{T}^B \tag{4}$$

Here again, $\widetilde{\boldsymbol{G}}^D$ is the propagation part of the downlink channel. That part of the channel is reciprocal and we have

$$\widetilde{\boldsymbol{G}}^D = \widetilde{\boldsymbol{G}}^{U^T} \tag{5}$$

If all transceivers are reciprocal, $\boldsymbol{R}^B = \boldsymbol{T}^B$ and $\boldsymbol{R}^M = \boldsymbol{T}^M$ and it follows that $\boldsymbol{G}^D = \boldsymbol{G}^{U^T}$. However, when non-reciprocity is present, $\boldsymbol{R}^B \neq \boldsymbol{T}^B$ and $\boldsymbol{R}^M \neq \boldsymbol{T}^M$ and it follows that $\boldsymbol{G}^D \neq \boldsymbol{G}^{U^T}$. This is a problem since implicit channel estimation is preferred: the channel is estimated in the uplink ($\boldsymbol{G}^U$) and this estimation is reused in the downlink to compute the downlink

precoder. Because of the non-reciprocity, the pre-coder will be fed with an estimate of the channel $\widehat{\boldsymbol{G}}^D = \boldsymbol{G}^{U^T} = \boldsymbol{T}^{M^T}\widetilde{\boldsymbol{G}}^{U^T}\boldsymbol{R}^{B^T} = \boldsymbol{T}^M\widetilde{\boldsymbol{G}}^D\boldsymbol{R}^B$ whereas it should use $\boldsymbol{G}^D = \boldsymbol{R}^M\widetilde{\boldsymbol{G}}^D\boldsymbol{T}^B$.

The uplink and downlink transmissions are illustrated in Figure 1. In this figure, the BS and MS transmitter and receiver frequency responses are represented by a complex scalar of the form $t_i^{B\ or\ M}$ or $r_i^{B\ or\ U}$. Those scalars are actually the diagonal entries of the diagonal matrices $\boldsymbol{R}^B$, $\boldsymbol{T}^B$, $\boldsymbol{R}^M$ or $\boldsymbol{T}^M$ as follows:

$$\boldsymbol{T}^B = \begin{bmatrix} t_1^B & 0 & \cdots & & 0 \\ 0 & \ddots & \ddots & & \\ \vdots & \ddots & t_m^B & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ 0 & & \cdots & 0 & t_M^B \end{bmatrix}, \boldsymbol{R}^B = \begin{bmatrix} r_1^B & 0 & \cdots & & 0 \\ 0 & \ddots & \ddots & & \\ \vdots & \ddots & r_m^B & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ 0 & & \cdots & 0 & r_M^B \end{bmatrix}$$

$$\boldsymbol{T}^M = \begin{bmatrix} t_1^M & 0 & \cdots & & 0 \\ 0 & \ddots & \ddots & & \\ \vdots & \ddots & t_k^M & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ 0 & & \cdots & 0 & t_K^M \end{bmatrix}, \boldsymbol{R}^U = \begin{bmatrix} r_1^M & 0 & \cdots & & 0 \\ 0 & \ddots & \ddots & & \\ \vdots & \ddots & r_k^M & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ 0 & & \cdots & 0 & r_K^M \end{bmatrix} \quad (6)$$

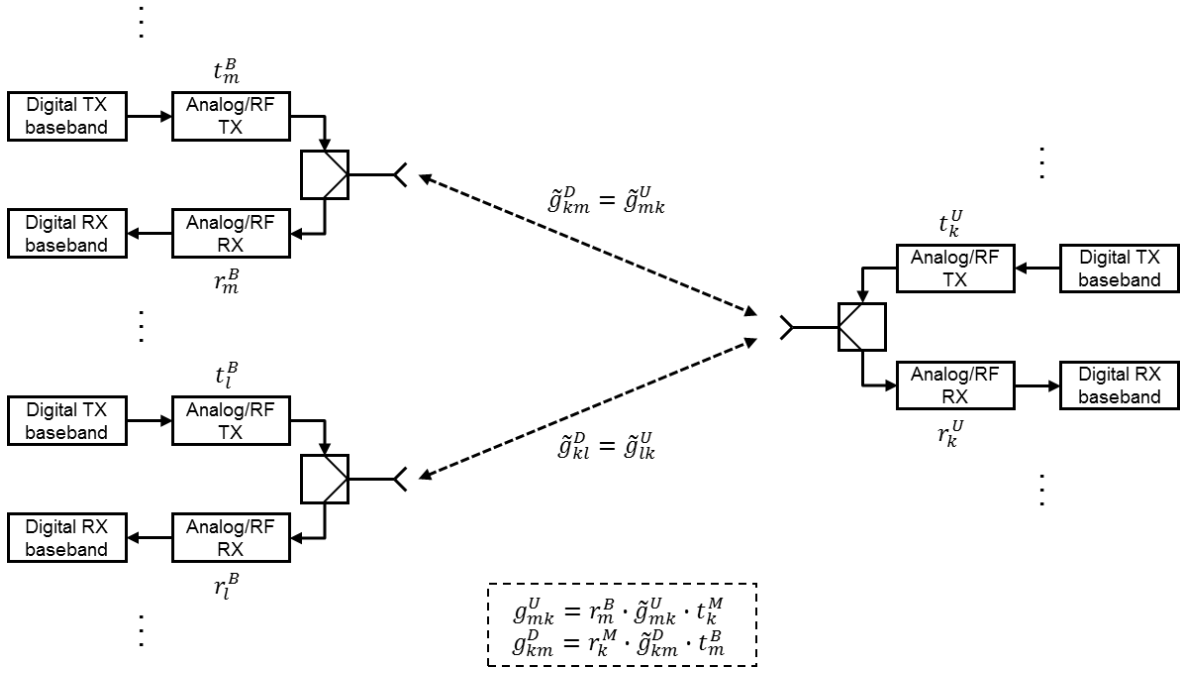

Figure 1 – Uplink and downlink transmission, showing the various frequency responses

2.2.3 Precoders including non-reciprocity

We will develop the three linear precoders (MRT, ZF and MMSE) taking non-reciprocity in the calculation. We will include the power-allocation matrix $\boldsymbol{P}$ for completeness and to show that it does not impact the non-reciprocity. We will also include the transmit power normalization so that the total transmit power is not affected by the pre-coder, regardless of the channel matrix.

2.2.3.1 Normalization of the precoder

The pre-coding matrix must be multiplied by the normalization factor α where α is the inverse of the Frobenius norm of the precoder matrix

$$\alpha = 1/\|\boldsymbol{W}\|_F = 1\Big/\sqrt{\sum_{m=1}^{M}\sum_{k=1}^{K}|w_{km}|^2} \tag{7}$$

This normalization factor must be compensated for at the receive side to yield an estimate of the transmitted symbols:

$$\hat{\boldsymbol{d}}^D = \frac{1}{\alpha}(\alpha \boldsymbol{G}^D \boldsymbol{W}\boldsymbol{P}\boldsymbol{d}^D + \boldsymbol{n}) = \boldsymbol{G}^D \boldsymbol{W}\boldsymbol{P}\boldsymbol{d}^D + \frac{1}{\alpha}\boldsymbol{n} \tag{8}$$

2.2.3.2 MRT pre-coder with non-reciprocity

The MRT pre-coder is given by:

$$\boldsymbol{W}_{MRT} = \left(\boldsymbol{T}^M \tilde{\boldsymbol{G}}^D \boldsymbol{R}^B\right)^H = \boldsymbol{R}^{B^*} \tilde{\boldsymbol{G}}^{D^H} \boldsymbol{T}^{M^*} \tag{9}$$

$$\alpha_{MRT} = \frac{1}{\|\boldsymbol{W}_{MRT}\|_F} \tag{10}$$

The MRT transmission model then reads:

$$\begin{aligned} \hat{\boldsymbol{d}}^D &= \boldsymbol{G}^D \boldsymbol{W}_{MRT} \boldsymbol{P}\boldsymbol{d}^D + \frac{1}{\alpha_{MRT}}\boldsymbol{n} \\ \hat{\boldsymbol{d}}^D &= \boldsymbol{R}^M \tilde{\boldsymbol{G}}^D \boldsymbol{T}^B \cdot \boldsymbol{R}^{B^*} \tilde{\boldsymbol{G}}^{D^H} \boldsymbol{T}^{M^*} \cdot \boldsymbol{P}\boldsymbol{d}^D + \frac{1}{\alpha_{MRT}}\boldsymbol{n} \end{aligned} \tag{11}$$

2.2.3.3 ZF pre-coder with non-reciprocity

The ZF pre-coder is given by:

$$\boldsymbol{W}_{ZF} = \left(\boldsymbol{T}^M \tilde{\boldsymbol{G}}^D \boldsymbol{R}^B\right)^{-1} = \boldsymbol{R}^{B^{-1}} \tilde{\boldsymbol{G}}^{D^{-1}} \boldsymbol{T}^{M^{-1}} \tag{12}$$

$$\alpha_{ZF} = \frac{1}{\|\boldsymbol{W}_{ZF}\|_F} \tag{13}$$

The ZF transmission model then reads:

$$\begin{aligned} \hat{\boldsymbol{d}}^D &= \boldsymbol{G}^D \boldsymbol{W}_{ZF} \boldsymbol{P}\boldsymbol{d}^D + \frac{1}{\alpha_{ZF}}\boldsymbol{n} \\ \hat{\boldsymbol{d}}^D &= \boldsymbol{R}^M \tilde{\boldsymbol{G}}^D \boldsymbol{T}^B \cdot \boldsymbol{R}^{B^{-1}} \tilde{\boldsymbol{G}}^{D^{-1}} \boldsymbol{T}^{M^{-1}} \cdot \boldsymbol{P}\boldsymbol{d}^D + \frac{1}{\alpha_{ZF}}\boldsymbol{n} \end{aligned} \tag{14}$$

2.2.3.4 MMSE pre-coder with non-reciprocity

The MMSE pre-coder is given by:

$$\boldsymbol{W}_{MMSE} = \left(\boldsymbol{T}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{R}^{B}\right)^{H}\left(\boldsymbol{T}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{R}^{B}\cdot\left(\boldsymbol{T}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{R}^{B}\right)^{H}+\sigma^{2}I_{K}\right)^{-1} \tag{15}$$

$$\alpha_{MMSE} = \frac{1}{\|\boldsymbol{W}_{MMSE}\|_{F}} \tag{16}$$

The MMSE transmission model then reads:

$$\begin{aligned}\widehat{\boldsymbol{d}}^{D} &= \boldsymbol{G}^{D}\boldsymbol{W}_{MMSE}\boldsymbol{P}\boldsymbol{d}^{D}+\frac{1}{\alpha_{MMSE}}\boldsymbol{n}\\ \widehat{\boldsymbol{d}}^{D} &= \boldsymbol{R}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B}\cdot\boldsymbol{R}^{B^{*}}\widetilde{\boldsymbol{G}}^{D^{H}}\left(\widetilde{\boldsymbol{G}}^{D}\boldsymbol{R}^{B}\boldsymbol{R}^{B^{*}}\widetilde{\boldsymbol{G}}^{D^{H}}+\sigma^{2}\boldsymbol{T}^{M^{-1}}\boldsymbol{T}^{M^{-H}}\right)^{-1}\boldsymbol{T}^{M^{-1}}\cdot\boldsymbol{P}\boldsymbol{d}^{D}+\frac{1}{\alpha_{MMSE}}\boldsymbol{n}\end{aligned} \tag{17}$$

2.2.4 Multi-user Interference due to non-reciprocity

The linear models (11),(14) and (17) presented in section 2.2.3 lead to several interpretations and highlight the origin of the degradation (MUI and other):

- Degradation:
  - **MRT**: equation (11) shows that, because of the product $\boldsymbol{T}^{B}\cdot\boldsymbol{R}^{B^{*}}$, the matched transmission is not ideal. Since ideal MRT is not interference free, there will be the natural MRT MUI and possibly some additional MUI due to non-reciprocity. There will also be a degradation of the SNR since the exact matched filter is not used.
  - **ZF**: equation (14) shows that the effect of the ZF precoder is altered by the two diagonal matrices ($\boldsymbol{T}^{B}\cdot\boldsymbol{R}^{B^{-1}}$) appearing in (14) between $\widetilde{\boldsymbol{G}}^{D}$ and $\widetilde{\boldsymbol{G}}^{D^{-1}}$. The product $\boldsymbol{T}^{B}\cdot\boldsymbol{R}^{B^{-1}}$ is generally not equal to the identity matrix nor to a scaled identity matrix, although this product is diagonal. Introducing a diagonal matrix in the middle of the product $\widetilde{\boldsymbol{G}}^{D}\widetilde{\boldsymbol{G}}^{D^{-1}}$ will "break" the attempted matrix inversion and cause a non-diagonal result, explaining the appearance of MUI.
  - **MMSE**: equation (17) shows two facts: first, the "matched filter part" of the MMSE precoder is not perfect (same as for the MRT precoder); second, the "whitening filter part" of the MMSE precoder is not perfectly matched to the channel (it only contains $\boldsymbol{R}^{B}$ whereas the effective channel also contains $\boldsymbol{T}^{B}$.
- Interestingly, for all three precoders, the terminal non-reciprocity (captured by $\boldsymbol{R}^{M}$ and $\boldsymbol{T}^{M}$) is not responsible for MUI. It will only scale the symbols individually; this effect can be simply compensated by an appropriate scaling at the terminal side. This will typically be part of the terminal equalization process.
- Similarly, the power allocation ($\boldsymbol{P}$) does not contribute to MUI.
- The propagation matrix $\widetilde{\boldsymbol{G}}^{D}=\widetilde{\boldsymbol{G}}^{U^{T}}$ in this model also includes the parts of the BS or MS's that are common to uplink and downlink, hence reciprocal. This is the case for the antennas (including antenna coupling) and for common components inserted between the antenna and the TX/RX switch.

2.2.5 Information required for non-reciprocity mitigation

We have already shown in the previous section that the terminal non-reciprocity does not introduce MUI and can easily be accounted for in the terminal equalization process.

Hence, only the BS hardware requires attention. By carefully observing the system models (11), (14) and (17), we see that, if the ratios of the TX and RX response is known for all BS antennas, then this ratio can be incorporated into the uplink channel estimation to compensate the non-reciprocity. This ratio can be expressed in matrix form as $\boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B}$. If we post-multiply the estimated downlink channel matrix ($\widehat{\boldsymbol{G}}^{D} = \widehat{\boldsymbol{G}}^{U^{T}}$) with this compensation, we obtain:

$$\hat{G}^{D}_{comp} = \widehat{\boldsymbol{G}}^{D}\boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B} = \boldsymbol{T}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{R}^{B} \cdot \boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B} = \boldsymbol{T}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B} \tag{18}$$

Basically, this boils down to cancelling the effect of the BS receivers and introducing the effect of the BS transmitter. The effect of this compensation on the three precoders is as follows (note that the normalization factors (α) are slightly different from those in Section 2.2.3 since the precoders are based on the compensated channel matrix):

$$\widehat{\boldsymbol{d}}^{D} = \boldsymbol{R}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B} \cdot \boldsymbol{T}^{B^{*}}\widetilde{\boldsymbol{G}}^{D^{H}}\boldsymbol{T}^{M^{*}} \cdot \boldsymbol{P}\boldsymbol{d}^{D} + \frac{1}{\alpha_{MRT}}\boldsymbol{n} \tag{19}$$

$$\widehat{\boldsymbol{d}}^{D} = \boldsymbol{R}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B} \cdot \boldsymbol{T}^{B^{-1}}\widetilde{\boldsymbol{G}}^{D^{-1}}\boldsymbol{T}^{M^{-1}} \cdot \boldsymbol{P}\boldsymbol{d}^{D} + \alpha_{ZF}\boldsymbol{n} = \boldsymbol{R}^{M}\boldsymbol{T}^{M^{-1}} \cdot \boldsymbol{P}\boldsymbol{d}^{D} + \frac{1}{\alpha_{ZF}}\boldsymbol{n} \tag{20}$$

$$\widehat{\boldsymbol{d}}^{D} = \boldsymbol{R}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B} \cdot \boldsymbol{T}^{B^{*}}\widetilde{\boldsymbol{G}}^{D^{H}}\left(\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B}\boldsymbol{T}^{B^{*}}\widetilde{\boldsymbol{G}}^{D^{H}} + \sigma^{2}\boldsymbol{T}^{M^{-1}}\boldsymbol{T}^{M^{-H}}\right)^{-1}\boldsymbol{T}^{M^{-1}} \cdot \boldsymbol{P}\boldsymbol{d}^{D} + \frac{1}{\alpha_{MMSE}}\boldsymbol{n} \tag{21}$$

The three precoders are now mathematically perfect in the sense that the channel used for the precoder is identical to the downlink channel (both of them equal to $\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B}$) except for the terminal transceiver responses ($\boldsymbol{R}^{M}$ and $\boldsymbol{T}^{M}$).

A final remark is that the ratios of the TX and RX responses need to be known up to a common complex scaling factor only since a common complex scaling factor will not introduce MUI and can be equalized easily at the terminal side. This can be useful is some calibration procedures where a common but unknown factor (e.g. a signal, an attenuation, ...) is used in the calibration.

2.2.6 Post equalization SINR

The post equalization signal to interference and noise ratio (SINR) can be computed as follows. We will assume uniform power allocation ($\boldsymbol{P} = \boldsymbol{I}_{\boldsymbol{K}}$). First, we collapse the product of the channel and precoder matrices into a single matrix $\boldsymbol{A}$ for compactness, yielding the following general form for the three precoders (note that $\boldsymbol{A}$ is square):

$$\widehat{\boldsymbol{d}}^{D} = \boldsymbol{A}\boldsymbol{d}^{D} + \frac{1}{\alpha}\boldsymbol{n} \tag{22}$$

Then, the post-equalization SINR of the kth terminal is easily shown to be:

$$\mathrm{SNIR}_{k} = \frac{\sigma_{d}^{2}\left[\mathrm{diag}(\boldsymbol{A}) \cdot \mathrm{diag}\left(\boldsymbol{A}^{\boldsymbol{H}}\right)\right]_{kk}}{\sigma_{d}^{2}\left[\overline{\mathrm{diag}}(\boldsymbol{A}) \cdot \overline{\mathrm{diag}}(\boldsymbol{A}^{\boldsymbol{H}})\right]_{kk} + \sigma_{n}^{2}/\alpha^{2}} \tag{23}$$

where σ_d^2 is the variance of the transmitted symbols (usually equal to 1) and σ_n^2 is the variance of the receiver AWGN component; $\mathrm{diag}(\boldsymbol{A})$ is a diagonal matrix containing only the principal diagonal of $\boldsymbol{A}$; $\overline{\mathrm{diag}}(\boldsymbol{A})$ is equal to $\boldsymbol{A} - \mathrm{diag}(\boldsymbol{A})$ i.e. it is equal to $\boldsymbol{A}$ except for the principal diagonal of $\boldsymbol{A}$ that is set to 0; $[.]_{kk}$ represents the kth element of the principal diagonal of a matrix.

Equation (23) is very useful since it allows to compute easily the post-equalization SINR with or without reciprocity.

Chapter 3 Impact analysis

3.1 Introduction

The non-reciprocity induced by the analog/RF transceivers is known to severely degrade conventional MU-MIMO systems ([3] Section 4.7). These systems operate usually at or near full system load and/or exploit high order constellations, which explains their high sensitivity to non-reciprocity.

Massive MIMO systems usually do not operate at high system load and are often used with moderate order constellations (e.g. QPSK or 16QAM). Therefore, it is necessary to estimate the impact of non-reciprocity on massive MIMO systems, given their typical operating conditions.
The SINR equation (23) derived in the previous section will be used for this purpose. We will use it to determine the sensitivity of the three linear pre-coders to different levels of non-reciprocity.

3.2 Assumptions for the impact assessment of non-reciprocity

We will consider a system with the parameters given in Table 1.

Table 1 – Parameters for assessment of the impact of non-reciprocity

Parameter	**Symbol**	**Value**
# BS antennas	M	100
# users	K	5, 10, 20, 50
SNR	-	-20 : 0.5 : 20
Non-reciprocity	-	-5, -10, -15, -25, -30
Channel type	-	Rayleigh, flat fading (SINR calculation) Rayleigh, multipath (SNR degradation)

The number of BS antennas (100) is typical of a massive MIMO system. The number of users (5, 10, 20 and 50) covers the scenarios of low system load (5% and10%), medium system load (20%) and high system load (50%).

The amount of reciprocity is defined as follows: for a given uplink channel coefficient g_{km}^{D} between the mth antenna of the BS and the kth user, non-reciprocity is modelled as a multiplicative error affecting the uplink channel coefficient g_{mk}^{U} used by the precoder:

$$g_{mk}^{U} = \left(1 + \epsilon_{r,m}\right) g_{km}^{D} \tag{24}$$

In matrix form, this can be expressed as follows:

$$\boldsymbol{G}^U = (\boldsymbol{I}_M + \boldsymbol{E}_r)\boldsymbol{G}^D \quad (25)$$

where $\boldsymbol{E}_r$ is a diagonal matrix with i.i.d. Gaussian entries $\epsilon_{r,m} \sim \mathcal{CN}(0, \sigma_r^2)$. The amount of non-reciprocity is conveniently expressed in decibels as $20 \cdot \log_{10}(\sigma_r)$.

3.3 SINR in flat fading channels

In this section, we will compute the theoretical SINR in flat fading channels based on formula (23).

3.3.1 SINR with reciprocal transceivers

With reciprocal transceivers, the SINR is affected by the multi-user interference and the noise variance. The SINR results are plotted in Figure 2 for the four different system loads. The horizontal axis shows the SNR that represents the ratio of the received energy over the receiver noise without interference from other users. The vertical axis represents the SINR that includes the interference from other users. We can draw some generic conclusions from those curves. We will assume that an SINR of 10 to 20dB is targeted, which is adequate to support modulation and coding schemes from QPSK rate 1/2 to 16QAM rate 3/4 approximately (see for example [4] and [5]).

- The MRT scheme should not be used beyond a load of 10%. At 10% load, the SINR hardly achieves 10dB, even when the receiver noise power tends to 0. This tends to show that only QPSK with good channel coding can be exploited at 10% load.
- The MMSE scheme as such (with the goal of trading of MUI vs receiver noise) is not beneficial (it has almost no effect on system load of 5, 10 and 20%) or not exploitable (it has a more visible effect at 50% system load but in a range of SNR/SINR that is too low (below 10dB). The MMSE could still be desired for regularization purposes but it will have a negligible impact on the SINR at the considered SINR ranges.
- From the curves at 50% load, we observe the “canonical” behaviour, namely that the MMSE converges to the MRT at very low SNR and to the ZF at high SNR, which gives confidence in the SNR and SINR assessment.
- The ZF scheme is the precoder of choice for a wide range of system loads, from 10+ to 50%. The SINR is good enough to support 16QAM with coding rate 3/4 (3bits/s/Hz).

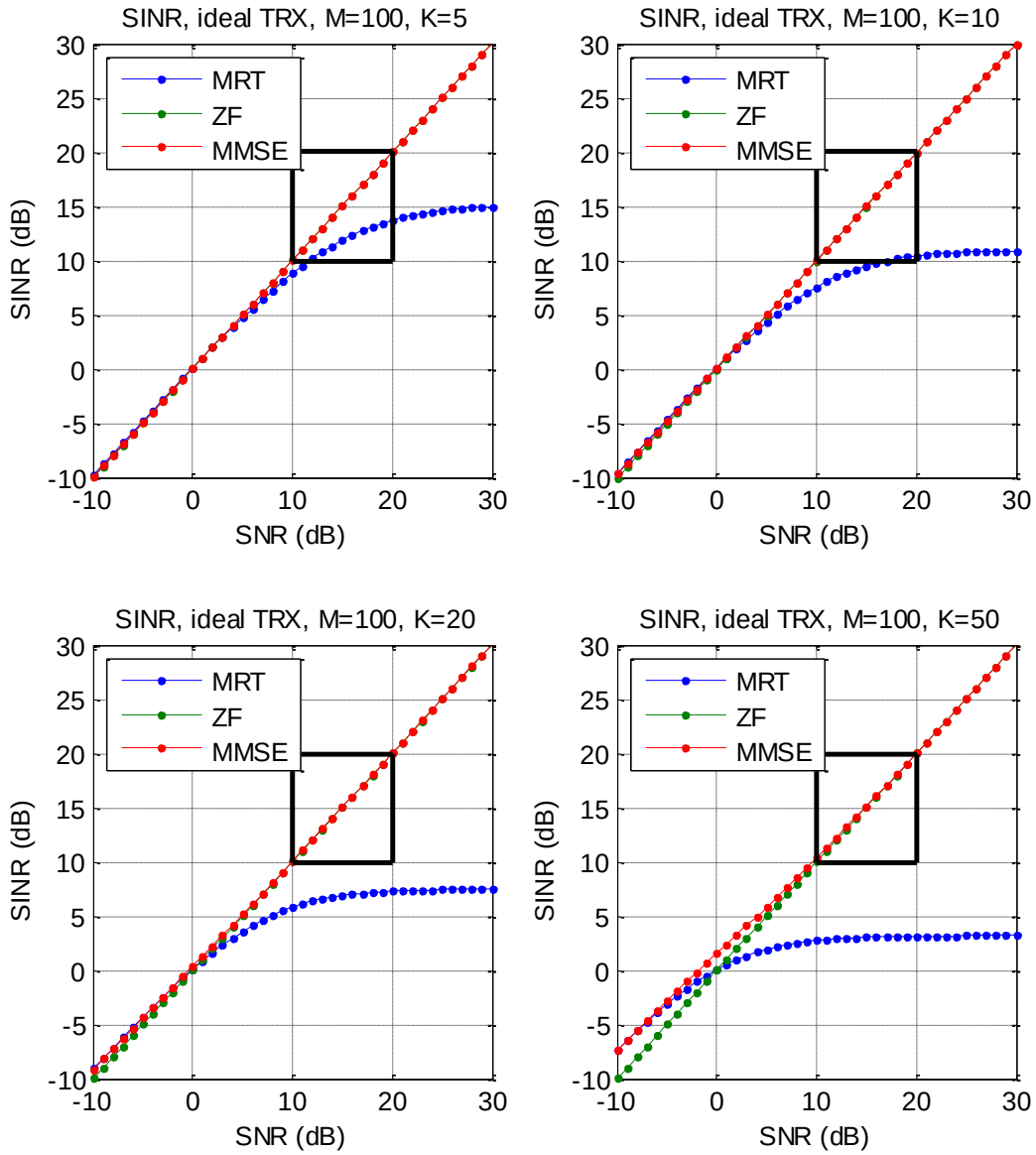

Figure 2 – SINR with reciprocal transceivers. A useful range of SNR and SINR is highlighted.

3.3.2 SINR with non-reciprocal transceivers

MRT

Non-reciprocity has very little impact on MRT precoding, for all system loads. A degradation of only 1dB is observed for a non-reciprocity of -5dB, which is a very high non-reciprocity. Again, as mentioned in Section 3.3.1, the MRT scheme is only suitable for low system loads (10% at most).

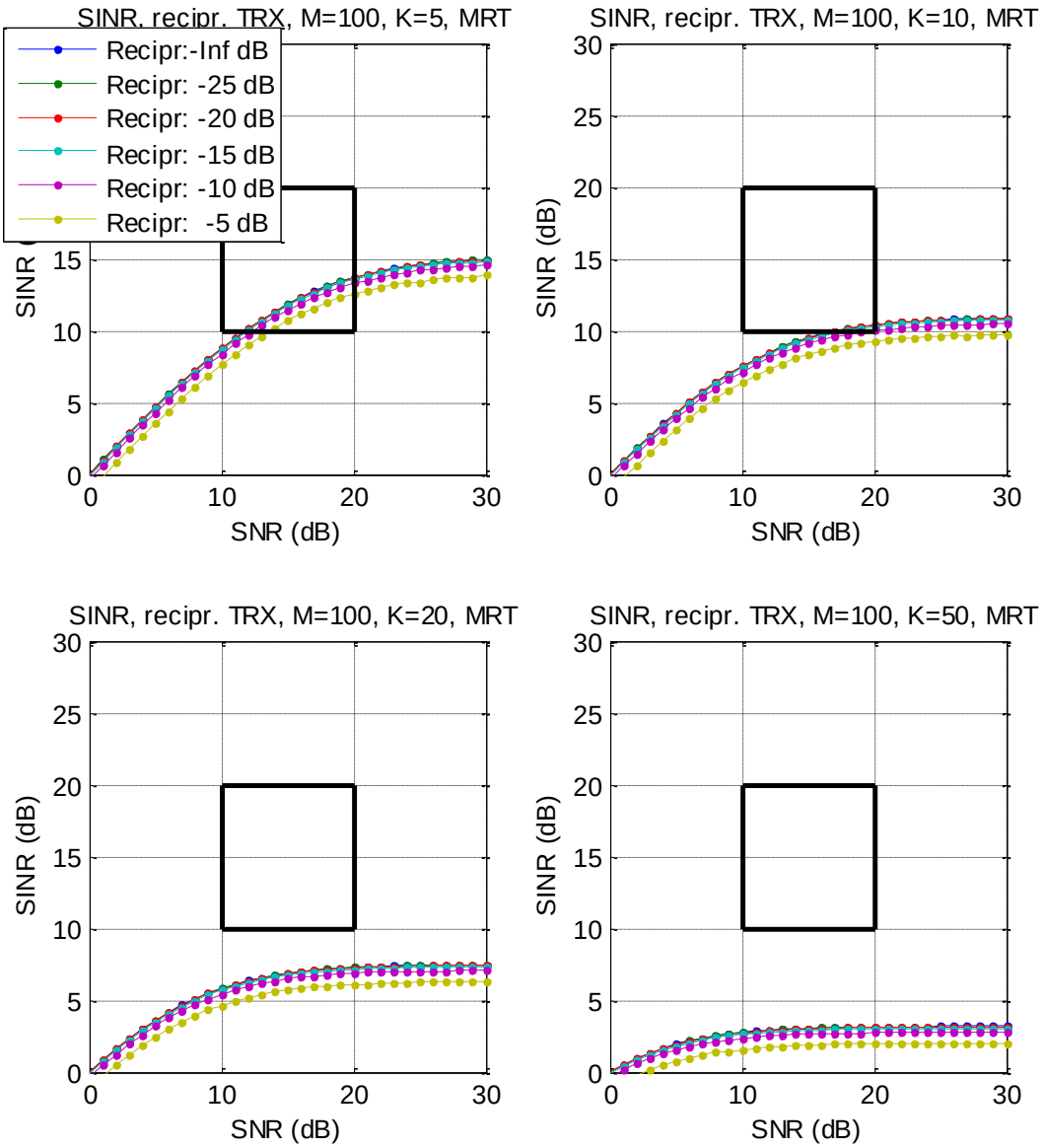

Figure 3 – SINR for non-reciprocal transceivers, MRT pre-coding

ZF and MMSE

The impact of non-reciprocity is shown for MMSE in Figure 4. The impact on ZF is very similar and is not shown. For the targeted SNR and SINR between 10 and 20dB, ZF and MMSE can be used. Here, the sensitivity to non-reciprocity is load dependent. To avoid a degradation higher than 1dB, the following levels of non-reciprocity should be:

- at 5 or 10% system load, the non-reciprocity should not exceed -15dB
- at 20% system load, the non-reciprocity should not exceed -20dB
- at 50% system load, the non-reciprocity should not exceed -25dB

These trends are not surprising. Whenever non-idealities are introduced in a massive MIMO system, the acceptable amount of non-ideality must be specified for a given system load to be meaningful.

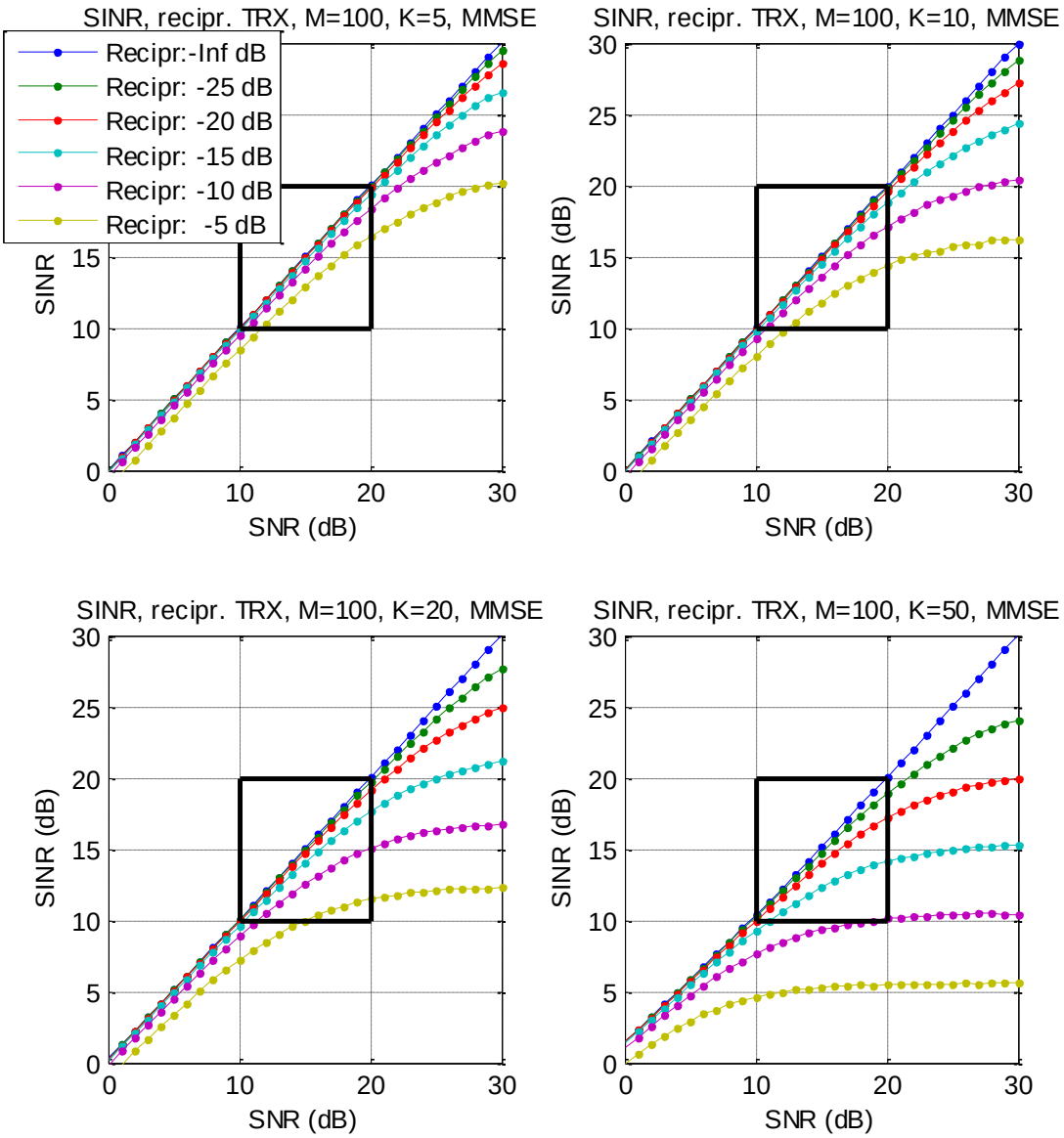

Figure 4 - SINR for non-reciprocal transceivers, MMSE pre-coding

3.3.3 SNR degradation in frequency selective channels

The results from the previous section are valid for frequency flat channels. We can expect similar results in frequency selective channels if the same channel coefficient distribution is valid per-subcarrier. This applies to e.g. multipath with Rayleigh fading per tap were OFDM results in a flat Rayleigh fading per sub-carrier. It has to be noted that the channel measurement campaigns in the frame of WP1 of MAMMOET have indicated that these theoretical models are not fully representative for reality. An update of the results will consider the new massive MIMO channel models (in WP4).

In addition, OFDM uses channel coding to mitigate the effect of deep fades in the frequency response that would otherwise limit severely the BER/PER performance. Assessing

analytically the SNR degradation or the equivalent SINR is hard for coded OFDM. We therefore use coded BER simulations to assess the SNR degradation due to non-reciprocity.

Figure 5 shows the BER for OFDM MRT in frequency selective channels with 100 antennas at the BS and 10 users in the ideal case (reciprocal channels) and with non-reciprocity values of -5, 10, -15 and -20 dB. Note that the horizontal axis is not calibrated in the same way as in Section 3.3.2 but we are here only concerned about relative values (degradation with respect to the ideal case). We observe in Figure 5 that coded OFDM with MRT is not very sensitive to non-reciprocity and that, in this scenario, a non-reciprocity of -10dB would only degrade the performance by less than 1 dB, which is in line with our conclusions from Section 3.3.2.

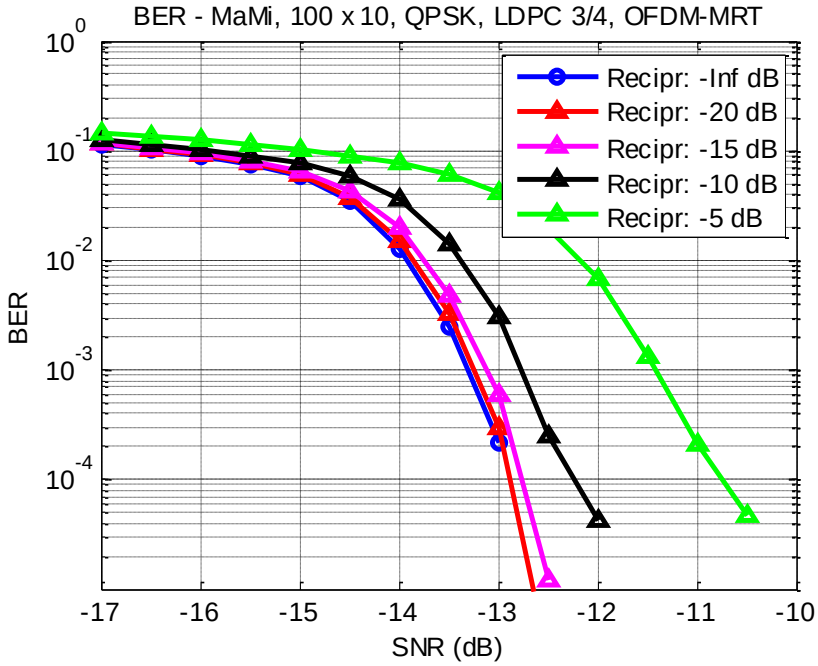

Figure 5 – BER for OFDM-MRT with non-reciprocity

Figure 6 shows the BER for coded OFDM ZF in frequency selective channels with 100 antennas at the BS and 25 users in the ideal case (reciprocal channels) and with non-reciprocity values of -5, 10, -15 and -20 dB. From this figure, a non-reciprocity of -15dB is acceptable since it causes a degradation smaller than 1dB. This could be compared to the results in Figure 4: for 16QAM with code rate 1/2, an SNR of 15dB is normally required. We can check in Figure 4 (20 users) that, at an SNR of 15dB, a non-reciprocity of -15dB causes an SINR degradation smaller than 1dB.

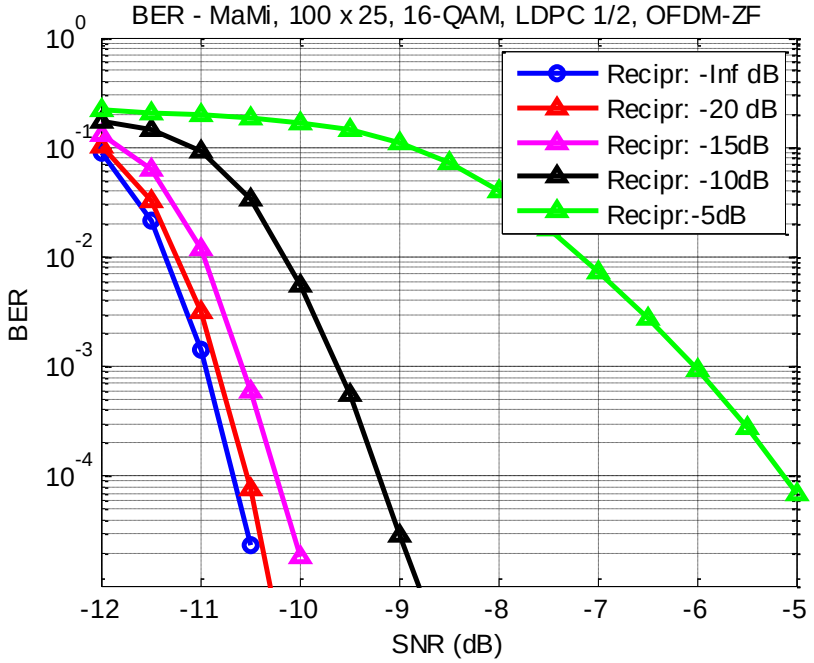

Figure 6 – BER for OFDM-ZF MRT with non-reciprocity

Chapter 4 Solution approaches

In this chapter, we review state-of-the-art methods for reciprocity calibration. Reciprocity calibration is covered to some degree in the literature. The three most prevalent approaches for reciprocity calibration in pre-coded MIMO systems are:

- Over-the-air (OTA) calibration, which requires cooperation from the MS but also enables to calibrate the terminal non-reciprocity;
- Calibration of the BS *with* extra hardware, where typically a reference transceiver is used to calibrate all other BS transceivers;
- Calibration of the BS *without* extra hardware, where typically coupling between the antenna elements is exploited for the BS reciprocity calibration.

As detailed in Section 2.2.5, the reciprocity calibration consist in estimating the diagonal matrix $\boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B}$ for the BS (and, if needed, the diagonal matrix $\boldsymbol{R}^{M^{-1}}\boldsymbol{T}^{M}$ for the MS) and to left multiply the channel matrix estimated in the uplink ($\widehat{\boldsymbol{G}}^{D} = \widehat{\boldsymbol{G}}^{U^{T}}$):

$$\hat{G}^{D}_{comp} = \widehat{\boldsymbol{G}}^{D}\boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B} = \boldsymbol{T}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{R}^{B} \cdot \boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B} = \boldsymbol{T}^{M}\widetilde{\boldsymbol{G}}^{D}\boldsymbol{T}^{B} \tag{26}$$

The whole purpose of the reciprocity calibration is thus to estimate the diagonal matrix $\boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B}$. We will refer to this matrix as $\boldsymbol{C}^{B} = \boldsymbol{R}^{B^{-1}}\boldsymbol{T}^{B}$ for the BS and $\boldsymbol{C}^{M} = \boldsymbol{R}^{M^{-1}}\boldsymbol{T}^{M}$ for the MS.

In the coming sections, we first describe the three calibration methods mentioned above and also the necessary algorithms to extract the desired calibration coefficients from the measurements. A section on the stability of the calibration over time follows. The chapter ends with a discussion on the feasibility of the different methods and on some additional technical considerations that are needed when selecting a reciprocity calibration method.

4.1 Over-the-air reciprocity calibration

OTA calibration involves channel estimation in both the downlink and uplink and collecting these two-way channel estimations at the pre-coder side, which is the BS in massive MIMO. The steps are typically as follows:

1. The BS sends a sounding signal (preamble or pilots) to the MS.
2. The MS estimates the downlink channel $\widehat{\boldsymbol{G}}^{D}$ based on the received sounding signal.
3. The MS sends an uplink packet, with a sounding signal and with the estimated downlink channel $\widehat{\boldsymbol{G}}^{D}$ encoded in the data portion of the packet.
4. The BS estimates the uplink channel $\widehat{\boldsymbol{G}}^{U}$ based on the received sounding signal.
5. The BS, with the knowledge of $\widehat{\boldsymbol{G}}^{D}$ and $\widehat{\boldsymbol{G}}^{U}$, can estimate the non-reciprocity calibration matrices $\boldsymbol{C}^{B}$ and $\boldsymbol{C}^{M}$ for both the BS and the MS.
6. The BS sends a downlink packet to the MS with the calibration matrix $\boldsymbol{C}^{M}$ in the data portion of the packet.

At the end of this procedure, both entities (BS and MS) know their own calibration coefficients and are thus able to apply precoding algorithms based on channel estimated during the most recent packet reception.

OTA calibration using such a procedure has been standardized in the IEEE 802.11n standard ([6] Section 9.29 and 20.3.12-13, [7]). This standard describes the sounding signals that must be transmitted but it does not, however, provide details on how to compute the calibration matrices $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$ from the knowledge of $\widehat{\boldsymbol{G}}^D$ and $\widehat{\boldsymbol{G}}^U$, since this is implementation dependent.

It is possible to recover $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$ as follows. Because of the properties of the (unknown) calibration matrices, we have the following equality (note that $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$ are diagonal matrices so they are equal to their own transpose):

$$\widehat{\boldsymbol{G}}^{U^T}\boldsymbol{C}^B = \boldsymbol{C}^M\widehat{\boldsymbol{G}}^D \tag{27}$$

This is an implicit equation with two unknown matrices so it cannot be solved in closed-form. The algorithm shown in Figure 7 can be used to iteratively estimate $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$. The algorithm starts with identity matrices as initial values for $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$. Then it alternates between solving for $\boldsymbol{C}^B$, assuming $\boldsymbol{C}^M$ is correct, and solving for $\boldsymbol{C}^M$, assuming $\boldsymbol{C}^B$ is correct. When equation (27) is verified up to a small error ϵ the algorithm stops. Given the structure of the problem and the fact that the calibration values never tend to extreme values (zero or infinity), the algorithm always converges in a few iterations, typically less than 10 iterations. Note that this is, at each iteration a least square solution to find the diagonal elements of $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$.

Initialize

$\boldsymbol{C}^B = \boldsymbol{I}_M$

$\boldsymbol{C}^M = \boldsymbol{I}_K$

Repeat until $\left|\widehat{\boldsymbol{G}}^{U^T}\boldsymbol{C}^B - \boldsymbol{C}^M\widehat{\boldsymbol{G}}^D\right| < \epsilon$

$\boldsymbol{A} = \widehat{\boldsymbol{G}}^{U^T}$

$\boldsymbol{B} = \boldsymbol{C}^M\widehat{\boldsymbol{G}}^D$

$[\boldsymbol{C}^B]_{mm} = \boldsymbol{A}_{(:,m)}{}^H\boldsymbol{B}_{(:,m)}/\left(\boldsymbol{A}_{(:,m)}{}^H\boldsymbol{A}_{(:,m)}\right)$, for $m = 1\cdots M$

$\boldsymbol{A} = \widehat{\boldsymbol{G}}^D$

$\boldsymbol{B} = \widehat{\boldsymbol{G}}^{U^T}\boldsymbol{C}^B$

$[\boldsymbol{C}^M]_{kk} = \boldsymbol{B}_{(k,:)}\boldsymbol{A}_{(k,:)}{}^H/\left(\boldsymbol{A}_{(k,:)}\boldsymbol{A}_{(k,:)}{}^H\right)$, for $k = 1\cdots K$

End

Figure 7 – Algorithm for OTA. Estimation of $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$.[1]

OTA reciprocity calibration has also been extensively studied in [8] in which a similar problem is solved. The authors also use an alternating iterative solution but exploit a total least square minimization, which improves slightly the solution when the known values (in this case $\widehat{\boldsymbol{G}}^D$ and $\widehat{\boldsymbol{G}}^U$) are noisy because they are estimated values.

[1] We use here the Matlab notation to designate parts of matrices: $\boldsymbol{A}_{(:,m)}$ means the m[th] column of $\boldsymbol{A}$ and $\boldsymbol{A}_{(k,:)}$ means the k[th] row of $\boldsymbol{A}$.

It should be noted that this OTA calibration does not need to be performed as often as the channel estimation itself: the calibration values are known to be varying relatively slowly and certainly much slower than the channel coefficients.

4.2 Reciprocity calibration of the BS with a reference front-end

Several authors ([9], [10], [11]) have proposed to use additional hardware to calibrate the ratio of the frequency responses ($\boldsymbol{C}^B$ and $\boldsymbol{C}^M$) of the BS transceivers. The underlying idea is to use an additional source/sink that does not need to be known accurately since it will appear as a scaling in the estimation of $\boldsymbol{C}^B$ and $\boldsymbol{C}^M$.

The extra hardware usually consists of a complete transceiver that is used for calibration purposes only. This additional transceiver can be coupled to all the BS transceivers via directional couplers [12] as in Figure 8 or via switches [9] as in Figure 9.

The calibration is done in two steps:

- Step 1: a known sounding signal is generated sequentially by all M transmitters and received by the reference receiver. This yields, after removal of the sounding signal, the measurement of the following transfer function in antenna m:

$$t_{1,m} = r_{ref}\beta_m t_m^B \tag{28}$$

  where β_m captures the (unknown) transfer function in the calibration path.

- Step 2: a known sounding signal is generated in the reference transmitter and received by all M receiver antenna branches simultaneously or sequentially. This yields, after removal of the sounding signal, the measurement of the following transfer function in antenna m:

$$t_{2,m} = r_m^B\beta_m t_{ref} \tag{29}$$

Both steps leading to $t_{1,m}$ and $t_{2,m}$ can of course be repeated and averaged to reduce the variance of the error in the measurement process. Finally, the first measurement is divided by the second measurement, yielding the desired calibration factor ($c_m^B = t_m^B/r_m^B$):

$$\frac{t_{1,m}}{t_{2,m}} = \frac{r_{ref}\beta_m t_m^B}{r_m^B\beta_m t_{ref}} = \frac{r_{ref}}{t_{ref}} \cdot \frac{t_m^B}{r_m^B} = \frac{r_{ref}}{t_{ref}} \cdot c_m^B \tag{30}$$

In this expression, the term t_m^B/r_m^B is the desired calibration factor (ratio of transmitter frequency response over receiver frequency response) while the term r_{ref}/t_{ref} comes from the unknown frequency response of the reference transceiver. However, since all calibration factors are multiplied by the same unknown term, this will not introduce MUI.

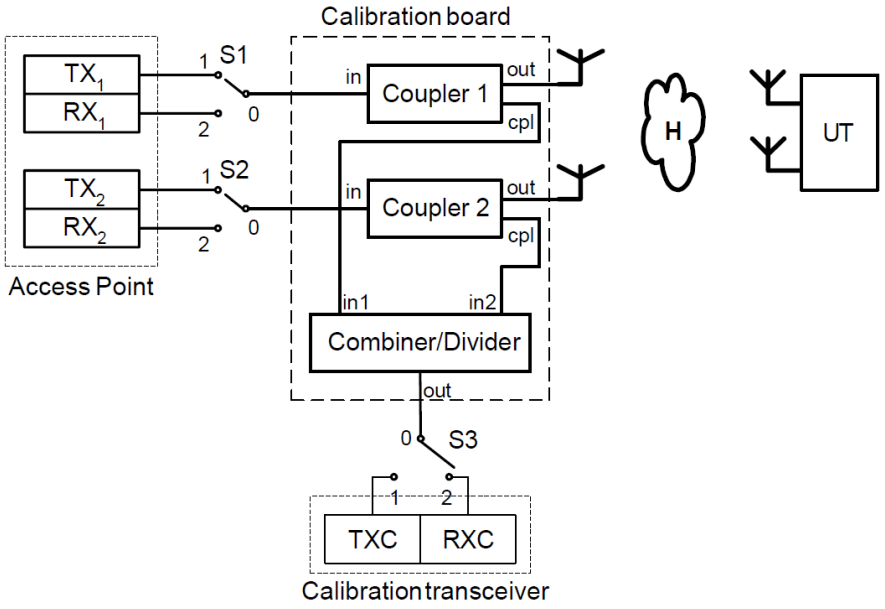

Figure 8 – Additional transceiver connected with directional couplers (from [9])

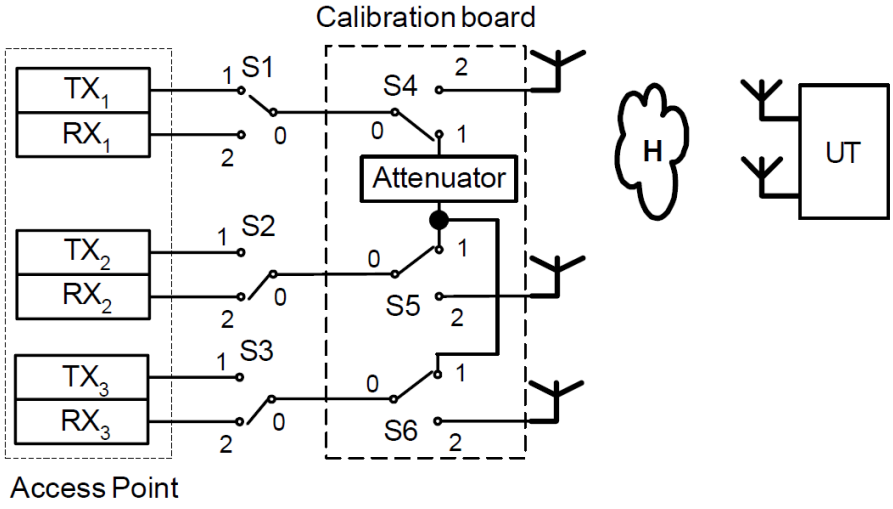

Figure 9 - Additional transceiver connected with switches and attenuators (from [9])

Both methods (directional couplers or switches) have their own merits and disadvantages:

- Connection with directional couplers
  - Pros:
    - The connection with directional couplers is less lossy. Indeed, a directional coupler has a fairly low insertion loss (typically a few tenths of a dB) in the signal path to the antenna.
    - A single attenuator (not shown in Figure 8) can be added in the common path to adjust the signal level during the calibration.
    - The coupling coefficient and the combiner/splitter losses need not be matched with high accuracy since they will cancel in the calibration.
  - Cons:

        - The splitter/combiner needs M ports which is quite large for massive MIMO and can result in high insertion loss (theoretical loss: $10 \cdot \log_{10}(M)$ (dB))
        - The solution does not scale easily for arbitrary scaling factor (adding e.g. one BS antenna means a totally different coupler/combiner must be used).
- Connection with switches
    - Pros:
        - This solution scales easily (adding one BS antenna is easy) although the large amount of switches and connections can be a hurdle.
        - A single attenuator (as shown in Figure 9) can be added in the common path to adjust the signal level during the calibration.
    - Cons:
        - The connection with switches is more lossy. According to the carrier frequency, a loss of 1 to 2dB in an RF switch is possible.
        - Impedance matching of the whole switching network may be a significant challenge due to the large number of open connections.

For both methods, one transceiver is devoted only to calibration and not used for real transmission. Given the large number of transceivers in massive MIMO, this is not a serious problem. Since the transmitters transmit sequentially to the reference receivers, it is important that all transceivers, including the reference transceivers are locked to the same crystal. Otherwise, random phase errors would appear in the measurements of the BS transmitters (step 1 above).

This method has been successfully implemented in a conventional MIMO system in the downlink of an OFDM-SDMA system in a WLAN context [12].

4.3 Reciprocity calibration exploiting antenna coupling

4.3.1 Description of the method

This method is similar to the method detailed in Section 4.2 except that it does not use a "wired" link between the reference transceiver and the transceivers to be measured. Rather, it uses the "natural" antenna coupling as a link [13], [14] [15]. This has two advantages: first, no extra hardware (one reference transceiver, one programmable attenuator and the couplers/splitter-combiner or switches) is needed; second, all transceiver can sequentially be used as reference transceiver for the other $M-1$ transceivers, which can provide an enhanced accuracy in the calibration process. Indeed, this provides measurements whereas the method in Section 4.2 provides M measurements.

To develop this method we need to introduce the "channels" h_{ml} and h_{lm} between antennas m and l (see illustration in Figure 10):

$$\begin{aligned} h_{lm} &= r_l^B \tilde{h}_{lm} t_m^B \\ h_{ml} &= r_m^B \tilde{h}_{ml} t_l^B \end{aligned} \tag{31}$$

h_{lm} refers to the channel from transmitter m to receiver l and h_{ml} refers to the channel from transmitter l to receiver m. The pure propagation part is reciprocal: $\tilde{h}_{ml} = \tilde{h}_{lm}$ in (31). This can also be put in a compact matrix form:

$$\boldsymbol{H} = \boldsymbol{R}^B \tilde{\boldsymbol{H}} \boldsymbol{T}^B \tag{32}$$

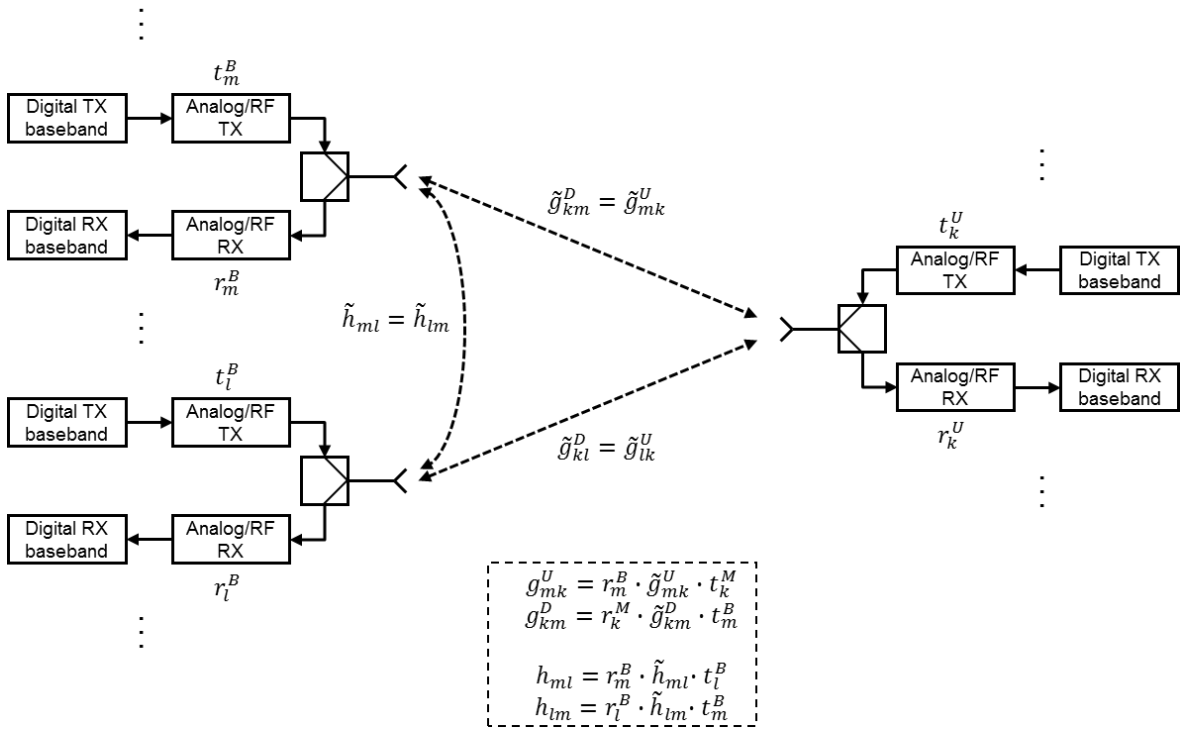

Figure 10 – Base station with inter-antenna coupling

When transmitting from antenna m to l (respectively l to m) the received signal is:

$$\begin{aligned} y_{lm} &= r_l^B \tilde{h}_{lm} t_m^B d + n_{lm} \\ y_{ml} &= r_m^B \tilde{h}_{ml} t_l^B d + n_{ml} \end{aligned} \tag{33}$$

This can be rewritten as:

$$\begin{aligned} y_{lm} &= r_l^B r_m^B \tilde{h}_{lm} \frac{t_m^B}{r_m^B} d + n_{lm} = \beta_{lm} c_m^B d + n_{lm} \\ y_{ml} &= r_l^B r_m^B \tilde{h}_{ml} \frac{t_l^B}{r_l^B} d + n_{ml} = \beta_{lm} c_l^B d + n_{ml} \end{aligned} \tag{34}$$

where $\beta_{lm} = \beta_{ml} = r_l^B r_m^B \tilde{h}_{lm}$. In this pair of equations, d is known, y_{lm} and y_{ml} are measured (hence known), β_{lm} is unknown and c_m^B and c_l^B are the two parameters that we wish to estimate. As such, since there are two equations and three unknowns, this system of two equations cannot be solved independently.

4.3.2 Least-squares solution (direct path)

If we take one antenna as a reference (without loss of generality antenna 1), we end up with the set of $2(M-1)$ equations and measurements:

$$y_{1m} = \beta_{1m} c_m^B d + n_{1m}$$
$$y_{m1} = \beta_{1m} c_1^B d + n_{m1} \tag{35}$$

Forming the following ratios and assuming that the measurement noise (n_{1m} and n_{m1}) is negligible or can be averaged out

$$\frac{y_{1m}}{y_{m1}} = \frac{\beta_{1m} c_m^B d + n_{1m}}{\beta_{1m} c_1^B d + n_{m1}} \cong \frac{c_m^B}{c_1^B} \tag{36}$$

we can directly find the desired values

$$c_m^B = c_1^B \frac{y_{1m}}{y_{m1}} \tag{37}$$

The values of c_m^B are all scaled by the unknown value c_1^B that does not affect the reciprocity as long as this value is bounded.

This solution can be shown to be an LS solution of the set of $2(M-1)$ equations (35). It is also very similar to the calibration method using a reference front-end (Section 4.2), the only difference being here that the reference front-end (here transceiver 1) is coupled to the other transceivers through the coupling between antennas.

4.3.3 Least-squares solution (generalized)

The direct path approach (Section 4.3.2) suffers from unbounded second moment [14]. A better approach to solve for all c_m^B is to exploit all possible $M(M-1)/2$ pairs of measurements as follows. If the observations in (34) were noiseless, we would have the property

$$c_l^B y_{lm} - c_m^B y_{ml} = 0 \tag{38}$$

Hence, a sensible approach is to define the following least-squares (LS) objective function [16]

$$J_{cal}(c_1^B, c_2^B, \ldots, c_M^B) = J_{cal}(\boldsymbol{c}^B) = \sum_{\substack{m,l \\ m \neq l}} \left| c_l^B y_{lm} - c_m^B y_{ml} \right|^2 \tag{39}$$

This cost function can be minimized in either of two ways.

- Solution 1: since (38) is equal to 0 for the trivial solution $\boldsymbol{c}^B = \mathbf{0}$, which is not a desirable solution, we also impose the constraint $|\boldsymbol{c}^B|^2 = 1$. To solve this constrained LS minimization, we define the Lagrangian

  $$F_{cal}(\boldsymbol{c}^B, \lambda) = J_{cal}(\boldsymbol{c}^B) - \lambda(|\boldsymbol{c}^B|^2 - 1) \tag{40}$$

  Differentiating $F_{cal}(\boldsymbol{c}^B, \lambda)$ with respect to ${c_m^B}^*$ and setting the partial derivatives to 0, we get after some elementary manipulations:

$$\frac{\partial}{\partial {c_m^B}^*} F_{cal}(\boldsymbol{c}^B, \lambda) = \sum_{\substack{l \\ l \neq m}} c_m^B |y_{ml}|^2 - c_l^B y_{lm} {y_{ml}}^* - \lambda\, c_m^B \tag{41}$$

This can be put in matrix form:

$$\boldsymbol{A}\boldsymbol{c}^B = \lambda \boldsymbol{c}^B \tag{42}$$

with

$$[\boldsymbol{A}]_{ml} = \begin{cases} |y_{ml}|^2 & \text{for} \quad m = l \\ -y_{lm}{y_{ml}}^* & \text{for} \quad m \neq l \end{cases} \tag{43}$$

The constrained LS solution $\boldsymbol{c}^B$ is a unit-norm eigenvector associated to the smallest eigenvalue of $\boldsymbol{A}$. In fact, this yields the direction in the domain $\mathbb{C}^M$ of $J_{cal}(\boldsymbol{c}^B)$ with slowest growth, such that the value of $J_{cal}(\boldsymbol{c}^B)$ is minimized at the intersection of the unit circle $|\boldsymbol{c}^B|^2 = 1$ and the eigenspace of $\boldsymbol{A}$.

- Solution 2: to minimize the LS problem, we compute the partial derivatives of the objective function $J_{cal}(\boldsymbol{c}^B)$ and set them equal to zero, which yields:

$$\boldsymbol{A}\boldsymbol{c}^B = \boldsymbol{0} \tag{44}$$

with $\boldsymbol{A}$ defined as in (43). To avoid the trivial solution $\boldsymbol{c}^B = \boldsymbol{0}$, we impose without loss of generality $c_1^B = 1$. We can then rewrite (44) as in [16]

$$[\boldsymbol{A}_{(:,1)} \quad \boldsymbol{A}_{(:,2:M)}] \begin{bmatrix} 1 \\ \boldsymbol{c}_{(2:m)}^B \end{bmatrix} = 0 \tag{45}$$

and the LS solution is then:

$$\boldsymbol{c}^B = \begin{bmatrix} 1 \\ -\left(\boldsymbol{A}_{(:,2:M)}{}^H \boldsymbol{A}_{(:,2:M)}\right)^{-1} \boldsymbol{A}_{(:,2:M)}{}^H \boldsymbol{A}_{(:,1)} \end{bmatrix} \tag{46}$$

It can be shown that solution 1 (the eigenvector of $\boldsymbol{A}$ corresponding to the smallest eigenvalue) and solution 2 (46) return the same solution for $\boldsymbol{c}^B$ up to a complex scaling factor.

4.3.4 Refinements of the LS solution

The LS objective function also allows introducing refinements in how the different pairs of measurements are used in $J_{cal}(\boldsymbol{c}^B)$ (39).

- Weighted LS: If some knowledge is available about the quality of a given pair of measurement (for example gain differences due to longer or smaller distance between pairs of antenna elements), this knowledge can be introduced in the objective function to obtain a weighted LS objective function:

$$J_{cal}(\boldsymbol{c}^B) = \sum_{\substack{m,l \\ m \neq l}} \gamma_{ml} \left| c_l^B y_{lm} - c_m^B y_{ml} \right|^2 \tag{47}$$

where γ_{ml} is a weighting factor $\in [0,1]$, 1 indicating highest confidence and 0 indicating no confidence. The weighting factor γ_{ml} must of course be taken into account to derive the LS solution as in Section 4.3.2. Example where the weighted LS is useful is when there are significant distance differences between antenna pairs or in the case of conformal antennas where some antenna pairs have a weak coupling due to the array topology.

- Selection-LS: only a subset of the $M(M-1)/2$ pairs of measurements is taken in the LS objective function. Typically, the closest pairs (Neighbour-LS in [14]) or the pairs exhibiting the better coupling are selected. The summation in the objective function (39) then only goes over the set of selected pairs. Note that Selection-LS can be interpreted as weighted LS where selected pairs are given a weight of 1 and discarded pairs are given weight of 0.

4.3.5 Performance

The reciprocity calibration with antenna coupling and the LS solution has been evaluated in [13], [14]. It is reported here again for completeness.

We simulated reciprocity calibration for the case of a 5x20 planar patch array. We used the antenna coupling loss model described in [14] and set the variance of the channel Rayleigh component to -50dB referred to the actual channel coefficient variance. One of the centre antenna elements of the array was defined as the reference. For the general case, modelling the statistics of RF chains responses is a hard task, thus we follow the same approach as [17], where both transmitter and receiver (i.e., t_m^B and r_m^B) have uniformly distributed phase within $[-2\pi, 2\pi[$ and uniformly distributed magnitude between $[1-\epsilon, 1+\epsilon]$ with ϵ such that $\sqrt{\mathrm{E}\{(|t_m^B|-1)^2\}} = \sqrt{\mathrm{E}\{(|r_m^B|-1)^2\}} = 0.1$. We focus on the distinct cases of neighbour antennas and furthest away antennas from the reference one. The latter are positioned at the array edges where coupling to the reference is practically null, thus being the hardest calibration case. Results for others antennas should, in principle, fall within these bounds. For all approaches, we choose to normalize all results with respect to the (calibration) signal-to-noise ratio $\mathrm{SNR}_{\mathrm{Cal}}$ of the neighbour antenna channel. With this normalization it is straightforward to see how different calibration methods "close the gap" between the best and worst calibration scenarios.

At low $\mathrm{SNR}_{\mathrm{Cal}}$ values, Figure 11 shows that the direct-path based estimator (Section 4.3.2) does not possess finite second moment, i.e., the simulated MSE does not converge as the number of simulation runs increases. As for the generalized estimators (Section 4.3.3), the generalized LS estimator shows the worst performance at low $\mathrm{SNR}_{\mathrm{Cal}}$. This is justified by the weak received signals being equally weighted in the cost function. The weighted LS estimator compensates for this, but has worse performance at high $\mathrm{SNR}_{\mathrm{Cal}}$ (by a small margin) since weights are not optimized in an MSE sense. Overall, the neighbour LS scheme works fairly well.

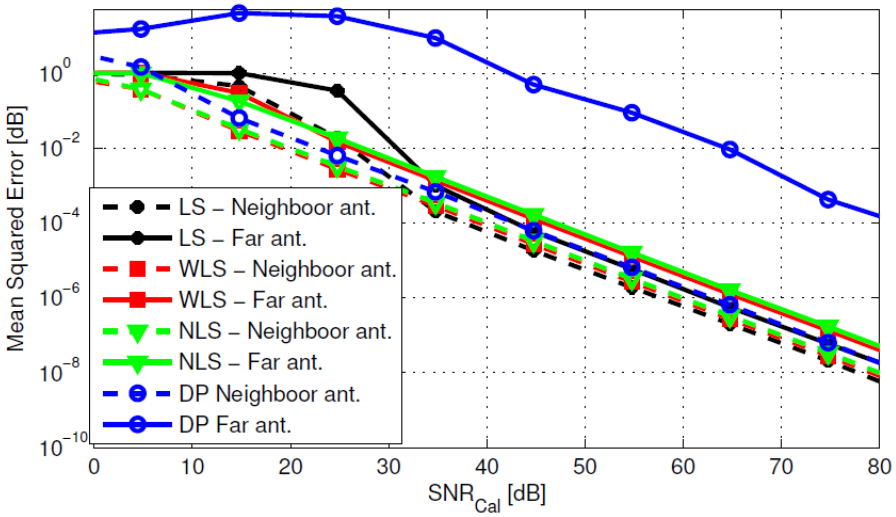

Figure 11 - Mean squared error (MSE) of the calibration coefficients computed for the neighbour and furthest antenna from the reference (from [13]).

4.4 Stability of the reciprocity calibration

The stability of the reciprocity calibration has been investigated among others in [15] and [18].

In [15], a plot of the amplitude and phase deviations over time shows that, after calibration, the average phase deviation is less than 2.6% (maximum 6.7%) and the average amplitude deviation is less than 0.7% (maximum 1.4%), over a period of 4 hours. Notably, these measurements were taken during the day with normal movement around the base station, indicating the calibration procedure is stable in real-world environments. The angle deviation is expressed in percent with respect to 180° i.e. 2.6% phase error is equivalent to 4.7°. This indicates that the calibration scheme can be performed very infrequently, i.e., once a day, and thus has negligible performance overhead. The method used in [15] corresponds to the method with antenna coupling (Section 4.3) and the direct path LS solution (Section 4.3.2).

The authors in [18] indirectly checked the stability requirement of the reciprocity calibration by measuring the stability of a transmit-receive link using transceivers of a Massive MIMO testbed as follows; the signal from a transmitter was split in 4 and input to four different receivers. Observation of the phase of the received signal in the four receivers (Figure 13) gives an indication of the stability of such links, hence of the reciprocity calibration. The deviation at the start of the recording is due to the warming up of the transmitters and receivers. After warm up, the phase drifts is limited to only a few degrees.

These measurement show that the reciprocity calibration can be performed very un-frequently e.g. every hour or even every few hours. It is well important to allow for a warm up period of 15 to 30 minutes before performing the reciprocity calibration. This also suggests that the reciprocity calibration might be sensitive to temperature variations (day-night differences, exposition to the sun, ...) and that this may need some consideration in the planning of the reciprocity calibration phase.

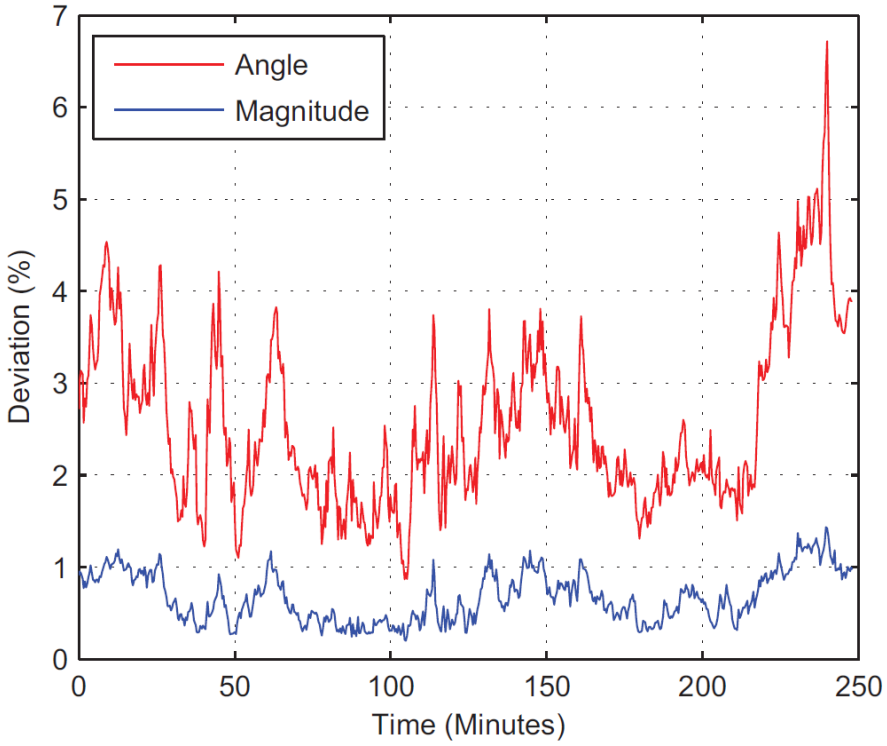

Figure 12 – Stability after calibration (from [15])

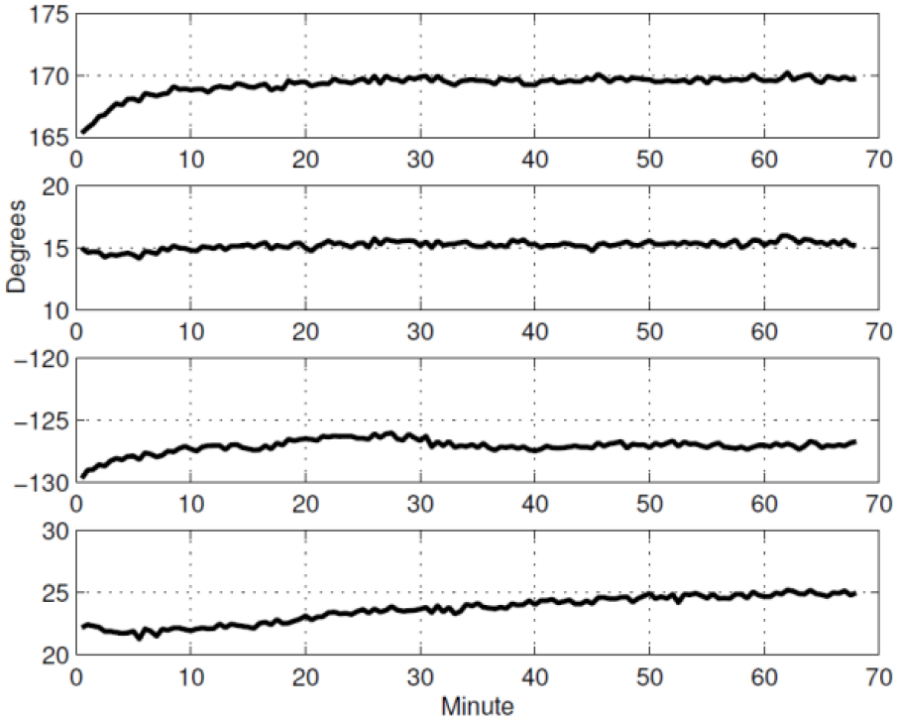

Figure 13 – Stability of a transmit-receive link using transceivers of a massive MIMO testbed (from [18])

4.5 Discussion on the feasibility of the proposed approaches

4.5.1 Over-the-air versus local calibration

Over-the-air calibration provides reciprocity calibration at both the BS and MS sides. In Massive MIMO systems, the precoding is typically only applied at the BS. Hence, the reciprocity calibration is only needed at the BS.

Over-the-air calibration is not the preferred option for the MAMMOET scenarios because:

- MAMMOET considers single-antenna terminals for which reciprocity calibration at the terminal side is not needed;
- OTA calibration requires more time to carry out since several UL and DL sounding and data packets must be exchanged;
- OTA requires more overhead in general.

It follows that local reciprocity at the BS, without intervention of the MS, is preferred. In this case, either the reference transceiver method or the antenna coupling method can be used. The choice between the two methods is the result of a trade-off since both methods have advantages and disadvantages.

4.5.2 Method with a reference transceiver:

- The overhead of one transceiver is marginal in a massive MIMO context since tens to hundreds of transmitters are used;
- The coupling hardware (splitters/combiner/couplers or switches) is a significant overhead since it must accommodate a large number of antennas. The scaling is an issue since adding antennas is not easy especially for the splitter/combiner/coupler configuration. For the switch configuration, the impedance matching on the node where tens or hundreds of switches are connected can be an issue.
- The method with a reference transceiver has a better control of the dynamic range since an (single) attenuator can be added to adapt to the PA power and /or receiver sensitivity level.
- It works equally well with any antenna of the array, irrespective of its location in the array.

4.5.3 Method with antenna coupling:

- There is virtually no hardware overhead for the calibration
- A rough idea of the coupling levels must be known for the weighted LS (although this can in principle be extracted from the magnitudes of the signals measured during the calibration).
- Knowledge of the antenna topology is needed to know which antenna pairs should be excluded from the LS solution computation.
- This method does not have a good control of the dynamic range.
- Since $M(M-1)$ measurements are available, this method should be more accurate than the method with a reference transceiver that takes only M measurements.
- This method can be sensitive to nearby reflectors that could create echoes in the receiving antennas of magnitude similar to the magnitude of the coupled signal but

with some delay due to the two-way time-of-flight of the signal. This could create dips in the measured frequency responses leading to severe inaccuracies.

4.5.4 Calibration of distributed arrays

For distributed arrays, the calibration can be done locally or globally:

- Locally: each sub-array calibrates itself with one of the methods describes earlier (e.g. through antenna coupling); this is the most straightforward way.
- Globally: the sub-arrays calibrate locally and also calibrate each other. The dynamic range of the received signals must be sufficient for such global calibration. The sub-arrays must also be perfectly synchronized for such global calibration but this is also a requirement for the precoding itself hence it is not an issue.

4.5.5 Impact of AGC and PA power levels

If the receiver employs an AGC that can be set dynamically to different gain levels (per antenna), these gain settings must not affect the reciprocity calibration. It might therefore be needed to execute the reciprocity calibration for all possible levels of receiver gain settings. If there are N_{AGC} gain levels, this would multiply the number of measurements by the same amount since all receive antenna can switch the gain setting simultaneously. The same reasoning applies for the transmitter if N_{PA} power amplifier settings are possible and must be calibrated. In total, this would increase the number of measurements and of calibration coefficients by ($N_{AGC} + N_{PA} + 1$).

A point of concern is related to the method with antenna coupling and the AGC and PA settings: it is not obvious that the coupling between antennas will allow all possible AGC settings (if for example the received signal is too weak or too strong) or all PA settings. This requires a specific analysis where the complete transmitter, receiver and antenna array have an impact.

Chapter 5 Conclusions

This deliverable covered in details the impact of non-reciprocity in Massive MIMO systems and how the non-reciprocity can be estimated and compensated.

After a section on the problem statement and system model of non-reciprocity, the impact section showed that

1. Non-reciprocity is only a problem for the BS
2. As for other non-idealities, non-reciprocity requirements are not severe for massive MIMO and depend on the system load, the precoding method and the targeted EVM (through the modulation and coding scheme choice).

Different methods were considered for the calibration. Over-the-air calibration, calibrating the MSs and the BS, does not seem attractive for Massive MIMO, certainly for single antenna MS. The scheme with a reference transceiver is attractive for its fine control capability of the dynamic range during the calibration but requires a non-negligible hardware for coupling the reference transceiver to the to-be-calibrated antennas. On the other hand, the scheme exploiting the natural coupling between antennas does not require any additional hardware but requires careful attention to cope with the measurement dynamic range.

We also pointed out that calibrating all AGC and PA operating points might be needed and that this may turn out to be a challenge for the method with natural antenna coupling.

In summary, the reciprocity calibration must be planned at system level when designing the BS hardware including the antenna array for the method with natural coupling.

As further steps, first of all the findings of this study will be further taken into account in the MAMMOET project. Next, Massive MIMO channel models including the newly observed features during the measurement campaign, will be included in the link level simulator. The state of the art will be further monitored and potential relevant new ideas will be investigated. Last but not least, the validation of the proposed calibration solutions will be considered. These aspects will be addressed in WP4.

Chapter 6 List of Abbreviations

BER	Bit Error Rate
BS	Base Station
CSI	Channel State Information
LS	Least Squares
MaMi	Massive MIMO
MIMO	Multi-input, Multi-output
MISO	Multi-input, Single-output
MMSE	Minimum Mean Square Error
MRT	Maximum Ratio Transmission
MS	Mobile Station
MUI	Multi-user Interference
OFDM	Orthogonal Frequency-division Multiplex
OTA	Over-the-Air
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RF	Radio Frequency
SDMA	Space-Division Multiple Access
SINR	Signal-to-interference-plus-noise Ratio
SISO	Single-input, Single-output
SNR	Signal-to-noise Ratio
TDD	Time-division Duplex
UE	User Equipment
WLAN	Wireless Local Area Network
ZF	Zero-forcing

Chapter 7 Bibliography

[1] E. Larsson, O. Edfors, F. Tufvesson et T. Marzetta, «Massive MIMO for next generation wireless systems,» *Communications Magazine, IEEE,* vol. 52, n° %12, pp. 186-195, February 2014.

[2] S. Thoen et et al, «Adaptive loading for OFDM/SDMA-based wireless networks,» *Communications, IEEE Transactions on,* vol. 50, n° %111, pp. 1798-1810, Nov. 2002.

[3] F. Horlin et A. Bourdoux, Digital Compensation for Analog Front-Ends - A New Approach to Wireless Transceiver Design, John Wiley & Sons, 2008.

[4] IEEE 802.11ac-2013, IEEE Standard for Information technology - Part 11: Wireless LAN MAC & PHY - Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz, IEEE Standards Association, Dec. 2013.

[5] IEEE 802.11ad-2012, IEEE Standard for Information technology - Part 11: Wireless LAN MAC & PHY - Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band, IEEE Standards Association, DEC. 2012.

[6] «IEEE 802.11-2012, IEEE Standard for Information technology - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,» IEEE Standards Association, March 2012.

[7] J. Ketchum et et al., «High-Throughput Enhancements for 802.11: PHY Supplement (doc.: IEEE 802.11-04/1449r1),» Nov. 2004.

[8] M. Guillaud, D. Slock et R. Knopp, «A practical method for wireless channel reciprocity exploitation through relative calibration,» chez *Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on*, Aug. 2005.

[9] J. Liu et et al., «A novel and low-cost analog front-end mismatch calibration scheme for MIMO-OFDM WLANs,» Radio and Wireless Symposium, 2006 IEEE, Jan. 2006.

[10] A. Bourdoux, B. Côme and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: impact and mitigation," in *IEEE Proc. RAWCON*, Aug. 2003.

[11] M. Petermann et et al., «Multi-User Pre-Processing in Multi-Antenna OFDM TDD Systems with Non-Reciprocal Transceivers,» *IEEE Trans. on Communications,* vol. 61, n° %19, pp. 3781-3793, Sept. 2013.

[12] M. Wouters et et al., «An Approach for Real-time Prototyping of MIMO-OFDM Systems,» chez *European Signal Processing Conference, EUSIPCO 2004*, Sept. 2004.

[13] «MAMMOET Deliverable D3.1, First Assessment of Baseband Processing,» Jan. 2015.

[14] J. Vieira, F. Rusek and F. Tufvesson, "Reciprocity calibration methods for Massive MIMO," in *Global Communications Conference (GLOBECOM), 2014 IEEE , vol., no., pp.3708,3712, 8-12 Dec. 2014*.

[15] C. Shepard et et al., «Argos: Practical Many-antenna Base Stations,» chez *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, Mobicom '12*, Aug. 2012.

[16] R. Rogalin et et al, «Hardware-impairment compensation for enabling distributed large-scale MIMO,» chez *Information Theory and Applications Workshop (ITA), 2013*, Febr 2013.

[17] R. Rogalin et et al., «Scalable Synchronization and Reciprocity Calibration for Distributed Multiuser MIMO,» *Wireless Communications, IEEE Transactions on,* vol. 13, n° %14, pp. 1815-1831, April 2014.

[18] F. Tufvesson et A. Bourdoux, «Massive MIMO for 5G: From Theory to Practice,» Tutorial at ICC2015, June 2015.

[19] F. Kaltenberger et et al., «Relative channel reciprocity calibration in MIMO/TDD systems,» chez *Future Network and Mobile Summit, 2010*, June 2010.

[20] J. Liu et et al., «OFDM-MIMO WLAN AP front-end gain and phase mismatch calibration,» chez *Radio and Wireless Conference, 2004 IEEE*, Sept. 2004.