UNIVERSITÉ CÔTE D'AZUR

EURECOM
Sophia Antipolis

École doctorale STIC
Unité de recherche : Automatique, Traitement du Signal et des Images

Thèse de doctorat

Présentée en vue de l'obtention du
grade de docteur en
Sciences et Technologies de l'Information et de la Communication
de
l'UNIVERSITÉ CÔTE D'AZUR

par
Xiwen JIANG

MIMO Massif : transformer le concept en réalité en exploitant la réciprocité du canal

Dirigée par **Florian Kaltenberger**
Co-encadrée par **Luc Deneire**

Soutenue le 4 Octobre 2017

Devant le jury composé de :

David Gesbert	Professeur, EURECOM	Président du jury
Fredrik Tufvesson	Professeur, Lund University	Rapporteur
Guillaume Villemaud	Maître de Conférences, INSA Lyon	Rapporteur
Mark Beach	Professeur, University of Bristol	Examinateur
Maxime Guillaud	Docteur, Huawei Technologies	Examinateur
Florian Kaltenberger	Maître de Conférences, EURECOM	Directeur de thèse
Luc Deneire	Professeur, Université Côte d'Azur	Codirecteur de thèse

UNIVERSITÉ CÔTE D'AZUR

EURECOM
Sophia Antipolis

École doctorale STIC
Specialization: Automatic, Signal and Image Processing

In Partial Fulfillment of the Requirements
for the Degree of Doctor in the scientific domain of
Information Technologies
from
UNIVERSITÉ CÔTE D'AZUR

by
Xiwen JIANG

Massive MIMO: turning concept into reality by exploiting the channel reciprocity

Supervised by **Florian Kaltenberger**
Co-supervised by **Luc Deneire**

Defended on October 4, 2017

Before a committee composed of:

David Gesbert	Professor, EURECOM	President of the Jury
Fredrik Tufvesson	Professor, Lund University	Reviewer
Guillaume Villemaud	Associate Professor, INSA Lyon	Reviewer
Mark Beach	Professor, University of Bristol	Examiner
Maxime Guillaud	Doctor, Huawei Technologies	Examiner
Florian Kaltenberger	Associate Professor, EURECOM	Thesis supervisor
Luc Deneire	Professor, Université Côte d'Azur	Thesis co-supervisor

Abstract

Massive multiple-input multiple-output (MIMO) is considered as one of the key technologies that will enable the next generation of wireless communications. In order to perform downlink (DL) beamforming algorithms with large antenna arrays, the biggest challenge is the acquisition of accurate channel state information at the transmitter (CSIT). To take up this challenge, time division duplex (TDD) is favorable to massive MIMO systems thanks to its channel reciprocity in DL and uplink (UL). However, while the physical channel in the air is reciprocal, the radio-frequency (RF) front-ends in transceivers are not; therefore, calibration should be used in practical systems to compensate the RF hardware asymmetry.

In this thesis, we focus on turning massive MIMO concept into reality based on TDD reciprocity calibration. The contributions can be summarized as follows. First, we propose a unified framework for reciprocity calibration, which generalizes various calibration methods existing in literature, providing a higher level view on the calibration problem as well as opening up possibilities of numerous innovations on calibration methods. Second, based on this general representation, we propose three new calibration schemes: a fast calibration method based on antenna grouping, a calibration scheme for hybrid beamforming architecture, as well as a calibration parameter tracking and system health monitoring mechanism which allows fast detection of parameter change. Third, we carried out measurements of calibration parameters on a real platform in order to reveal the hardware properties. Fourth, we study, from a system point of view, how accurately a TDD massive MIMO system should be calibrated. Last but not least, enabled by TDD reciprocity calibration, we build up an open source long term evolution (LTE) compatible massive MIMO testbed based on the OpenAirInterface platform, which can directly provide Internet service to a commercial device. The testbed demonstrates the feasibility of integrating massive MIMO into current 3rd Generation Partnership Project (3GPP) standards and its usage in 5G can be a smooth evolution from current 4G systems.

Acknowledgments

I still remember my first day at EURECOM. When Florian Kaltenberger, my Ph.D. advisor, guided me around the laboratory, I was astonished by the equipment in front of my eyes, as I was expecting a Ph.D. student's life would be accompanied by papers rather than devices, cables and antennas. Life is full of surprises, and that is the exciting part about it. It later turns out that a research, not only about reading and writing but also about doing something concrete that bridges theory and practice, is a great adventure. This enjoyment would never be possible without Florian's guidance, advices and support. He is always ready to listen to my new ideas, sometimes crazy or stupid ones, then challenges me and helps me go deeper. His impressive qualities in the academy as well as his cool personality gave me a flexible and enjoyable environment to fully explore my potential in accomplishing the research work.

I am also grateful to my co-supervisor Luc Deneire for offering me the opportunity to start the work. His inspiring advices and continued support have greatly helped me solve a lot of challenging problems and make progress in my research.

During the last three years, I had the chance to work with many highly talented researchers, professors and students from both Eurecom and Huawei Technologies (Paris). The joyful and fruitful discussion and cooperation with them either have brought many of my primary ideas to mathematically sound theories or have helped me in realizing and implementing new concepts. I would like to thank Kalyana Gopala, Dirk Slock, Maxime Guillaud, Alexis Decurninge, Raymond Knopp, Cedric Roux for their help. I feel honored to have had the chance to work with them.

I was lucky to meet many friends at EURECOM and more generally in the south of France. Their friendship and the colorful moments we had together will be unforgettable in my life. I would like to express my thankfulness to them: Jingjing, Junting, Robins, Elena, Haifan, Xinping, Yongchao, Shengyun, Qianrui, Chia-Yu, Luhan, Shuang, Pierre-Mayeul, Jianshu and many other friends.

I would like to thank my girlfriend, Yuyao, for her love and spiritual support. Her encouragement makes me go through many difficult moments and pursue higher goals.

Finally, my profound gratitude goes to my parents. They have made tremendous sacrifices throughout their lives to support their son in going out from a small town and realizing his dream. Their unconditional love and long-standing support makes me ambitious to achieve great things in my life.

Contents

List of Figures

List of Tables

Acronyms

Here are the main acronyms used in this document. The meaning of an acronym is usually indicated once, when it first appears in the text.

3GPP	Third Generation Partnership Project
ADC	Analog-to-digital converter
AML	Alternative Maximum Likelihood
AWGN	Additive White Gaussian Noise
BS	Base Station
CDD	Cyclic Delay Diversity
CN	Core Network
CP	Cyclic Prefix
CRB	Cramér-Rao Bound
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CoMP	Coordinated Multiple Points
DAC	Digital-to-Analog Converter
DMRS	Demodulation Reference Signal
EPC	Enhanced Packet Core
FDD	Frequency-Division Duplex
FFT	Fast Fourier Transform
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
i.i.d.	Independent and Identically Distributed
LS	Least Squares
LTE	Long Term Evolution
LTI	Linear Time-Invariant
LNA	Low noise amplifier
MAC	Multiple Access Channel
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MRT	Maximum Ratio Transmission
MSE	Mean Square Error
MU	Multiuser

OAI	OpenAirInterface
OFDM	Orthogonal Frequency Division Multiplexing
OTA	Over-The-Air
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RAN	Radio Access Network
RB	Resource Block
PBCH	Physical Broadcast Channel
PUCCH	Uplink Shared Channel
PUSCH	Uplink Control Channel
RE	Resource Element
RF	Radio Frequency
RRC	Radio Resource Control
RS	Reference Signal
Rx	Receiver
SIMD	Single instruction multiple data
SINR	Signal to Interference and Noise Ratio
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SRS	Sounding Reference Signal
SVD	Singular Value Decomposition
SU	Single-user
TDD	Time Division Duplex
TM	Transmission Mode
TLS	Total Least Squares
Tx	Transmitter
UDP	User Datagram Protocol
UE	User Equipment
ZF	Zero Forcing

Notations

The main notations used in this thesis are list as below. Their meanings will be recalled in each chapter when they are first used.

a	Variable
$\mathbf{a}$	Vector
$\mathbf{A}$	Matrix
a^*	Conjugate of a variable
$\mathbf{A}^*$	Conjugate of a matrix
$\mathbf{A}^T$	Transpose of a matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of a matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of a matrix $\mathbf{A}$
$\mathbf{A}^\dagger$	Moore-Penrose pseudo inverse
$\mathbf{A}^\perp$	The orthogonal complement of the column space of $\mathbf{A}$
$\mathbf{A}_{ij}$	The ij-th entry of a matrix $\mathbf{A}$
$\mathrm{Tr}\{\mathbf{A}\}$	The trace of a matrix $\mathbf{A}$
$\mathrm{Rank}\{\mathbf{A}\}$	The rank of a matrix $\mathbf{A}$
$\mathrm{diag}\{a_1, a_2, \ldots, a_n\}$	A diagonal matrix with $a_1, a_2, \ldots, a_n$ on its diagonal
$\mathrm{vec}(\mathbf{A})$	The vectorization of $\mathbf{A}$ formed by stacking the columns of $\mathbf{A}$ into a single column vector
$\mathbb{E}[\mathbf{A}]$	Expected value of the random matrix $\mathbf{A}$
$\lvert a\rvert$	Absolute value of the complex number a
$\lvert\mathbf{A}\rvert$	The determinant of a matrix $\mathbf{A}$
$\lVert\mathbf{a}\rVert$	The norm of a vector $\mathbf{a}$
$\lVert\mathbf{A}\rVert_F$	The Frobenius norm of a matrix $\mathbf{A}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrix $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} * \mathbf{B}$	Khatri–Rao product (column-wise Kronecker product) of $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{0}$	Zero vector
$\mathbf{I}_K$	$K \times K$ identity matrix
$\mathcal{CN}(0, \sigma^2)$	Complex circularly symmetric Gaussian distribution with zero mean and variance σ^2

Chapter 1

Résumé [Français]

1.1 Abrégé

Entrées multiples, sorties multiples (MIMO) massif est considéré comme l'une des technologies clés de la prochaine génération de communications sans fil. Afin d'effectuer des algorithmes de formation de faisceau en liaison descendante (DL) avec un grand réseau d'antennes, le plus grand défi est l'acquisition d'informations précises d'état de canal à l'émetteur (CSIT). Pour relever ce défi, le duplex à division temporelle (TDD) est favorable aux systèmes MIMO massif grâce à sa réciprocité de canal de la DL et la liaison montante (UL). Cependant, alors que le canal physique dans l'air est réciproque, les front-ends de radiofréquence (RF) dans les émetteurs-récepteurs ne le sont pas; par conséquent, calibration devrait être utilisée dans des systèmes pratiques pour compenser l'asymétrie matérielle RF.

Dans cette thèse, nous nous efforçons de transformer le concept MIMO massif en réalité en utilisant la calibration de la réciprocité TDD. Les contributions peuvent être résumées comme suit. Tout d'abord, nous proposons un cadre unifié pour la calibration de la réciprocité, qui généralise diverses méthodes de calibration existant dans la littérature, offrant une vue supérieure sur le problème de calibration ainsi que l'ouverture de nombreuses innovations sur les méthodes de calibration. Deuxièmement, sur la base de cette représentation générale, nous proposons trois nouveaux schémas de calibration: une méthode de calibration rapide basée sur le groupement d'antennes, un schéma de calibration pour l'architecture hybride de formation de faisceau, ainsi qu'un mécanisme de suivi des paramètres de calibration et de surveillance de la santé du système qui permet une détection rapide du changement de paramètre. Troisièmement, nous avons effectué des mesures des paramètres de calibration sur une plate-forme réelle afin de révéler les propriétés matérielles. Quatrièmement, nous étudions, du point de vue du système, avec quelle précision un système MIMO massif TDD devrait être calibré. Enfin, grâce à la calibration de réciprocité TDD, nous avons construit un banc d'essai pour MIMO massif, qui est compatible avec l'évolution à long terme (LTE) basé sur la plate-forme « open source » OpenAirInterface, et peut directement fournir un service Internet à un appareil commercial. Le banc d'essai démontre la faisabilité d'intégrer le MIMO massif dans les normes actuelles du projet de partenariat de troisième génération (3GPP) et son utilisation dans le 5G peut être une évolution à partir des systèmes 4G actuels.

1.2 Introduction

Les communications sans fil posent les bases du monde connecté. Il permet la connexion de millions d'appareils mobiles qui façonnent notre vie quotidienne. Avec l'augmentation rapide du nombre d'appareils connectés ainsi que l'émergence de nouvelles applications telles que la réalité virtuelle, la réalité augmentée, la conduite autonome et les villes intelligentes, les systèmes de communication sans fil actuels sont confrontés aux défis de débit, latence et couverture. Prenant le débit de données comme exemple, la prochaine génération de systèmes de communication sans fil devrait avoir une augmentation de 1000 fois dans la capacité du réseau [1] pour répondre aux demandes des appareils connectés. Il existe principalement trois approches pour relever ces défis: 1) densification du déploiement cellulaire, 2) recherche de plus de spectre, et 3) augmentation de l'efficacité spectrale. Le déploiement massif de petites cellules et l'utilisation d'ondes millimétriques (mmWave) dans les systèmes de communication sont deux exemples concrets principaux de la première et de la deuxième approche. Pour le troisième, parmi les diverses propositions, entrées multiples, sorties multiples (MIMO) massif est une technologie révolutionnaire qui peut augmenter considérablement l'efficacité spectrale. MIMO massif, initialement conçu dans [2,3], consiste à utiliser un grand nombre d'antennes à la station de base (BS) pour servir simultanément plusieurs ≪ user equipments ≫ (UEs) à travers le spectre de fréquences alloué entier. Dans la transmission en liaison descendante (DL), les antennes de la BS ajustent leurs poids de précodage de sorte que les signaux d'émission provenant d'antennes différentes se renforcent mutuellement à la position de l'UE ciblé et s'annulent à l'emplacement des autres UEs. En liaison montante (UL), la BS applique des procédés similaires de traitement de signaux afin de détecter les signaux multiplex des différents UEs.

L'augmentation du nombre d'antennes fournit naturellement une capacité de réseau plus élevée, réduit la puissance rayonnée grâce au gain de formation de faisceau et augmente le nombre d'UEs simultanément servis. Cependant, plutôt que d'étendre simplement le MIMO multi-utilisateur classique, le MIMO massif utilise un nombre excessif d'antennes à la BS par rapport aux UEs, ce qui conduit à un phénomène appelé ≪ durcissement du canal ≫ et simplifie grandement le traitement du signal et l'allocation des ressources. Le canal vu par un UE est composé de canaux physiques à partir d'antennes différentes à la BS. Lorsque le nombre d'antennes à la BS est grand, grâce à la loi des grands nombres, le canal composite vu à l'UE tend à être un canal de scalaire simple avec un bruit additif. Ce canal équivalent à évanouissements uniformes simplifie grandement l'allocation des ressources et le traitement du signal. En outre, des schémas de codage et modulation standard conçus pour des canaux à bruit blanc gaussien additif (AWGN) peuvent être utilisés avec la possibilité de réduire la latence sur la couche physique introduite par de grands entrelaceur dans les systèmes actuels.

En outre, le MIMO massif est complémentaire aux d'autres innovations technologiques. La recherche dans la coopération des petites cellules est conforme à un système MIMO massif avec une topologie distribuée, nommé ≪ MIMO massif distribué ≫ ou réseau d'accès radio centralisé, collaboratif, nuisible et propre (C-RAN). Systèmes mmWave sont également favorable à être combiné avec le MIMO massif afin de compenser la forte atténuation de la radio en raison de l'absorption atmosphérique lors de sa propagation dans l'air.

Compte tenu de ses grands potentiels, le MIMO massif est considéré comme un transformateur de paradigme des communications sans fil.

1.2.1 Motivation

Le défi fondamental pour libérer totalement le potentiel du MIMO massif réside dans l'acquisition d'informations précises d'état de canal à l'émetteur (CSIT) en DL. L'approche utilisée dans les systèmes traditionnels est de laisser la BS envoyer des pilotes aux UEs qui retournent le CSIT mesuré. Lorsque le nombre d'antennes à la BS augmente, cette approche n'est plus possible puisque les frais généraux dans la UL sont si importants que, au moment où la BS reçoit le retour, les informations sur le canal pourraient déjà être obsolètes. Par conséquent, le MIMO massif est initialement conçu pour fonctionner dans le mode duplex à division temporelle (TDD), car la BS peut obtenir le CSIT via la réciprocité de canal en DL et UL [4]. Pour le duplex à division fréquentielle (FDD), l'acquisition de CSIT pour MIMO massif est toujours une question ouverte, même si des efforts, tels que la réduction des feedbacks en UL [5] ou l'interpolation spatiale [6], ont été faits pour avancer l'état de l'art. Idéalement, on attribue à chaque UE un pilote orthogonal en UL de sorte que la BS apprend le canal en DL pour les UEs indépendamment. Cependant, comme le nombre de séquences pilotes orthogonales est limité par le temps de cohérence du canal divisé par la dispersion du retard de canal de transmission, il est inévitable de réutiliser les pilotes d'une cellule à l'autre, ce qui rend l'estimation du canal en UL pour un UE contaminé par l'autre UE assigné avec la même séquence pilote UL dans une autre cellule [7]. Ce phénomène, appelé ≪contamination de pilote≫, constitue la limite ultime de performance théorique lorsque le nombre d'antennes passe à l'infini. En optimisant les allocations de pilote [8], concevant intelligemment des algorithmes d'estimation de canal [9,10] ou en utilisant les schémas de précodage qui tiennent compte de la structure du réseau [11], les effets causés par la contamination de pilote peuvent être atténués ou éliminés. La réciprocité de canal dans les systèmes TDD signifie que les canaux physiques dans l'air sont les mêmes pour UL et DL dans le temps de cohérence du canal. Cependant, comme l'estimation du canal est une fonction d'un récepteur dans le domaine numérique, le canal qu'il voit contient non seulement le canal physique dans l'air, mais aussi les front-ends à fréquence radio (RF), incluant les composants du convertisseur numérique-analogique (DAC) aux antennes à l'émetteur (Tx) et la partie correspondante, des antennes au convertisseur analogique-numérique (ADC), au récepteur (Rx). Étant donné que les front-ends RF dans Tx et Rx sont différents, le canal d'un point de vue du traitement du signal numérique n'est pas réciproque. Sans tenant compte de l'asymétrie matérielle entraînera une inexactitude dans l'estimation CSIT et, par conséquent, dégradera gravement les performances de formation de faisceau en DL [12–15].

Afin de compenser l'asymétrie matérielle et d'obtenir une réciprocité de canal complet, des techniques de calibration sont nécessaires. Ce sujet a été exploité bien avant l'apparition du concept MIMO massif. Dans [16–20], les auteurs suggèrent d'ajouter dans les émetteurs-récepteurs des composants matériels supplémentaires dédiés à la calibration. Cette méthode compense l'asymétrie de Tx et Rx front-ends RF dans les émetteurs-récepteurs, mais ne parait pas une solution avec un bon rapport coût-efficacité. [21–24] proposent donc des schémas de calibration ≪ over-the-air ≫, où les coefficients de calibration sont estimés en utilisant des méthodes de traitement de signal basées sur l'estimation de canal bidirectionnel entre la BS et l'UE. Puisqu'ils restent relativement stables, une fois que ces coefficients sont obtenus dans la phase d'initialisation du système (phase de calibration), ils peuvent être utilisés plus tard pour ajuster les estimations instantanées des canaux en UL pour évaluer le CSIT. Les méthodes traditionnelles de calibration ≪ over-the-air ≫ incluent UE dans le processus de calibration, ce qui implique que les UE doivent retourner leur canaux en DL estimés. Bien que les coefficients de calibration soient assez stables et que le système

ne répète pas la calibration très fréquemment, le retour des informations de canal en DL pour toutes les antennes BS est encore un processus lourd pour les systèmes MIMO massifs. Une nouvelle catégorie de méthodes qui peuvent calibrer le système en interne à la BS sans aucune assistance d'UE sont donc proposées dans [25–29] pour relever ce défi. Ces méthodes sont appelées «calibration interne». Malgré cet état de l'art, de nombreuses questions sont encore ouvertes :

- Les méthodes existantes sont indépendantes l'une de l'autre ou sont-elles liées de façon inhérente?
- Quelle est la meilleure façon de calibrer un système MIMO massif TDD?
- Quel est le moyen le plus rapide d'atteindre un état calibré?
- Est-il possible de ne pas interrompre le service de données lorsque le système se calibre?
- Comment calibrer un système MIMO massif avec une structure hybride de formation de faisceau?
- Comment surveiller un système dans un état calibré, suivre en continu l'évolution des coefficients de calibration et détecter rapidement les coefficients changés?
- À quoi ressemblent ces paramètres de calibration dans un système pratique?
- Avec quelle précision devons-nous calibrer un système MIMO massif?

De plus, les méthodes de calibration doivent être vérifiées dans un système réel afin de comprendre les questions suivantes du point de vue de l'implémentation:

- Comment utiliser la calibration de la réciprocité pour faire fonctionner un système réel de MIMO massif TDD ?
- Est-il possible d'intégrer le MIMO massif dans les standards 3GPP actuels?
- Quels sont les autres défis et les solutions correspondantes dans la construction d'un banc d'essai de MIMO massif en temps réel compatible à l'évolution à long terme (LTE)?

Motivé par ces questions, cette thèse aborde ces problèmes en utilisant des approches théoriques et pratiques. Dans la prochaine section, nous résumons les résultats obtenus et les contributions.

1.2.2 Contributions

Cette thèse concerne la mise en œuvre du concept MIMO massif. Les contributions peuvent être résumées en deux aspects. Tout d'abord, nous mettons l'accent sur les méthodes de calibration de la réciprocité du canal TDD, y compris la généralisation et l'innovation sur les algorithmes de calibration. Nous avons également effectué des mesures sur les paramètres de calibration de la réciprocité et effectué une étude sur la précision de DL CSIT obtenue. Deuxièmement, grâce à la calibration de la réciprocité, nous avons construit un banc d'essai de MIMO massif conforme à la norme LTE, qui peut directement fournir des services Internet en temps réel aux terminaux mobiles commerciaux.

1.2.3 Méthodes de calibration de la réciprocité du canal TDD

Nous proposons un cadre général pour la calibration de la réciprocité de canal TDD sous le principe de la ≪partition du réseau d'antennes≫. Presque toutes les méthodes de calibration existant dans la littérature peuvent être représentées par ce cadre général. Par exemple, si nous partitionnons le réseau d'antennes en une antenne de référence et un groupe contenant toutes les autres antennes, effectuant une transmission bidirectionnelle conduit à la méthode de calibration d'Argos dans [25]. La méthode dans [26] définir des groupes avec une seule antenne chacun, alors qu'une calibration complète d'Avalanche [29] équivaut à partitionner le réseau d'antennes en groupes avec $\max\{1, i-1\}$ de éléments où i est l'indice du groupe d'antennes. Ce travail combine un effort conjoint de collègues d'Eurecom et de Huawei Technology (Paris). La proposition initiale sur le cadre général par l'auteur de cette thèse a ensuite été grandement améliorée par les contributions de Dirk Slock, Kalyana Gopala, Maxime Guillaud et Alexis Decunringe, avec une représentation plus mathématiquement sonore.

Ces résultats sont présentés dans Chapitre 5 et ont conduit à

- X. Jiang, A. Decunringe, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock et L. Deneire, ≪ A Framework for Over-the-air Reciprocity Calibration for TDD Massive MIMO Systems ≫, submitted to IEEE Trans. on Wireless Commun.

La partition du réseau d'antenne ouvre également de nombreuses nouvelles possibilités pour la calibration de la réciprocité TDD. Dans cette thèse, nous décrivons trois de nos méthodes proposées: calibration rapide, calibration de la réciprocité pour le système hybride de formation de faisceau et méthode de suivi des paramètres de calibration avec une capacité de détection rapide du changement des paramètres. La calibration rapide vise à réduire le nombre de ressources nécessaires pour accomplir le processus de calibration. On peut prouver que, pour un réseau d'antennes avec un nombre d'éléments égal à M, si le canal ne change pas pendant le processus de calibration, le nombre d'utilisation de canal nécessaire pour accomplir la calibration de la réciprocité sont $\mathcal{O}(\sqrt{2M})$ si la calibration est effectué entre les groupes, alors que les méthodes Argos et [26] nécessitent un nombre d'utilisation de canal de l'ordre de $\mathcal{O}(M-1)$. La calibration rapide est un exemple pour illustrer la valeur de la représentation uniforme. Les détails sont présentés dans Chapitre **??**. Pour les structures hybrides, la manière courante pour la transmission consiste à effectuer une formation de faisceau sélectionné dans un ensemble de faisceaux prédéfinis, ce qui entraîne une perte de performance significative dans la plupart des scénarios. Nous proposons ici de partitionner l'ensemble des antennes en deux groupes et d'utiliser multiple transmissions de pilote entre les deux groupes pour obtenir l'estimation du canal entre différentes antennes. Il est alors possible de formuler un problème de moindres carrés (LS) afin d'estimer les paramètres de calibration. Sur la base de cette méthode, nous pouvons acquérir un CSIT presque parfait pour les systèmes de faisceau hybride.

Ces résultats sont présentés dans Chapitre 6 et publiés dans

- X. Jiang and F. Kaltenberger, "TDD channel reciprocity calibration in hybrid beamforming massive MIMO systems", Technical Report TD(17)04085, COST CA-15104 IRACON, Lund, Sweden, May 2017.
- X. Jiang and F. Kaltenberger, "TDD channel reciprocity calibration in hybrid beamforming massive MIMO systems", submitted to IEEE Journal on Sel. Topics in Sig. Proc.

Une autre application de la partition du réseau d'antenne consiste en une méthode de suivi des paramètres de calibration. Comme le matériel ne peut pas éviter les phénomènes physiques tels que la surtension ou l'interruption de courant, il faut un mécanisme pour surveiller la santé du système, surtout la stabilité des paramètres de calibration. Si nous partageons l'ensemble du réseau d'antennes en deux groupes, il est possible d'effectuer une transmission bidirectionnelle entre ces deux groupes et de définir une métrique pour détecter si tous les paramètres restent identiques à ceux de la calibration précédente. Dans le cas où certains paramètres ont changé, nous divisons les deux groupes, adoptons la même idée et utilisons un algorithme d'arbre binaire de recherche pour déterminer le paramètre de quelle chaîne RF a changé. L'avantage du suivi des paramètres est que le système peut éviter d'effectuer le processus de calibration si aucun paramètre ne change et a besoin de corriger seulement les paramètres modifiés autrement. Ces résultats sont présentés dans Chapitre 7 et ont conduit à

- X. Jiang, F. Kaltenberger, A. Decunringe, M. Guillaud, "Antenna array calibration allowing fast correction of sudden impairments variations", submitted to European Patent, application number: PCT/EP2017/056303.

Nous avons effectué des mesures sur les paramètres de calibration de la réciprocité afin de révéler les propriétés matérielles. En particulier, nous avons vérifié l'hypothèse largement adoptée dans la littérature selon laquelle la matrice de calibration peut être supposée diagonale, c'est-à-dire que les coefficients de calibration introduits par la diaphonie RF et le couplage mutuel de l'antenne sont très faibles et peuvent être ignorés dans un schéma de calibration pratique. Les résultats de l'expérience révèlent également l'évolution des amplitudes et des phases des paramètres de calibration dans le domaine fréquentielle. Nous utilisons ensuite les coefficients de calibration obtenus pour estimer le canal en DL à partir de l'estimation de canal en UL et ensuite effectuer une formation de faisceau avec le CSIT. Nous mesurons le rapport signal sur bruit (SNR) et comparons avec le cas où la calibration n'est pas utilisée pour illustrer l'impact de la calibration de la réciprocité sur la performance de la formation du faisceau.

Ces résultats sont présentés dans Chapitre 8 et publiés dans

- X. Jiang, M. Čirkić, F. Kaltenberger, E. G. Larsson, L. Deneire, and R. Knopp, "MIMO-TDD reciprocity and hardware imbalances: experimental results," in Proc. IEEE Intern. Conf. on Commun. (ICC), London, UK, Jun. 2015, pp. 4949–4953.

Nous avons également étudié la précision de CSIT en DL obtenue à partir de la calibration de la réciprocité TDD. Comme le CSIT est calculé en utilisant à la fois le canal en UL estimé et les coefficients de calibration, nous étudions l'impact de ces deux facteurs sur la précision CSIT. Ces résultats sont présentés dans Chapitre 9 et publiés dans

- X. Jiang, F. Kaltenberger, and L. Deneire, "How accurately should we calibrate a massive MIMO TDD system?" in Proc. IEEE ICC 2016 Workshops: Workshop on 5G RAN DESIGN, Kuala Lumpur, Malaysia, May 2016.

1.2.4 Prototype de MIMO massif

Nous avons construit un prototype MIMO massif TDD compatible à LTE basé sur OpenAirInterface, qui peut directement fournir un service Internet à un appareil commercial. Le banc d'essai met en œuvre le mode de transmission (TM) 7 défini dans 3GPP, version

8, où un nombre arbitraire d'antennes physiques peut être mappé sur un port d'antenne logique en utilisant un précodage non-prédéfini. La calibration interne de la réciprocité à la BS est utilisé pendant la phase d'initialisation et les coefficients de calibration obtenus sont directement appliqués sur l'estimation de UL afin d'acquérir le CSIT qui est ensuite utilisé pour calculer les poids de précodage pour la transmission en DL. Nous utilisons des signaux de référence (RS) spécifiques à l'UE qui sont précodés avec les mêmes poids que les données pour la démodulation de TM7. Les techniques de parallélisation de « thread » et « données à plusieurs instructions multiples » (SIMD) sont utilisées pour accélérer la modulation de l'« orthogonal frequency-division multiplexing » (OFDM) afin de permettre une exécution en temps réel. En outre, nous avons étendu la « Radio Resource Control » (RRC) dans OpenAirInterface afin de supporter la transmission TM7. Le banc d'essai MIMO massif OpenAirInterface démontre qu'il est possible d'intégrer la technologie MIMO massif dans la norme 4G actuelle et son évolution vers 5G peut être une évolution à partir de la norme LTE. Dans le cadre de la plate-forme OpenAirInterface, le banc d'essai ouvre ses sources logicielles afin d'aider la recherche universitaire et le développement industriel dans la technologie MIMO massif.

Ces résultats sont présentés dans Chapitre 10 et publiés dans

- X. Jiang, F. Kaltenberger, R. Knopp, and H. Maatallah, "Openairinterface massive mimo testbed : A 5g innovation platform", OpenAirInterface Software Alliance White Paper and Video, http://www.openairinterface.org/?page_id=1760, Sep. 2016.
- X. Jiang and F. Kaltenberger, "OpenAirInterface Massive MIMO Testbed: A 5G Innovation Platform", Technical Report TD(16)02044, COST CA-15104 IRACON, Durham, UK, Oct. 2016.
- X. Jiang and F. Kaltenberger. "Demo: an LTE compatible massive MIMO testbed based on OpenAirInterface", 21st Intern. ITG Workshop on Smart Antennas (WSA), Berlin, Germany, Mar. 2017.
- F. Kaltenberger, X. Jiang and R. Knopp, "From massive MIMO to C-RAN: the OpenAirInterface 5G testbed", accepted in 51st Asilomar Conf. on Sig., Sys. and Comp.

1.3 Conclusion

Cette thèse traite du problème de transformer le MIMO massif d'un concept théorique à des systèmes pratiques, en mettant l'accent sur la calibration de la réciprocité du canal TDD. Correctement et efficacement calibrer l'asymétrie Tx et Rx sur les chaines RF est essentiel pour l'acquisition CSIT basée sur la réciprocité dans les systèmes MIMO massif TDD. Nous proposons un cadre général basé sur l'échange de pilote pour la calibration « over-the-air ». Différentes méthodes de calibration existant dans la littérature, en particulier, ceux qui concernent la calibration interne à la BS, adaptées aux systèmes MIMO massif, peuvent être représentées par ce cadre général à l'aide d'une partition du réseau d'antennes. Estimation optimale et la borne Cramér-Rao (en tant que référence de performance) sont dérivée. Nous avons également discuté sur accumulation cohérente et non cohérente et soulignons qu'il est possible d'intégrer le processus de calibration au service de données, de sorte que la consommation de ressource de calibration disparaisse. Le cadre général ouvre de nombreuses possibilités d'innovations sur la calibration de la réciprocité.

Tout d'abord, effectuer des échanges de pilotes à l'aide de groupes d'antennes plutôt que d'utiliser des éléments individuels peut accélérer le processus de calibration. Nous avons montré que, pour calibrer un réseau d'antennes avec un nombre d'antennes M à l'aide d'une accumulation cohérente, le minimum d'utilisation de canal nécessaire est sur l'ordre de $\mathcal{O}(\sqrt{M})$ plutôt que sur $\mathcal{O}(M)$ pour d'autres méthodes existantes dans la littérature. Deuxièmement, pour un système de formation de faisceau hybride analogique-numérique fonctionnant en mode TDD, nous proposons de partitionner le réseau d'antennes en deux groupes et de calibrer le système en utilisant une transmission de pilote bidirectionnelle entre eux[1]. Un système de formation de faisceau hybride basé sur la calibration de réciprocité donne la possibilité d'acquérir un CSIT parfait en DL, et donc surpasse de manière significative le cas où nous opérons le système en modes traditionnels par entraînement de faisceau basé sur un ensemble de faisceaux prédéfinis. Troisièmement, nous proposons également une méthode de calibration qui peut suivre l'évolution des coefficients de calibration. Il fournit un mécanisme pour surveiller le système dans un état calibré et lorsqu'il y a un changement sporadique et soudain, il permet une détection rapide sur ce paramètre changé. En plus de ces nouvelles méthodes de calibration, nous avons également réalisé une campagne de mesure pour les coefficients de calibration de la réciprocité sur les matériels RF réels. Les résultats révèlent les propriétés de ces coefficients et vérifient, dans un système MISO à petite échelle, l'hypothèse diagonale sur la matrice de calibration qui est largement adoptée dans la littérature. Nous avons également étudié la précision du CSIT obtenu impacté par la précision de la matrice de calibration et de le canal en UL instantanément estimé.

Grâce à la calibration de la réciprocité TDD, nous avons construit un banc d'essai de MIMO massif compatible à LTE sur la plate-forme OpenAirInterface. Il peut interagir avec des appareils commerciaux, ce qui démontre la possibilité d'intégrer un MIMO massif dans les systèmes 4G actuels. Il montre également la faisabilité d'utiliser la calibration pour l'acquisition de CSIT dans un system réel. La mise en œuvre identifie d'autres défis pour la construction de systèmes et propose des solutions correspondantes. Bien que des résultats intéressants aient été établis dans cette thèse, certains problèmes ne sont pas résolus. Dans Chapitre 5 et 6, nous avons comparé les performances de calibration du groupement d'antennes entrelacées et non entrelacées. Cependant, la meilleure façon de regrouper les antennes est toujours une question ouverte. En outre, dans le cadre de la calibration unifiée, une autre dimension de l'innovation réside dans la conception du pilote. Nous n'avons pas pleinement exploré ce sujet dans la thèse actuelle, mais utilisons simplement un pilote Fourier à amplitude constante dans la plupart des simulations. En termes d'expérience et de mesure, une question essentielle est la modélisation des canaux intra-réseau d'antennes. Le canal de champ proche entre deux éléments dans le même réseau d'antennes dépend fortement du type d'antenne et de l'arrangement des éléments d'antenne. Des mesures du monde réel sont nécessaires pour révéler les propriétés et pour établir un modèle approprié. Certains résultats intéressants sont rapportés dans [28]. Nous travaillons également de notre côté pour faire avancer l'état de l'art sur cette recherche. Une autre question concerne la diaphonie RF et l'accouplement mutuel de l'antenne. Dans Chapitre **??**, nous avons vérifié l'hypothèse diagonale sur la matrice de calibration dans un système MIMO à petite échelle. Bien que la modélisation théorique dans [19], ainsi que l'expérience pratique dans la plupart des littératures [25, 28], tous soulignent que la diaphonie RF et l'accouplement mutuel d'antenne peuvent être ignorés, comme nous

[1]Cette méthode est valable pour une architecture de ≪ subarray ≫. Pour une architecture ≪ fully connected ≫, la transmission bidirectionnelle doit être effectuée avec un dispositif assistant ou un UE.

l'avons supposé dans Chapitre **??**-**??**, il vaut toujours une grande valeur pour augmenter l'échelle de l'expérience dans Chapitre **??** pour vérifier cette hypothèse dans un contexte MIMO massif. Du point de vue du développement du banc d'essai et de la mise en œuvre du système, plusieurs aspects doivent être faits pour faire évoluer la version actuelle. Aujourd'hui, le banc d'essai ne peut créer qu'un faisceau étroit vers un seul UE avec TM7. Pour servir simultanément plusieurs UEs, les efforts de développement sont nécessaires pour permettre aux TM8, TM9 ou aux nouveaux TMs qui pourraient apparaître dans les futures normes 3GPP. En outre, la calibration de la réciprocité est effectuée à l'aide de scripts Octave lorsque le système est initialisé, il est intéressant de voir comment ce processus peut être intégré à l'opération en temps réel. D'ailleurs, le MIMO massif co-localisé effectue le traitement du signal dans un PC, ce qui lui permet de traiter tous les calculs localement. L'évolution de cette architecture vers une topologie distribuée, alignée avec l'architecture C-RAN, permet d'affecter une partie des tâches de calcul au cloud alors que d'autres tâches aux unités de radio distribuées. Cela permet de relâcher la contrainte de calcul en temps réel dans le banc d'essai.

Chapter 2

Introduction

Wireless communications lay the foundation of today's connected world. It enables the connection of millions of mobile devices that are shaping our daily life. With the rapidly increase in the number of connected devices as well as the emergence of new applications such as virtual reality, augmented reality, autonomous driving and smart cities, the current wireless communication systems are challenged in data throughput, latency and coverage. Taking the data throughput as an example, the next generation of wireless communication systems is expected to have 1000-fold increase in the network capacity [1] to meet the demand for connected devices.

There mainly exist three approaches to take up these challenges: 1) densification of the cell deployment, 2) search for more spectrum, and 3) increasse in the spectral efficiency. The massive deployment of small cells and use of millimeter waves (mmWave) in communication systems are two main concrete examples of the first and second approaches. For the third, among various proposals, massive multiple-input multiple-output (MIMO) appears to be a breakthrough technology that can dramatically increase the spectral efficiency.

Massive MIMO, originally conceived in [2,3], consists in using a large number of antennas at the base station (BS) to simultaneously serve multiple user equipments (UEs) through the entire allocated frequency spectrum. In the downlink (DL) transmission, the BS antennas adjust its precoding weights so that transmit signals from different antennas reinforce each other at the target UE and cancel out each other at the locations of other UEs. In the uplink (UL), the BS applies similar signal processing methods in order to de-multiplex signals from different UEs.

Increasing the number of antennas naturally provides a higher network capacity, reduces the radiated power thanks to the beamforming gain and increases the number of simultaneously served UEs. However, rather than simply scaling up the classical multi-user MIMO, massive MIMO uses an excess number of BS antennas compared with UEs, which leads to a phenomenon called "channel hardening" and thus greatly simplifies the signal processing and resource allocation. In fact, the channel seen by a UE is composed of physical channels from different antennas at the BS. When the number of antennas at the BS is large, due to the law of large numbers, the composite channel seen at the UE tends to be a simple scalar channel with additive noise. This equivalent frequency flat fading channel greatly simplifies resource allocation and signal processing. Additionally, standard coding and modulation schemes designed for additive white Gaussian noise (AWGN) channels can be used, with the possibility of reducing the latency on the physical layer introduced

by large inter-leavers.

Moreover, massive MIMO is complementary to other technology innovations. The research in small cells cooperation is in line with a massive MIMO system with a distributed topology, named "distributed massive MIMO" or Centralized, Collaborative, Cloud and Clean Radio Access Network (C-RAN). mmWave systems are also favorable to be combined with massive MIMO in order to alleviate the strong radio attenuation due to the atmospheric absorption when propagating in the air.

Given its great potentials, massive MIMO is considered as a paradigm shifter of wireless communications.

2.1 Motivation

The fundamental challenge to fully release the potential of massive MIMO lies in the acquisition of accurate DL channel state information at the transmitter (CSIT). The approach used in traditional systems is to let the BS send pilots to UEs who then feed back the measured DL CSIT to the BS. When the number of antennas at the BS grows, this approach is no longer feasible since the overhead in the UL is so large that, at the time when the BS receives the feedback, the channel information might already be outdated. Therefore, massive MIMO is initially conceived to work in time division duplex (TDD) mode, as the BS can obtain the CSIT through the DL and UL channel reciprocity [4]. For frequency division duplexing (FDD), acquiring accurate enough CSIT for massive MIMO BS is still an open question, although efforts, such as reducing the UL feedback [5] or performing spatial interpolation [6], were made to advance the state of the art.

Ideally, every UE is assigned an orthogonal UL pilot so that the BS learns the DL channel for UEs independently. However, as the number of orthogonal pilot sequences is limited by the channel coherence time divided by the channel delay spread, it is unavoidable to reuse the pilots from one cell to another, making the UL channel estimation for one UE contaminated by the other UE assigned with the same UL pilot sequence in another cell [7]. This phenomenon, named as "pilot contamination" constitutes a theoretical ultimate performance limit when the number of antennas goes to infinity. By optimizing the pilot allocations [8], smartly designing channel estimation algorithms [9, 10] or using precoding schemes that take into account network structure [11], the effects caused by pilot contamination can be mitigated or eliminated.

Channel reciprocity in TDD systems means that the physical channels in the air are the same for UL and DL within the channel coherence time. However, as channel estimation is a function of a receiver in the digital domain, the channel it sees contains not only the physical channel in the air but also radio frequency (RF) front-ends, including the hardware from digital-to-analog converter (DAC) to transmit antennas at the transmitter (Tx) and the corresponding part, from receiving antennas to analog-to-digital converter (ADC), at the receiver (Rx). Since the RF front-ends in Tx and Rx are different, the channel from a digital signal processing point of view is not reciprocal. Not accounting for hardware asymmetry will cause inaccuracy in the CSIT estimation and, as a consequence, seriously degrade the DL beamforming performance [12–15].

In order to compensate the hardware asymmetry and achieve a full channel reciprocity, calibration techniques are needed. This topic has been exploited long before the appearance

of the massive MIMO concept. In [16–20], authors suggest to add additional hardware components in transceivers which are dedicated to calibration. This method compensates the Tx and Rx RF asymmetry in the transceivers, but does not appear to be a cost-effective solution. [21–24] thus put forward "over-the-air" calibration schemes, where the calibration coefficients are estimated using signal processing methods based on bi-directional channel estimation between BS and UE. Since hardware properties can be expected to remain relatively stable, once these coefficients are obtained in the initialization phase of the system (calibration phase), they can be used later to adjust instantaneous UL channel estimations to assess the CSIT.

Traditional "over-the-air" calibration methods involves UE in the calibration process, implying that UEs have to feed their estimated DL channel back to the BS. Although the calibration coefficients are quite timely stable, and the system does not repeat the calibration very frequently, feeding back DL channel information for all BS antennas is still a heavy process for massive MIMO systems. A new category of methods that can calibrate the system internally at BS without any assistance of UE are thus proposed in [25–29] to take up this challenge. These methods are named as "BS internal calibration". Despite these achievements, many questions are still unclear and need to be further investigated:

- Are those existing methods independent with each other or inherently related?
- What is the best way to calibrate a TDD massive MIMO system?
- What is the fastest way to achieve a calibrated status?
- Is it possible to not interrupt the data service when the system is calibrating itself?
- How to calibrate a hybrid beamforming structure massive MIMO systems?
- How to monitor a system in a calibrated status, continuously follow the evolution of the calibration coefficients and quickly detect changed coefficients?
- What do those calibration parameters look like in a practical system?
- How accurately should we calibrate a massive MIMO system?

Additionally, calibration methods should be verified in a real system in order to understand the following questions from a system implementation point of view:

- How to use reciprocity calibration to enable a real world TDD massive MIMO system?
- Is it possible to integrate massive MIMO into current 3GPP standards?
- What are other challenges and corresponding solutions in building up a Long Term Evolution (LTE) compatible real time massive MIMO testbed?

Motivated by these questions, this thesis addresses these problems using both theoretical and practical approaches. In the next section, we summarize the achieved results and the contributions.

2.2 Contributions

This thesis mainly concerns bringing the massive MIMO concept into reality. The contributions can be summarized in two aspects. First, we put a focus on the TDD channel

reciprocity calibration methods, including generalizing and innovating on the calibration algorithms. We also carried out measurements on reciprocity calibration parameters and performed a study on the obtained DL CSIT accuracy. Second, based on reciprocity calibration, we built up an LTE standard compliant massive MIMO testbed, which can directly provide real time Internet services to commercial mobile terminals.

2.2.1 TDD channel reciprocity calibration methods

We propose a general framework for TDD channel reciprocity calibration under the principle of "antenna partition". Almost all existing calibration methods in literature can be represented by this general framework. For example, if we partition the array into a reference antenna and a group containing all other antennas, performing bi-directional transmission using timely orthogonal pilots, leads to the Argos calibration method in [25]. The method [26] consists in defining groups each with one antenna, whereas a full Avalanche calibration [29] is equivalent to partition the antenna array into groups with $\max\{1, i-1\}$ where i is the index of the antenna group. This work combines a joint effort from colleagues in Eurecom and Huawei Technology (Paris). The initial proposal on the general framework from the author of this thesis was later greatly improved by contributions from Dirk Slock, Kalyana Gopala, Maxime Guillaud and Alexis Decunringe with a more mathematically sound representation.

These results are presented in Chapter 5 and has lead to

- X. Jiang, A. Decunringe, K. Gopala, F. Kaltenberger, M. Guillaud, D. Slock, and L. Deneire, "A Framework for Over-the-air Reciprocity Calibration for TDD Massive MIMO Systems", submitted to IEEE Trans. on Wireless Commun.

Antenna partition also opens up many new possibilities for TDD reciprocity calibration. In this thesis, we describe three of our proposed methods: Fast calibration, reciprocity calibration for hybrid beamforming system and calibration parameter tracking with change detection.

Fast calibration aims at reducing the number of channel uses (time slots) needed to accomplish the calibration process. It can be proven that for an antenna array with M antennas, if the channel does not change during the calibration process, the channel uses needed to accomplish the reciprocity calibration is $\mathcal{O}(\sqrt{2M})$ if the calibration is performed between groups whereas both Argos method and the method in [26] need a number of channel uses on the order of $\mathcal{O}(M-1)$. The details are presented in Chapter 5 as an example to illustrate the value of the uniform representation.

For hybrid structures, the common way of enabling beamforming transmission is to perform beam training in a pre-defined beam set, which leads to significant performance loss in most scenarios. We propose here to partition the whole antenna array into two groups and to use multiple pilot transmission between the two antenna groups to achieve the channel estimation between different antennas. It is then possible to formulate a least squares (LS) problem in order to estimate calibration parameters. Based on this method,we can achieve near perfect CSIT for hybrid beamforming systems.

These results are presented in Chapter 6 and published in

- X. Jiang and F. Kaltenberger, "TDD channel reciprocity calibration in hybrid beamforming massive MIMO systems", Technical Report TD(17)04085, COST CA-15104

IRACON, Lund, Sweden, May 2017.

- X. Jiang and F. Kaltenberger, "TDD channel reciprocity calibration in hybrid beamforming massive MIMO systems", submitted to IEEE Journal on Sel. Topics in Sig. Proc.

Another application using antenna partition consists in calibration parameter tracking. As hardware can not avoid physical phenomenons such as power surge or power interruption, a mechanism to monitor the system health including the consistency of calibration parameters is needed. If we partition the whole antenna array into two groups, it is possible to perform one bi-directional transmission between these two groups and define a metric to detect if all parameters stay the same as previous calibration. In the case where some parameters changed, we further partition the two groups, adopt the same idea and use a binary search tree algorithm to find out which RF chain has its calibration parameter changed. The benefit of parameter tracking is that the system can avoid performing the calibration process if no parameter changes, and needs to only recalibrate the changed parameters otherwise.

These results are presented in Chapter 7 and has lead to

- X. Jiang, F. Kaltenberger, A. Decunringe, M. Guillaud, "Antenna array calibration allowing fast correction of sudden impairments variations", submitted to European Patent, application number: PCT/EP2017/056303.

We carried out measurements on reciprocity calibration parameters in order to reveal the hardware properties. Especially, we verified the widely adopted assumption in literature that the reciprocity calibration matrix can be assumed diagonal, i.e. calibration coefficients introduced by RF crosstalk and antenna mutual coupling are very small and can be ignored in a practical calibration scheme. The results from the experiment also reveal the evolution of the calibration parameters' amplitudes and phases on frequency. We then use the obtained calibration coefficients to estimate the DL channel from UL channel estimation and then perform beamforming with the CSIT. We measure the beamforming signal-to-noise ratio (SNR) and compare with the case where calibration is not used in order to illustrate the impact of the reciprocity calibration on the beamforming performance.

These results are presented in Chapter 8 and published in

- X. Jiang, M. Čirkić, F. Kaltenberger, E. G. Larsson, L. Deneire, and R. Knopp, "MIMO-TDD reciprocity and hardware imbalances: experimental results," in Proc. IEEE Intern. Conf. on Commun. (ICC), London, UK, Jun. 2015, pp. 4949–4953.

We also studied the DL CSIT accuracy obtained from TDD reciprocity calibration. As the CSIT is calculated using both estimated UL channel and calibration coefficients, we study the impact of both factors on the CSIT accuracy.

These results are presented in Chapter 9 and published in

- X. Jiang, F. Kaltenberger, and L. Deneire, "How accurately should we calibrate a massive MIMO TDD system?" in Proc. IEEE ICC 2016 Workshops: Workshop on 5G RAN DESIGN, Kuala Lumpur, Malaysia, May 2016.

2.2.2 Massive MIMO prototyping

We built up an LTE compatible TDD massive MIMO prototype based on OpenAirInterface, which can directly provide Internet service to a commercial device. The testbed implements Transmission Mode (TM) 7 defined in 3rd Generation Partnership Project (3GPP) Release 8, where an arbitrary number of physical antennas can be mapped to a logical antenna port using non-codebook-based precoding. BS internal reciprocity calibration is used during the initialization phase and the obtained calibration coefficients are directly applied on the UL channel estimation in order to acquire the DL CSIT which is further used to calculate the DL precoding weights used in logical to physical antenna mapping. We use UE specific reference signals (RS) which are precoded with the same weights as the data for TM7 demodulation. Thread parallelization and "single instruction multiple data" (SIMD) techniques are used to speed up the precoding and orthogonal frequency-division multiplexing (OFDM) modulation in order to enable the testbed to run in real time. Additionally , we extended the Radio Resource Control (RRC) signaling in OpenAirInterface in order to support TM7 transmission.

The OpenAirInterface massive MIMO testbed demonstrates that it is possible to integrate massive MIMO technology into current 4G standard and its evolution towards 5G can be a smooth evolution starting from the LTE standard. As a part of OpenAirInterface platform, the testbed opens its software sources in order to help the academic research and industrial development in massive MIMO technology.

These results are presented in Chapter 10 and published in

- X. Jiang, F. Kaltenberger, R. Knopp, and H. Maatallah, "Openairinterface massive mimo testbed : A 5g innovation platform", OpenAirInterface Software Alliance White Paper and Video, http://www.openairinterface.org/?page_id=1760, Sep. 2016.
- X. Jiang and F. Kaltenberger, "OpenAirInterface Massive MIMO Testbed: A 5G Innovation Platform", Technical Report TD(16)02044, COST CA-15104 IRACON, Durham, UK, Oct. 2016.
- X. Jiang and F. Kaltenberger. "Demo: an LTE compatible massive MIMO testbed based on OpenAirInterface", 21st Intern. ITG Workshop on Smart Antennas (WSA), Berlin, Germany, Mar. 2017.
- F. Kaltenberger, X. Jiang and R. Knopp, "From massive MIMO to C-RAN: the OpenAirInterface 5G testbed", accepted in 51st Asilomar Conf. on Sig., Sys. and Comp.

Chapter 3

Massive MIMO Background

3.1 Fundamentals of massive MIMO

Multiple antenna technology, also known as MIMO consists in exploiting the spatial dimension to increase the spectral efficiency per time and frequency resource. In this section, we firstly give a brief review on the evolution of MIMO technology, starting from single-input single-out (SISO), going through single-user MIMO (SU-MIMO), multi-user MIMO (MU-MIMO) and arriving to massive MIMO. We then present the fundamentals, benefits and challenges of the massive MIMO technology. This overview mainly follows the content in [3, 7, 30, 31].

3.1.1 From SISO to MIMO

In the most basic wireless communication system, where both the transmitter and receiver have a single antenna, which is known as SISO system, the fundamental signal model under the point-to-point deterministic channel with Gaussian noise is given by

$$y = \sqrt{\rho}x + n, \tag{3.1}$$

where n is the receiver noise following a standard circularly symmetric complex Gaussian distribution $\mathcal{CN}(0, 1)$ [32]. x is the transmitted signal with a zero mean and unit power constraint such as $\mathbb{E}[x] = 0$ and $\mathbb{E}\{|x|^2\} \leq 1$. ρ is a constant scaling the transmit signal. Since in our model both n and x have normalized variance, ρ is interpreted as the SNR. The Shannon capacity [33] in terms of bit/s/Hz is

$$C = \log(1 + \rho), \tag{3.2}$$

which is achieved when the input signal x follows a Gaussian distribution.

The introduction of multiple antennas, named MIMO, offers an approach to substantially improve the special efficiency. The usage of multiple antennas can be classified into three main categories: transmit/receive diversity, spatial multiplexing and beamforming. In transmit/receive diversity, transceivers usually exploit the spatial diversity using space-time or space-frequency coding to enhance the transmission robustness in a fading channel. Spatial multiplexing consists in using multiple antennas to transmit multiple data schemes in parallel to increase the channel capacity. For beamforming, precoding is used on different

antennas in order to create a virtual antenna pattern to enhance the received signal on the target user and to reduce interference to other users.

The first way to use MIMO for spatial multiplexing is the SU-MIMO, also known as the point-to-point MIMO, where the BS equipped with an antenna array send multiple data streams to a multi-antenna UE. Assuming M and K being the number of antennas at the BS and UE, the DL[1] channel capacity in the presence of additive white Gaussian noise at the receiver is

$$C_{dl} = \log_2 \left| \mathbf{I}_K + \frac{\rho_{dl}}{M} \mathbf{G}^H \mathbf{G} \right| \overset{(a)}{=} \log_2 \left| \mathbf{I}_M + \frac{\rho_{dl}}{M} \mathbf{G}\mathbf{G}^H \right| \tag{3.3}$$

where $\mathbf{G}$ is the $M \times K$ channel matrix. ρ_{dl} is the DL SNR, proportional to the total transmission power at the BS. $\mathbf{I}_M$ is an identity matrix of size M. In (a), we used Slyvecter's determinant theorem $|\mathbf{I} + \mathbf{A}\mathbf{A}^H| = |\mathbf{I} + \mathbf{A}^H\mathbf{A}|$. To achieve the capacity in (3.3), the receiver needs to have the knowledge of $\mathbf{G}$ whereas the transmitter does not need to know $\mathbf{G}$. Note that if the transmitter has the DL channel information, the channel capacity can be improved by optimizing the power allocation on the transmit signal following the water-filling principle, however, this is seldomly used in practice since the benefits would almost never justify the effort in obtaining the CSIT.

Theoretically, C_{dl} scales linearly with $\min(M, K)$, implying that by simultaneously scaling up the number of antennas at the BS and UE, it is possible to increase the channel capacity. However, in practice, due to the limitation of antenna array size and the line-of-sight (LoS) propagation condition, it is difficult to have $\min(M, K)$ independent channels for parallel data streams transmission. Especially, at the UE side, the limitation of the device size and cost constraint prohibits the usage of a large number of antennas and independent RF chains, as well as real time data processing with high complexity. Additionally, near the cell edge, where the SNR is low due to high path loss, the spectral efficiency scales slowly with $\min(M, K)$.

The size limitation to host many antennas at the UE can be solved by breaking the K antennas into multiple UEs, which yields the MU-MIMO, where the BS serves multiple UEs using the same time and frequency resource. The DL sum spectral efficiency is given by

$$C_{dl} = \max_{\mathbf{v}} \log_2 \left| \mathbf{I}_M + \rho_{dl} \mathbf{G}\mathbf{D}_{\mathbf{v}}\mathbf{G}^H \right| \tag{3.4}$$

$\mathbf{D}_{\mathbf{v}} = \text{diag}\{v_1, v_2, \cdots, v_K\}$ with $v_k \geq 0$ and $\sum_{k=1}^{K} v_k \leq 1$ is the power allocation strategy at the BS constructed. ρ_{dl} is the DL SNR. Both the BS and UE needs to have the CSI to achieve the DL performance. Note that, assuming an equal power allocation constraint at the BS, $\mathbf{D}_{\mathbf{v}} = \frac{1}{M}\mathbf{I}_K$, (3.4) collapses to (3.3).

MU-MIMO is much less sensitive to assumptions on the propagation environment thanks to the fact that multiple UEs can be physically located in different places, e.g., LoS conditions are stressing for SU-MIMO but not for MU-MIMO. Besides, it is possible to use single antenna UE in MU-MIMO, enabling low cost and legacy UEs to benefit from the MIMO technology. However, to achieve (3.4), both BS and UE must know $\mathbf{G}$, which implies that substantial resources are needed for pilot transmission. At the same time both BS and UE requires complicated signal processing in MU-MIMO. These two facts limits the practical use of MU-MIMO in its original form.

[1] In this thesis, we concentrate on the DL transmission which is more important and challenging in terms of system performance.

3.1.2 Massive MIMO fundamentals

Massive MIMO, originally introduced in [2, 3], is an extension of MU-MIMO. In this technology, the BS equipped with a large number of antennas serves a cell with a large number of UEs. The fundamental distinctions between massive MIMO and MU-MIMO can be summarized in three aspects: first, only the BS learns the CSI $\mathbf{G}$; second, the number of BS antennas, M, is typically much larger than K; third, simple linear signal processing, such as maximum ratio transmission (MRT) or zero-forcing (ZF), is used both on the UL and on the DL.

Let us assume $\mathbf{G} \in \mathbb{C}^{M\times K}$ is a product of a small scale fading matrix $\mathbf{H} \in \mathbb{C}^{M\times K}$ and a large scale fading matrix $\mathbf{D}_\beta^{1/2} \in \mathbb{C}^{K\times K}$ accounting for path loss and shadow fading, such as $\mathbf{G} = \mathbf{H}\mathbf{D}_\beta^{1/2}$, where the elements of $\mathbf{H}$ have magnitudes of one and $\mathbf{D}_\beta^{1/2}$ is a diagonal matrix with the vector β as its diagonal, $\mathbf{D}_\beta^{1/2} = \mathrm{diag}\{\beta_1, \beta_2, \cdots, \beta_K\}$. As M becomes large ($M \gg K$), the column-vectors of the propagation matrix are asymptotically orthogonal, thus

$$\frac{1}{M}(\mathbf{G}^H\mathbf{G}) = \mathbf{D}_\beta^{1/2}\frac{1}{M}(\mathbf{H}^H\mathbf{H})\mathbf{D}_\beta^{1/2} \approx \mathbf{D}_\beta \tag{3.5}$$

When M grows without bound, the asymptotic sum rate is

$$\begin{aligned} C_{dl} &= \max_{\mathbf{v}} \log_2 \left|\mathbf{I}_M + \rho_{dl}\mathbf{G}\mathbf{D}_\mathbf{v}\mathbf{G}^H\right| \\ &= \max_{\mathbf{v}} \log_2 \left|\mathbf{I}_K + \rho_{dl}\mathbf{D}_\mathbf{v}^{1/2}\mathbf{G}^H\mathbf{G}\mathbf{D}_\mathbf{v}^{1/2}\right| \\ &\approx \max_{\mathbf{v}} \log_2 \left|\mathbf{I}_K + M\rho_{dl}\mathbf{D}_\mathbf{v}\mathbf{D}_\beta\right| \\ &= \max_{\mathbf{v}} \sum_{k=1}^{K} \log_2(1 + M\rho_{dl}v_k\beta_k), \end{aligned} \tag{3.6}$$

which is the sum of the spectral efficiencies of the K terminals. It is shown in [31] that even with the simplest MRT, we can achieve this sum-capacity.

Massive MIMO does not only increase the sum spectral efficiencies, but also brings many other benefits. For example, as the channel seen by each terminal is a combination of the channels from a large number of BS antennas, the law of large numbers leads the combined channel to be deterministic with small-scale fading and frequency dependence disappear when M grows. This phenomenon is known as "channel hardening". It makes many traditional system designs against small scale fading or fading in the frequency domain irrelevant. The main benefits of massive MIMO are summarized as follows [7] :

- Significantly increasing the network capacity and improving the radiated energy efficiency;
- Being possible to be built with inexpensive, low-power hardware components, especially with low cost power amplifiers since the output signal per antenna is small;
- Reducing the latency on the air interface, since small-scale fading disappears;
- Simplifying the multiple access as the channel hardens in the frequency domain and each terminal can be given the whole bandwidth;
- Increasing the robustness against interference with many excess degrees of freedom.

To bring the massive MIMO concept to reality, there are many challenges to address, among which [7] :

- **Channel reciprocity calibration**

  To fully exploit the potential of the large number of antennas, BS needs to have accurate CSI. Massive MIMO relies on the channel reciprocity in TDD to acquire the DL CSI based on UL channel estimation. However, as the transmit and receive hardware in transceivers are not symmetric, reciprocity calibration is essential for CSIT acquisition. Calibrating massive MIMO and maintaining the system in a calibrated status is an important topic and will be the focus of this thesis.

- **Pilot contamination**

  As the BS relies on the UL channel estimation for CSIT acquisition, in a multi-cell scenario, it is unavoidable to reuse the UL pilot from one cell to another, since the maximum number of orthogonal pilot sequences is upper-bounded by the duration of the coherence interval divided by the channel delay spread. Pilot reuse lead to the fact that the UL channel estimation for one UE is contaminated by UEs in other cells sharing the same UL pilot sequences. Pilot contamination constitutes an ultimate limit on performance when M becomes large.

- **Fast and distributed coherent signal processing**

  With the increase of network capacity, the massive MIMO BS has to process a huge amount of data in real time. This is very challenging in system implementation as we'll show in our own implementation experience in Section 10.5. Designing signal processing algorithms with low complexity and the possibility to distribute the calculation to different modules in the system is another important question.

There are many other challenges, such as designing the system against hardware impairments, characterizing massive MIMO channel, building up massive MIMO with low cost hardware, reducing the baseband processing power consumption, system operation with no or limited CSI, enabling massive MIMO in FDD, etc.

Research efforts in recent years have greatly pushed the frontier of the study of massive MIMO and many solutions were proposed to address these various challenges with different approaches. In this thesis, we mainly focus on the channel reciprocity calibration, based on which we built up a 3GPP standard compatible massive MIMO system, which can further make its contribution in turning massive MIMO into reality.

3.2 MIMO and massive MIMO in 3GPP standards

Although the usage of beamforming can be tracked back to more than 60 years ago in military applications, the introduction of MIMO technology into commercial wireless networks starts around the year 2000 [34]. The first adoption of MIMO in 3GPP standards happens in Release 7 version of High Speed Downlink Packet Access (HSDPA). Soon later, from Release 8 on, MIMO has been considered as a key technology on the physical layer in the LTE standardization.

3.2.1 Transmission modes and antenna ports

LTE defines 10 different transmission modes (TMs) for different usage of multiple antennas (transmit diversity, spatial multiplexing and beamforming), as shown in Table 3.1, where TM 1-7 are defined in Release 8, TM 8 is added in Release 9 and in Release 10 and 11, 3GPP introduced TM 9 and TM 10, respectively.

Among these TMs, TM 1 is the single antenna scheme (SISO), TM 2 corresponds to transmit diversity, TM 3-6 are spatial multiplexing schemes and TM 7-10 uses beamforming[2]. For TM 7-10, 3GPP does not specify the precoding weights and leaves the flexibility to system implementation. Later, in this thesis, we show that as it is possible to acquire very accurate CSIT by exploiting the TDD channel reciprocity, we can take full use of this flexibility and calculate the beamforming weights based on the CSIT.

Transmission modes	**Description**	**DCI Format (Main)**	**Comment**
1	Single transmit antenna	1/1A	Single antenna port 0
2	Transmit diversity	1/1A	2 or 4 antennas ports 0,1 (2, 3)
3	Open loop spatial multiplexing with cyclic delay diversity (CDD)	2A	2 or 4 antennas ports 0,1 (2, 3)
4	Closed loop spatial multiplexing	2	2 or 4 antennas ports 0,1 (2, 3)
5	Multi-user MIMO	1D	2 or 4 antennas ports 0,1 (2, 3)
6	Closed loop spatial multiplexing using a single transmission layer	1B	1 layer (rank 1), 2 or 4 antennas ports 0,1 (2, 3)
7	Beamforming	1	Single antenna port, port 5 (virtual antenna port, actual antenna configuration depends on implementation)
8	Dual-layer beamforming	2B	Dual-layer transmission, antenna ports 7 and 8
9	8 layer transmission	2C	Up to 8 layers, antenna ports 7 - 14
10	8 layer transmission	2D	Up to 8 layers, antenna ports 7 - 14, Extension of TM9 for Coordinated Multi-Point (CoMP)

Table 3.1: Downlink transmission modes in LTE Release 12.

Transmission modes usually comes together with the "antenna port", another notion defined in 3GPP. An "antenna port" represents a logical antenna rather than a physical antenna. Different TMs can use different antenna ports for its transmission, e.g., TM1 uses antenna port 0, TM2 uses antenna port 0, 1, if two antenna ports are configured. TM 7 uses antenna port 5, a virtual antenna port for the data transmission. Similar to TM7, TM8 uses virtual antenna port 7 and 8 whereas TM9 and TM10 uses antenna port 7-14. Logical antenna ports are mapped to physical antennas. We show in Section 10.5 how this mapping is possible in a practical system.

In Table. 3.1, the third column shows the main Downlink Control Indicator (DCI) formats

[2]TM 6 can also be regarded as beamforming, TM 8-10 can be seen as a combination of beamforming and spatial multiplexing where the BS is capable of beamforming at least two data schemes to one or multiple UEs.

used for each TM. DCI is the DL control information carrying resource allocation, Power Control Command, CSI Report Request or Channel Quality Indicator (CQI) Report Request, etc. As for different TMs, the control information needed are different, specific DCI formats indicating different set of information the corresponding DCI can carry are specified for each TM. Note that there can be other possible DCI formats associated to each TM. Table. 3.1 list the main DCI formats.

3.2.2 Towards massive MIMO in 3GPP standardization

The enhancement of MIMO in 3GPP is given the official name of Full-Dimension MIMO (FD-MIMO), targeting scaling up the number of antennas at the transmitter side. The antennas with a two dimensional (2D) array structure was used as a starting point [35]. A 2D antenna array can accommodate a large number of antenna elements in a compact space, making the installation and deployment easier. More importantly, it gives the antenna the flexibility to control the transmit signal in both elevation and azimuth dimensions, enabling beamforming in 3D space.

The discussion of FD-MIMO was started in Release 13[3], from which onwards, the specifications are given the name of LTE-Advanced Pro. (The standardization of 5G New Radio (NR) will be from Release 15 onwards). In Release 13, the CSI Reference Signal (CSI-RS) [4] was extended to 16 although no new transmission mode is added. In release 14, the CSI-RS are further extended to 32.

Apart from that, the main discussions on FD-MIMO in 3GPP can be found in its study items, a process done before a formal standardization process. [35] gives an overview of these discussions. The main points are summarized as follows.

- **Deployment scenarios**

  Two typical deployment scenarios are chosen for the design and evaluation of FD-MIMO systems: 3D urban macro scenario (3D-UMa) and 3D micro scenario (3D-UMi). In the first scenario, antennas are installed on rooftops, higher than surrounding buildings. Diffraction over the rooftop is dominant, thus down-tilt transmission is desirable. In the second scenario, antennas are lower than surrounding buildings and users can be both higher or lower than the base station depending on the situation, so both down-tilt and up-tilt transmission need to be considered.

- **Antenna configurations**

  Active antennas with gain control on each antenna element will be used to replace traditional passive antennas. The antenna arrangement can be described by three parameters: the number of elements in the vertical direction, the number of elements in the horizontal direction, and the polarization degree (co-polarization or cross-polarization).

- **TXRU architecture**

  The TXRU architecture is also known as hybrid beamforming structure, where the TXRU (digital chains) are connected with the antenna array via a radio distribution

[3] Release 12 has mainly focused on the use of LTE technology for emergency and security services, Small cells and Network densification, Device to Device (D2D) communications, etc. MIMO enhancement is rarely addressed in that release.

[4] CSI-RS is a type of DL RS added in Release 10 for DL channel estimation for up to 8 layers transmission.

network (RDN). Two options are suggested: array-partition and array-connected architecture. For more details about this architecture, please refer to Chapter 6, where the two options are named subarray architecture and fully (or not necessarily fully) connected architecture.

- **New CSI-RS transmission and CSI feedback strategy**

  Different TXRU architecture can use different CSI-RS transmission and CSI feedback strategy. For partitioned antenna arrays, non precoded CSI-RS are transmitted from each partition. Upon the reception of the CSI-RS, UE decides the precoding weights from a codebook and feeds back the information to the BS. In order to reduce the feedback payload, the codebook can be designed as a composite codebook combined by vertical and horizontal codebooks. For connected antenna arrays, beamformed CSI-RS can be transmitted to the UE. The UE then selects the best beam index and feeds it back to the BS. Besides, a combination of non-precoded CSI-RS transmission and beamforming CSI-RS transmission can be smartly combined to have benefits from both strategies.

3.3 Summary

In this chapter, we reviewed the historical evolution of MIMO technology and the fundamentals of massive MIMO. The main benefits and challenges of massive MIMO are summarized. We then gave a glance of the ongoing discussions in 3GPP standardization about FD-MIMO, an official name of massive MIMO.

In the sequel, we try to combine theory and practice, by enabling massive MIMO in the current 3GPP 4G standard, in order to see the possibility of its evolution towards 5G. To this end, we firstly focus on CSIT acquisition by exploiting TDD channel reciprocity and then describe our massive MIMO testbed implementation on the OpenAirInterface platform.

Chapter 4

State of the Art on Over-the-air Reciprocity Calibration

Accurate DL CSIT is essential for MIMO precoding in order to fully release the potential of multiple antenna technology. In a TDD system, channel reciprocity can be exploited to obtain DL CSIT from UL channel estimation. However, CSIT acquisition and MIMO precoding, without taking into account the RF asymmetry, will cause severe performance degradation [36] [14]. Various solutions to this issue have been proposed. One possible solution is to design a robust precoding scheme [37], at the cost of extra implementation complexity. Another possibility is to use a calibration scheme to compensate the RF asymmetry. One of these methods is the (absolute) calibration of each RF circuit involved [16], requiring additional hardware. Another alternative, limited to low-power transmission, is to use a specially crafted transceiver where the same op-amp is used for both transmitting and receiving [38]. A third alternative [21, 22], termed "over-the-air" calibration, achieves the same effect as normal calibration without the requirement for extra hardware. Assuming that the impairments due to imperfect RF chains can be modeled as linear time-invariant (LTI) filters, it was shown in [21] that the channels measured in both directions could be related by a simple convolution operation in the time domain (multiplication operation in the frequency domain), thus alleviating the need for hardware-based calibration. Appearing to be a cost-effective solution that attracts most attention from the research community, "over-the-air" calibration will be the focus of this thesis. In this chapter, we describe the basic idea of "over-the-air" reciprocity calibration in a practical TDD system.

4.1 System Model

Let us consider a system as in Fig. 4.1, where A represents a BS and B is a UE, each containing M_A and M_B antennas. The DL and UL channel seen in the digital domain are noted by $\mathbf{H}_{A\to B}$ and $\mathbf{H}_{B\to A}$. In the frequency domain, they can be represented by

$$\begin{cases} \mathbf{H}_{A\to B} = \mathbf{R}_B \mathbf{C}_{A\to B} \mathbf{T}_A \\ \mathbf{H}_{B\to A} = \mathbf{R}_A \mathbf{C}_{B\to A} \mathbf{T}_B, \end{cases} \tag{4.1}$$

where $\mathbf{T}_A$, $\mathbf{R}_A$, $\mathbf{T}_B$, $\mathbf{R}_B$ represent the transmit and receive RF front-ends for BS and UE respectively. The size 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 transmitter and receiver coefficients of RF front-ends, whereas the off-diagonal elements correspond to RF crosstalk and antenna mutual coupling[1]. It is worth noting that if there is no (or negligible) RF crosstalk or antenna mutual coupling, $\mathbf{T}_A$, $\mathbf{R}_A$, $\mathbf{T}_B$, $\mathbf{R}_B$ all become diagonal matrices. Both $\mathbf{H}_{A\to B}$ and $\mathbf{H}_{B\to A}$ can be represented by $M_A \times M_B$ independent scalar equations, each equation corresponds to the single-input single-output (SISO) channel between one antenna element in A and another one in B.

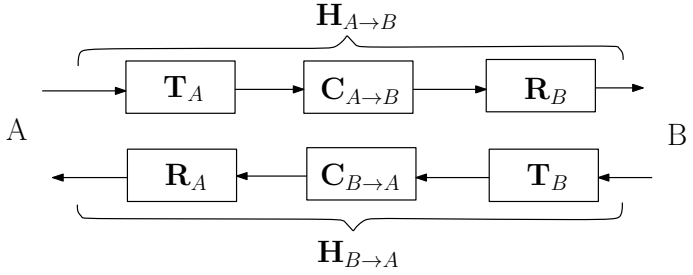


Figure 4.1: Reciprocity model.

As the system is operating under TDD mode, the DL and UL physical channels in the air enjoy reciprocity within the channel coherence time, i.e. $\mathbf{C}_{A\to B} = \mathbf{C}_{B\to A}^T$, we can easily obtain the relationship of the bi-directional channels:

$$\mathbf{H}_{A\to B} = \mathbf{R}_B(\mathbf{R}_A^{-1}\mathbf{H}_{B\to A}\mathbf{T}_B^{-1})^T\mathbf{T}_A \quad = \underbrace{\mathbf{R}_B\mathbf{T}_B^{-T}}_{\mathbf{F}_B^{-T}}\mathbf{H}_{B\to A}^T\underbrace{\mathbf{R}_A^{-T}\mathbf{T}_A}_{\mathbf{F}_A} = \mathbf{F}_B^{-T}\mathbf{H}_{B\to A}^T\mathbf{F}_A. \tag{4.2}$$

where $\mathbf{F}_A = \mathbf{R}_A^{-T}\mathbf{T}_A$ and $\mathbf{F}_B = \mathbf{R}_B^{-T}\mathbf{T}_B$ of size $M_A \times M_A$ and $M_B \times M_B$, respectively, including both transmit and receive hardware properties are called calibration matrices. The diagonal elements of those calibration matrices represent the main M_A and M_B calibration coefficients whereas the off-diagonal elements come from the RF cross-talk and antenna mutual coupling.

Although transmitting and receiving antenna mutual coupling is not generally reciprocal [41], it is worth noting that theoretical modeling [19] and practical experience in [25, 28, 42] both indicate that, RF crosstalk and antenna mutual coupling can be ignored for reciprocity calibration, which implies that $\mathbf{T}_A$, $\mathbf{R}_A$, $\mathbf{T}_B$, $\mathbf{R}_B$ can all be considered as diagonal matrices. Thus $\mathbf{F}_A$ and $\mathbf{F}_B$ are also diagonal, (4.1) can be represented by $M_A \times M_B$ independent scalar equations. In this thesis, Section 8 is dedicated to the experiment validation on the fact that $\mathbf{F}_A$ and $\mathbf{F}_B$ are diagonal.

A TDD reciprocity based MIMO system normally has two phases for its function. Firstly, during the initialization of the system, the system activates the reciprocity calibration process in order to estimate $\mathbf{F}_A$ and $\mathbf{F}_B$. Then during the data transmission phase, these calibration coefficients are used together with the instantaneous measured UL channel $\hat{\mathbf{H}}_{B\to A}$ to estimate the CSIT $\mathbf{H}_{A\to B}$, based on which advanced beamforming algorithms can

[1]In this thesis, the term "antenna mutual coupling" is used to describe parasitic effects that two nearby antennas have on each other, when they are either both transmitting or receiving [19, 39]. However, this is different to the channel between transmitting and receiving elements of the same array, which we call the intra-array channel. This definition is slightly different to the one used in [28] and [40] where the term mutual coupling is also used for the intra-array channel.

be performed. Since the calibration coefficients stay stable during quite a long time [25], the calibration process doesn't have to be done very frequently.

"Over-the-air" calibration consists in estimating calibration matrices using channel measurements. According to whether UE is involved in the calibration process, we can classify various different existing methods into UE involved calibration and BS internal calibration.

4.2 UE involved calibration

Traditional methods for calibrating TDD systems, such as [21–23, 43], are to enable bi-directional transmission of pilots between the BS and the UE. Both sides estimate the channels from the other side and then the BS can calculate the calibration coefficients based on its own estimated UL channel the DL channel information fed back by the UE. These methods can be mainly classified into totally least squares (TLS) estimation and dot division estimation. TLS estimation consists in accumulating multiple bi-directional transmissions between node A and B and estimating the calibration coefficients by minimizing the sum of errors in all the accumulated channel estimations in both directions. This method allows for estimating the full calibration matrix including antenna mutual coupling and RF crosstalk, thus enabled our experiment validation on the diagonal assumption of calibration matrix in Chapter 8, where we give a detailed description. In this section, we give a brief presentation on the dot division method, as the principles appear to be simpler.

As the calibration matrix is diagonal, a MIMO system can be viewed as $M_A M_B$ SISO systems. Let us consider $\mathbf{F}_A = \text{diag}\{f_{A,1}, \ldots, f_{A,M_A}\}$ and $\mathbf{F}_B = \text{diag}\{f_{B,1}, \ldots, f_{B,M_B}\}$, where diag$\{\cdot\}$ represents a diagonal matrix composed of given elements. (4.2) can be written as

$$h_{A_k \to B_l} = f_{B,l}^{-1} h_{B_l \to A_k} f_{A,k} \tag{4.3}$$

where $\hat{h}_{A_k \to B_l}$ and $\hat{h}_{B_l \to A_k}$ are the bi-directional channel estimations between the k^{th} antenna at the BS and the l^{th} antenna element at the UE.

The most intuitive and easy way to estimate $\frac{f_{A,k}}{f_{B,l}}$ is the element wise division between the corresponding channel estimates, such as

$$\frac{\hat{f}_{A,k}}{\hat{f}_{B,l}} = \frac{\hat{h}_{A_k \to B_l}}{\hat{h}_{B_l \to A_k}}, \quad k = 1, \ldots, M_A, \quad l = 1, \ldots, M_B, \tag{4.4}$$

Assuming $\hat{f}_{B,1} = 1$, it is possible to calculate all $\hat{f}_{A,k}$ w.r.t $\hat{f}_{B,1}$, then it is also possible to obtain all $\hat{f}_{B,l}$ for $l \neq 1$. Since all obtained coefficients are relative values w.r.t. a reference, i.e., $\hat{f}_{B,1}$, such a method is also called "relative calibration". Note that although when we apply the estimated calibration coefficients to 4.2, the obtained $\mathbf{H}_{A \to B}$ has an ambiguity up to a scalar value, when we use this CSIT to calculate the beamforming weights, e.g. MRT precoding, the scalar ambiguity will not change the final generated beam pattern.

There are three weakness in UE involved reciprocity calibration. First, as channel measurements happen between BS and UE, it is difficult to ensure the quality of channel measurement. Second, the UE has to feed back its measured DL channel to BS where there would be unavoidable quantization error. When the number of antenna at the BS grows, the UL feedback will become very heavy. Third, BS and UE has an unknown

frequency offset, which makes the calibration more difficult, although some solutions were proposed in [44].

4.3 BS internal calibration

With the advent of massive MIMO, the weakness of UE involved calibration used in classic MIMO scheme become more severe, especially, the UE has to feed back a great amount of DL CSI for all BS antennas, making the calibration a heavy process. Meanwhile, [45, 46] point out that in a practical TDD system, it is mainly the calibration at the BS side, which restore the hardware asymmetry and help achieve the multi-user MIMO performance, whereas the benefit brought by the calibration on the UE side is not necessarily justified. In additional, it was observed in [43] that the calibration factor at the BS side is the same for all channels from the BS to any UE. To overcome the drawbacks of UE involved schemes while taking into account minor impact of the benefit of UE side calibration, BS internal calibration methods were proposed in [25–29]. In [25], the authors reported the first massive MIMO prototype Argos, where they enable their system with a BS internal calibration by replacing the UE with a reference antenna. By performing bi-directional transmission between the reference antenna and the rest of the antenna array, it is possible to estimate the calibration coefficients up to a common scalar ambiguity. The Argos system takes a big step in moving massive MIMO from a theoretical concept to practical systems, however, its calibration algorithm is sensitive to the location of the reference antenna, and as one of the consequences, is not suitable for distributed massive MIMO. This concern motivated the introduction by Rogalin *et al.* in [26] of a method whereby calibration is not performed w.r.t. a reference antenna but within different antenna pairs. Removing the reference antenna makes the calibration scheme a good method for antenna array having a distributed topology. It is worth noting that it can also be applied to collocated massive MIMO, e.g., LuMaMi massive MIMO prototype [47] from Lund University is enabled by a weighted version of the estimator as presented in [27] whereas a maximum likelihood (ML) estimator is presented in [28]. Moreover, a fast calibration method named Avalanche was proposed in [29]. The principle is to use a calibrated sub-array to calibrate uncalibrated elements. The calibrated array thus grows during the calibration process in a way similar to the phenomenon of avalanche. In [48], the author provides an idea to perform system health monitoring on the calibrated reciprocity. Under the assumption that the majority of calibration coefficients stay calibrated and only a minority of them change, the authors propose a compressed sensing enabled detection algorithm to find out which calibration coefficient has changed based on the sparsity in the vector representing the coefficient change. In [40], a calibration method dedicated to MRT is proposed.

The detail of these existing BS internal will later be exposed in Section 5.2 using our proposed unified calibration framework.

Chapter 5

A General Framework of Over-the-air Calibration in TDD Massive MIMO Systems

In this chapter, we provide a unified framework for "over-the-air" reciprocity calibration in a TDD system using antenna partition[1]. We reveal that although existing calibration methods appear at first sight to be different approaches, they can be modeled under a general pilot based calibration framework; different ways to partition the array into transmit and receive elements during successive training phases yield different schemes. The unified representation shows the relationship between these methods, provides alternative ways to obtain corresponding estimators. As this framework gives a general and high level understanding of the TDD calibration problem in massive MIMO systems, it opens up possibilities of new calibration methods. As an example, we present a novel family of calibration scheme based on antenna grouping, which can greatly speed up the calibration process. We will show that our proposed method greatly outperforms the Avalanche method [29] in terms of calibration accuracy, yet is equally fast. In order to evaluate the performance of calibration schemes, we derive Cramér-Rao bounds (CRB) on the accuracy of calibration coefficients estimation. Another important contribution of this work is the introduction of non-coherent accumulation. We will see that calibration does not necessarily have to be performed in an intensive manner during a single channel coherence interval, but can rather be executed using time resources distributed over a relatively long period. This enables TDD reciprocity calibration to be interleaved with the normal data transmission or reception, leaving it almost invisible for the whole system.

5.1 General calibration framework

Let us consider an antenna array of M elements partitioned into G groups denoted by $A_1, A_2, \ldots, A_G$, as in Fig. 5.1. We do not distinguish whether the antenna is on BS or UE side, but consider them equivalently. Group A_i contains M_i antennas such that

[1]This chapter combines a joint effort from colleagues in Eurecom and Huawei Technology (Paris). The initial proposal on the general framework from the thesis author was later greatly improved by contributions from Dirk Slock, Kalyana Gopala, Maxime Guillaud and Alexis Decunringe, especially on Sections 5.4, 5.5, 5.6.

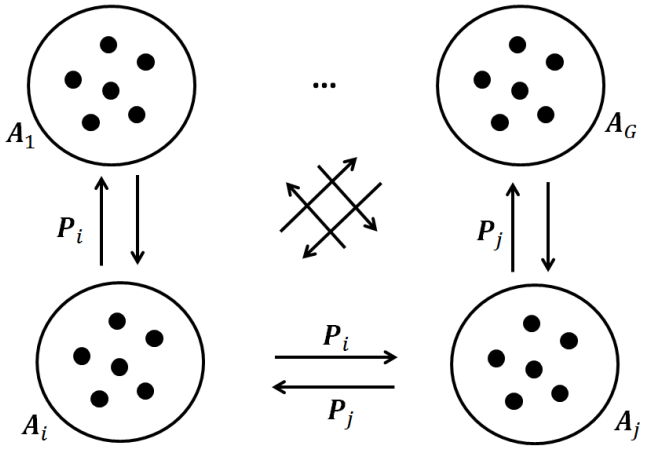


Figure 5.1: Bi-directional transmission between antenna groups.

$\sum_{i=1}^{G} M_i = M$. 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

$$\begin{cases} \mathbf{Y}_{i\to j} = \mathbf{R}_j \mathbf{C}_{i\to j} \mathbf{T}_i \mathbf{P}_i + \mathbf{N}_{i\to j} \\ \mathbf{Y}_{j\to i} = \mathbf{R}_i \mathbf{C}_{j\to i} \mathbf{T}_j \mathbf{P}_j + \mathbf{N}_{j\to i} \end{cases} \tag{5.1}$$

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 induces that $\mathbf{C}_{i\to j} = \mathbf{C}_{j\to i}^T$, thus for two different groups $1 \le i \ne j \le G$ in (5.1), by eliminating $\mathbf{C}_{i\to j}$ we have,

$$\mathbf{P}_i^T \mathbf{F}_i^T \mathbf{Y}_{j\to i} - \mathbf{Y}_{i\to j}^T \mathbf{F}_j \mathbf{P}_j = \widetilde{\mathbf{N}}_{ij}, \tag{5.2}$$

where the noise component $\widetilde{\mathbf{N}}_{ij} = \mathbf{P}_i^T \mathbf{F}_i^T \mathbf{N}_{j\to i} - \mathbf{N}_{i\to j}^T \mathbf{F}_j \mathbf{P}_j$, while $\mathbf{F}_i = \mathbf{R}_i^{-T} \mathbf{T}_i$ and $\mathbf{F}_j = \mathbf{R}_j^{-T} \mathbf{T}_j$ are the calibration matrices for groups i and j. The calibration matrix $\mathbf{F}$ is diagonal, and thus takes the form of

$$\mathbf{F} = \mathrm{diag}\{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_G\}. \tag{5.3}$$

Note that estimating $\mathbf{F}_i$ or $\mathbf{F}_j$ from (5.2) for a given pair (i, j) does not exploit all relevant received data. An optimal estimation considering jointly all received signals for all (i, j) will be proposed in Section 5.4. Note also that the proposed framework also allows to consider using only subsets of the received data which corresponds to some of the methods found in the literature.

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 (5.2) into

$$(\mathbf{Y}_{j\to i}^T * \mathbf{P}_i^T)\mathbf{f}_i - (\mathbf{P}_j^T * \mathbf{Y}_{i\to j}^T)\mathbf{f}_j = \widetilde{\mathbf{n}}_{ij}, \tag{5.4}$$

where $*$ denotes the Khatri–Rao product (or column-wise Kronecker product[2]), where we have used the equality $\text{vec}(\mathbf{A}\,\text{diag}(\mathbf{x})\,\mathbf{B}) = (\mathbf{B}^T * \mathbf{A})\,\mathbf{x}$. Note that, if we do not suppose that every $\mathbf{F}_i$ is diagonal, (5.4) holds more generally by replacing the Katri–Rao products by Kronecker products and $\mathbf{f}_i$ by $\text{vec}(\mathbf{F}_i)$. Finally, stacking equations (5.4) for all $1 \leq i < j \leq G$ yields

$$\boldsymbol{\mathcal{Y}}(\mathbf{P})\mathbf{f} = \widetilde{\mathbf{n}} \tag{5.5}$$

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

$$\boldsymbol{\mathcal{Y}}(\mathbf{P}) = \underbrace{\begin{bmatrix} (\mathbf{Y}_{2\to1}^T * \mathbf{P}_1^T) & -(\mathbf{P}_2^T * \mathbf{Y}_{1\to2}^T) & 0 & \dots \\ (\mathbf{Y}_{3\to1}^T * \mathbf{P}_1^T) & 0 & -(\mathbf{P}_3^T * \mathbf{Y}_{1\to3}^T) & \dots \\ 0 & (\mathbf{Y}_{3\to2}^T * \mathbf{P}_2^T) & -(\mathbf{P}_3^T * \mathbf{Y}_{2\to3}^T) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{(\sum_{j=2}^{G}\sum_{i=1}^{j-1} L_iL_j)\times M}. \tag{5.6}$$

5.1.1 Parameter identifiability and pilot design

Before proposing an estimator for $\mathbf{f}$, we raise the question of the problem identifiability which corresponds to the fact that (5.5) admits a unique solution in the noiseless scenario

$$\boldsymbol{\mathcal{Y}}(\mathbf{P})\mathbf{f} = \mathbf{0}. \tag{5.7}$$

Obviously, $\mathbf{f} = \mathbf{0}$ is a solution of (5.7) which means that $\mathbf{f}$ is defined up to a complex scalar factor. This indeterminacy can be resolved by fixing one of the calibration parameters, say $f_1 = \mathbf{e}_1^H\mathbf{f} = [1\,0\,\cdots\,0]\mathbf{f} = 1$ or by a norm constraint, for example $\|\mathbf{f}\| = 1$.

Then, the identifiability is related to the dimension of the kernel of $\boldsymbol{\mathcal{Y}}(\mathbf{P})$ in the sense that the problem is fully determined if and only if the kernel of $\boldsymbol{\mathcal{Y}}(\mathbf{P})$ is of dimension 1. Since the true $\mathbf{f}$ is a solution to (5.7), we know that the rank of $\boldsymbol{\mathcal{Y}}(\mathbf{P})$ is at most M-1. We will assume furthermore in the following that the pilot design is such that the rows of $\boldsymbol{\mathcal{Y}}(\mathbf{P})$ are linearly independent as long as the number of rows is less $M-1$. Note that this condition depends on the internal channel realization $\mathbf{C}_{i\to j}$ and on the pilot matrices $\mathbf{P}_i$. However, sufficient conditions of identifiability expressed on these matrices are out of the scope of this work. Under rows independence, (5.5) may be read as the following sequence of events:

1. Group 1 broadcasts its pilots to all other groups using L_1 channel uses;
2. After group 2 transmits its pilots, we can formulate L_2L_1 equations of the form (5.4);
3. After group 3 transmits its pilots, we can formulate $L_3L_1 + L_3L_2$ equations;
4. After group j transmits its pilots, we can formulate $\sum_{i=1}^{j-1} L_jL_i$ equations.

[2]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, $\mathbf{A} * \mathbf{B} = \begin{bmatrix}\mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \dots & \mathbf{a}_M \otimes \mathbf{b}_M\end{bmatrix}$ [49].

This process continues until group G finishes its transmission, and the whole calibration process finishes. During this process of transmission by the G antenna groups, we can start forming equations as indicated, that can be solved recursively for subsets of unknown calibration parameters, or we can wait until all equations are formed to solve the problem jointly.

By independence of the rows, we can state that the problem is fully determined if and only if $\sum_{1\leq i<j\leq G} L_j L_i \geq M-1$.

5.1.2 LS calibration parameter estimation

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

$$\begin{aligned} \hat{\mathbf{f}} &= \arg\min \|\boldsymbol{\mathcal{Y}}(\mathbf{P})\,\mathbf{f}\|^2 \\ &= \arg\min \textstyle\sum_{i<j} \|(\mathbf{Y}_{j\to i}^T * \mathbf{P}_i^T)\mathbf{f}_i - (\mathbf{P}_j^T * \mathbf{Y}_{i\to j}^T)\mathbf{f}_j\|^2 \end{aligned} \tag{5.8}$$

where $\boldsymbol{\mathcal{Y}}(\mathbf{P})$ is defined in (5.6). If we assume $\mathbf{e}_1^H\mathbf{f}=1$, the minimization problem can be solved by deriving the corresponding Lagrangian w.r.t $\mathbf{f}$, where $\mathbf{f}$ and $\mathbf{f}^*$ are regarded as independent [50]. The solution of (5.8) is given by

$$\begin{aligned} \widehat{\mathbf{f}} &= \arg\min_{\mathbf{f}:\mathbf{e}_1^H\mathbf{f}=1} \|\boldsymbol{\mathcal{Y}}(\mathbf{P})\,\mathbf{f}\|^2 \\ &= \frac{1}{\mathbf{e}_1^H(\boldsymbol{\mathcal{Y}}(\mathbf{P})^H\boldsymbol{\mathcal{Y}}(\mathbf{P}))^{-1}\mathbf{e}_1}(\boldsymbol{\mathcal{Y}}(\mathbf{P})^H\boldsymbol{\mathcal{Y}}(\mathbf{P}))^{-1}\mathbf{e}_1. \end{aligned} \tag{5.9}$$

Assuming a unit norm constraint on the other hand yields

$$\arg\min_{\mathbf{f}:\|\mathbf{f}\|=1} \|\boldsymbol{\mathcal{Y}}(\mathbf{P})\,\mathbf{f}\|^2 \;=\; \widehat{\mathbf{f}} = V_{min}(\boldsymbol{\mathcal{Y}}(\mathbf{P})^H\boldsymbol{\mathcal{Y}}(\mathbf{P})) \tag{5.10}$$

where $V_{min}(\mathbf{X})$ denotes the eigenvector of matrix $\mathbf{X}$ corresponding to its eigenvalue with the smallest magnitude.

5.2 Existing calibration techniques

Different choices for the partitioning of the M antennas and the pilots matrices exposed in Section 5.1 lead to different calibration algorithms. We will now see how different estimators of the calibration matrix can be derived from (5.2). In order to ease the description, we assume that the channel is constant during the whole calibration process, this assumption will later be relaxed and discussed in Section 5.5. Note also that in this section, we mainly focus on BS-side internal calibration since it is more suitable for massive MIMO systems, although our framework described above is not limited to this scenario. For UE-aided calibration, it suffices to set 2 groups such as A_1 and A_2, representing the BS and the UE, respectively.

5.2.1 Argos

The calibration method from the Argos prototype [25] consists in performing bi-directional transmission between a carefully chosen reference antenna and the rest of the antenna array. This can be recast in our framework by considering $G=2$ sets of antennas, with

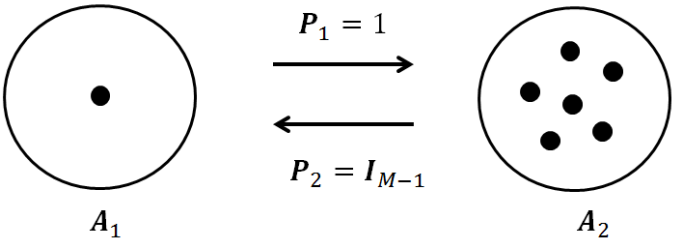


Figure 5.2: Argos calibration.

set A_1 containing only the reference antenna, i.e., $M_1 = 1$ and set A_2 containing all other antenna elements with $M_2 = M - 1$, as shown in Fig. 5.2. In slot 1, pilot 1 is broadcasted from the reference antenna to all antennas in set A_2, thus $L_1 = 1$, $\mathbf{P}_1 = 1$ and $\mathbf{f}_2 = \left[f_2, \ldots, f_M\right]^T$. From slot 2 to slot M, antennas in set A_2 successively transmit pilot 1 to the reference antenna, thus $L_2 = M - 1$ and $\mathbf{P}_2 = \mathbf{I}_{M-1}$. (5.4) thus becomes

$$f_1 \mathbf{y}_2^T = \text{diag}(\mathbf{y}_1^T)\mathbf{f}_2 + \tilde{\mathbf{n}}, \tag{5.11}$$

where $\mathbf{y}_1 = \begin{bmatrix} y_{1\to 2} & y_{1\to 3} & \cdots & y_{1\to M} \end{bmatrix}^T$ and $\mathbf{y}_2 = \begin{bmatrix} y_{2\to 1} & y_{3\to 1} & \cdots & y_{M\to 1} \end{bmatrix}^T$ with $y_{i\to j}$ representing the signal transmitted from antenna i and received at antenna j. (5.11) can be decomposed into $M - 1$ independent equations as

$$f_1 y_{i\to 1} = f_i y_{1\to i} + \tilde{n}_i, \tag{5.12}$$

where $\tilde{n}_i$ is the i^{th} element in the noise vector $\tilde{\mathbf{n}}$. The LS estimator for each element is thus given by

$$f_i = f_1 \frac{y_{i\to 1}}{y_{1\to i}}, \quad \text{where} \quad i = 2, 3, \ldots, M. \tag{5.13}$$

5.2.2 Methods based on successive single-antenna transmissions followed by joint estimation

The method from Rogalin et al. presented in [26,51] and further analyzed in [28] is based on single-antenna transmission at each channel access; all received signals are subsequently taken into account through joint estimation of the calibration parameters. In order to represent this method within the unified framework, we define each set A_i as containing antenna i, i.e., $M_i = 1$ for $1 \leq i \leq M$, as in Fig. 5.3.

Since we assume that the channel is constant, this calibration procedure can be performed in a way that antennas can broadcast pilot 1 in a round-robin manner to all other antennas. In total, M channel uses are needed to finish the transmission, making the pilots to be $\mathbf{P}_i = 1$ (with $L_i = 1$). With these pilot exchanges, and under the fact that the calibration matrix is diagonal, (5.2) degrades to

$$f_i y_{j\to i} = f_j y_{i\to j} + \tilde{n}. \tag{5.14}$$

Note that the calibration coefficient in [26] is defined as the inverse of that in the current chapter, in order to ease the comparison, we let $u_i = f_i^{-1}$ and rewrite (5.14) as

$$u_i^{-1} y_{j\to i} = u_j^{-1} y_{i\to j} + \tilde{n}. \tag{5.15}$$

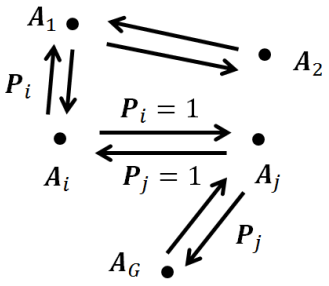


Figure 5.3: Calibration methods based on successive single-antenna transmissions follows by joint estimation. Not all links between elements are plotted.

Estimating $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \dots & u_M \end{bmatrix}^T$ may be performed by solving the minimization problem

$$\hat{\mathbf{u}} = \arg \min_{u_1, u_2, \dots, u_M} \sum_{i<j} |u_j y_{j\to i} - u_i y_{i\to j}|^2, \tag{5.16}$$

which is the same as defined in [51]. By fixing $u_1 = 1$, the solution of (5.16) is given by

$$\hat{\mathbf{u}} = -(\mathbf{A}_1^H \mathbf{A}_1)^{-1} \mathbf{A}_1^H \mathbf{a}_1 u_1, \tag{5.17}$$

where $\mathbf{A} = [\mathbf{a_1} | \mathbf{A}_1]$, with $\mathbf{a}_1$ representing the first column and $\mathbf{A}_1$ the remaining columns, and $\mathbf{A}$ is given by

$$A_{i,j} = \begin{cases} \sum_{k \neq i} |y_{k\to i}|^2 & \text{for } j = i, \\ -y_{i\to j}^* y_{j\to i} & \text{for } j \neq i. \end{cases} \tag{5.18}$$

(5.16) can also be solved by assuming a unit norm on $\mathbf{u}$, the solution is then given by $V_{min}(\mathbf{A})$, the eigenvector corresponding to the eigenvalue with the smallest magnitude [26].

Other methods following the same single antenna partition scenario can be viewed as variants of the above method. For example, by allowing only the transmission between two neighboring antennas (antenna index difference is 1), (5.15) becomes $f_i y_{i-1\to i} = f_{i-1} y_{i\to i-1} + \tilde{n}$. Thus, $f_i = \frac{y_{i\to i-1}}{y_{i-1\to i}} f_{i-1} + \tilde{n}$. By setting the first antenna as the reference antenna with $f_1 = 1$, we can obtain a daisy chain calibration method as in [20], although the original was presented as a hardware-based calibration.

Another variant considered in [26] and [27] consists in weighting the error metric in (5.16) such as $|\beta_{j\to i} u_j y_{j\to i} - \beta_{i\to j} u_i y_{i\to j}|^2$ where the weights $\beta_{j\to i}$ and $\beta_{i\to j}$ are based on the SNR of the intra-array channel between antenna element i and j.

5.2.3 Avalanche

Avalanche [29] is a family of fast recursive calibration methods. The algorithm successively uses already calibrated parts of the antenna array to calibrate uncalibrated antennas which, once calibrated, are merged into the calibrated array. A full Avalanche calibration may be expressed under the unified framework by considering $M = \frac{1}{2}G(G-1)+1$ antennas where G is the number of groups of antennas partitioning the set of antenna elements as follows: group A_1 contains antenna 1, group A_2 contains antenna 2, group A_3 contains antennas 3 and 4, etc. until group A_G that contains the last $G-1$ antennas. In other terms, group

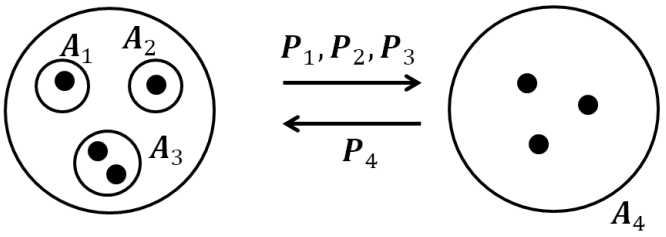


Figure 5.4: Example of full Avalanche calibration with 7 antennas partitioned into 4 groups. Group 1, 2, 3 have already been calibrated, and group 4 is to be calibrated.

A_i contains $M_i = \max(1, i-1)$ antennas. Moreover, in the method proposed in [29], each group A_i uses $L_i = 1$ channel use by sending the pilot $\mathbf{P}_i = \mathbf{1}_{M_i \times 1}$. An example with 7 antenna elements partitioned into 4 antenna groups, where we use group 1, 2, 3 (already been calibrated) to calibrate group 4, is shown in Fig. 5.4. In this case, (5.4) then becomes

$$(\mathbf{y}_{j\to i}^T * \mathbf{P}_i^T)\mathbf{f}_i - (\mathbf{P}_j^T * \mathbf{y}_{i\to j}^T)\mathbf{f}_j = \widetilde{\mathbf{n}}_{ij}. \tag{5.19}$$

In [29], the authors exploited an online version of the LS estimator using previously estimated calibration parameters $\hat{\mathbf{f}}_1, \ldots, \hat{\mathbf{f}}_{i-1}$ by minimizing

$$\begin{aligned} \hat{\mathbf{f}}_i &= \arg\min_{\mathbf{f}_i} \sum_{j=1}^{i-1} \left\| (\mathbf{y}_{j\to i}^T * \mathbf{P}_i^T)\mathbf{f}_i - (\mathbf{P}_j^T * \mathbf{y}_{i\to j}^T)\hat{\mathbf{f}}_j \right\|^2 \\ &= (\mathbf{Y}_i^H \mathbf{Y}_i)^{-1} \mathbf{Y}_i^H \mathbf{a}_i, \end{aligned} \tag{5.20}$$

where $\mathbf{Y}_i = \begin{bmatrix} \mathbf{y}_{1\to i} & \mathbf{y}_{2\to i} & \cdots & \mathbf{y}_{i-1\to i} \end{bmatrix}^T \in \mathbb{C}^{(i-1)\times M_i}$, and $\mathbf{a}_i = [(\mathbf{P}_1^T * \mathbf{y}_{i\to 1}^T)\hat{\mathbf{f}}_1, \ldots, (\mathbf{P}_{i-1}^T * \mathbf{y}_{i\to i-1}^T)\hat{\mathbf{f}}_{i-1}] \in \mathbb{C}^{(i-1)\times 1}$. Two things should be noted, firstly, $\mathbf{f}_1, \ldots, \mathbf{f}_{i-1}$ are replaced by their estimated version which cause error propagation (as will be seen in Section 5.6): estimation errors on a given calibration coefficient will later propagate to subsequently calibrated antenna elements. Secondly, in order for (5.20) to be well-defined, i.e. in order for $\mathbf{Y}_i^H \mathbf{Y}_i$ to be invertible, it is necessary that $M_i \leqslant i - 1$. Note that this necessary condition is specific to the considered online LS estimator (5.20) and is more restrictive than the identifiability condition exposed in Section 5.1.1.

5.3 Fast Calibration: optimal antenna grouping

The general calibration framework in Section 5.1 opens up possibilities for new calibration schemes, since new ways to group up antennas leads to new calibration schemes. In this section we show that considering group of antennas can potentially reduce the total number of channel accesses necessary for calibration; we derive the theoretical limit on the smallest number of groups needed to perform calibration.

We first address the problem on the smallest number of groups by finding the best choices for the L_i in order to see to what extent optimizing the group based calibration can speed up the calibration process. Let us consider the case where the total number of time slots available for calibration is fixed to K. We derive the number of pilot transmissions for each group, $L_1, \ldots, L_G$, that would maximize the total number of antennas that can be

calibrated, i.e.

$$\max_{(L_1,\ldots,L_G)} \left[\sum_{j=2}^{G} \sum_{i=1}^{j-1} L_j L_i + 1 \right], \text{ subject to } \sum_{i=1}^{G} L_i = K. \tag{5.21}$$

As shown in Appendix 5.A, the solution of this discrete optimization problem is attained when the number of pilot transmissions for each group is equal to 1, i.e. $L_i = 1$ for any i and $G = K$; note that the Avalanche approach is optimal in that sense. In this case, the number of antennas that can be calibrated is $\frac{1}{2}K(K-1)+1$. Thus, for a given array size M, the calibration time grows only on the order of $\sqrt{M}$, which is faster than $\mathcal{O}(M)$ in Argos and the method of Rogalin et al.[3] [26]. Remark also that it is not necessary for the groups to be of equal size.

5.4 Optimal estimation and performance limits

In order to derive estimation error bounds for the reciprocity parameters, we should not exclude a priori any data obtained during the training phase, which is what we shall assume here. In this section, we derive the CRB and associated ML estimation for the unified calibration scheme based on antenna partition. In order to obtain tractable results, we rely on a bilinear model to represent the calibration process. From (5.1), we have

$$\begin{aligned} \mathbf{Y}_{i\to j} &= \mathbf{R}_j \mathbf{C}_{i\to j} \mathbf{T}_i \mathbf{P}_i + \mathbf{N}_{i\to j} \\ &= \underbrace{\mathbf{R}_j \mathbf{C}_{i\to j} \mathbf{R}_i^T}_{\boldsymbol{\mathcal{H}}_{i\to j}} \mathbf{F}_i \mathbf{P}_i + \mathbf{N}_{i\to j}, \end{aligned} \tag{5.22}$$

where $\mathbf{F}_i = \mathbf{R}_i^{-T}\mathbf{T}_i$ is the calibration matrix for group i. We define $\boldsymbol{\mathcal{H}}_{i\to j} = \mathbf{R}_j \mathbf{C}_{i\to j} \mathbf{R}_i^T$ to be a auxiliary internal channel (it does not correspond to any physically measurable quantity) that appears as a nuisance parameter in the estimation of the calibration parameters. Note that the auxiliary channel $\boldsymbol{\mathcal{H}}_{i\to j}$ inherits the reciprocity from the propagation channel $\mathbf{C}_{i\to j}$: $\boldsymbol{\mathcal{H}}_{i\to j} = \boldsymbol{\mathcal{H}}_{j\to i}^T$. Upon applying the vectorization operator for each bidirectional transmission between groups i and j, we have, similarly to (5.5)

$$\operatorname{vec}(\mathbf{Y}_{i\to j}) = (\mathbf{P}_i^T * \boldsymbol{\mathcal{H}}_{i\to j})\, \mathbf{f}_i + \operatorname{vec}(\mathbf{N}_{i\to j}). \tag{5.23}$$

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

$$\operatorname{vec}(\mathbf{Y}_{j\to i}^T) = (\boldsymbol{\mathcal{H}}_{i\to j}^T * \mathbf{P}_j^T)\mathbf{f}_j + \operatorname{vec}(\mathbf{N}_{j\to i})^T. \tag{5.24}$$

Alternatively, (5.23) and (5.24) may also be written as

$$\begin{cases} \operatorname{vec}(\mathbf{Y}_{i\to j}) = \left[(\mathbf{F}_i\mathbf{P}_i)^T \otimes \mathbf{I}\right] \operatorname{vec}(\boldsymbol{\mathcal{H}}_{i\to j}) + \operatorname{vec}(\mathbf{N}_{i\to j}) \\ \operatorname{vec}(\mathbf{Y}_{j\to i}^T) = \left[\mathbf{I} \otimes (\mathbf{P}_j^T\mathbf{F}_j)\right] \operatorname{vec}(\boldsymbol{\mathcal{H}}_{i\to j}) + \operatorname{vec}(\mathbf{N}_{j\to i}). \end{cases} \tag{5.25}$$

[3]The slots needed by Rogalin et al. method in [26] is M if we perform round-robin broadcasting for each antenna under the assumption that the all channels between antennas are constant during the whole calibration process whereas it would be $\mathcal{O}(M^2)$ if we perform bi-directional transmission independently for each antenna pair with enough good internal channels. Please refer to Section 5.5 for more details.

Stacking these observations into a vector $\mathbf{y} = \left[\text{vec}(\mathbf{Y}_{1\to2})^T \text{vec}(\mathbf{Y}_{2\to1}^T)^T \text{vec}(\mathbf{Y}_{1\to3})^T \dots\right]^T$, the above two alternative formulations can be summarized into

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\mathcal{H}}(\mathbf{h}, \mathbf{P})\mathbf{f} + \mathbf{n} \\ &= \boldsymbol{\mathcal{F}}(\mathbf{f}, \mathbf{P})\mathbf{h} + \mathbf{n}, \end{aligned} \tag{5.26}$$

where $\mathbf{h} = \left[\text{vec}(\boldsymbol{\mathcal{H}}_{1\to2})^T \text{vec}(\boldsymbol{\mathcal{H}}_{1\to3})^T \text{vec}(\boldsymbol{\mathcal{H}}_{2\to3})^T \dots\right]^T$, $\mathbf{n}$ is the corresponding noise vector. The composite matrices $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{F}}$ are given by,

$$\boldsymbol{\mathcal{H}}(\mathbf{h}, \mathbf{P}) = \begin{bmatrix} \mathbf{P}_1^T * \boldsymbol{\mathcal{H}}_{1\to2} & 0 & 0 & \dots \\ 0 & \boldsymbol{\mathcal{H}}_{1\to2}^T * \mathbf{P}_2^T & 0 & \dots \\ \mathbf{P}_1^T * \boldsymbol{\mathcal{H}}_{1\to3} & 0 & 0 & \dots \\ 0 & 0 & \boldsymbol{\mathcal{H}}_{1\to3}^T * \mathbf{P}_3^T & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\boldsymbol{\mathcal{F}}(\mathbf{f}, \mathbf{P}) = \begin{bmatrix} \mathbf{P}_1^T\mathbf{F}_1 \otimes \mathbf{I} & 0 & 0 & 0 & \dots \\ \mathbf{I} \otimes \mathbf{P}_2^T\mathbf{F}_2 & 0 & 0 & 0 & \dots \\ 0 & \mathbf{P}_1^T\mathbf{F}_1 \otimes \mathbf{I} & 0 & 0 & \dots \\ 0 & \mathbf{I} \otimes \mathbf{P}_3^T\mathbf{F}_3 & 0 & 0 & \dots \\ 0 & 0 & \mathbf{P}_2^T\mathbf{F}_2 \otimes \mathbf{I} & 0 & \dots \\ 0 & 0 & \mathbf{I} \otimes \mathbf{P}_3^T\mathbf{F}_3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{5.27}$$

The scenario is now identical to that encountered in some blind channel estimation scenarios and hence we can take advantage of some existing tools [52], [53], which we summarize next.

5.4.1 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 (5.26) as

$$\mathbf{J} = \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\mathcal{H}} & \boldsymbol{\mathcal{F}} \end{bmatrix}^H \begin{bmatrix} \boldsymbol{\mathcal{H}} & \boldsymbol{\mathcal{F}} \end{bmatrix}. \tag{5.28}$$

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 as the received data (5.26) involves the product of the channel and the calibration factors, $\boldsymbol{\mathcal{H}}\mathbf{f} = \boldsymbol{\mathcal{F}}\mathbf{h}$. As a result the FIM has the following null space [54]

$$\mathbf{J} \begin{bmatrix} \mathbf{f} \\ -\mathbf{h} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\mathcal{H}} & \boldsymbol{\mathcal{F}} \end{bmatrix}^H (\boldsymbol{\mathcal{H}}\mathbf{f} - \boldsymbol{\mathcal{F}}\mathbf{h}) = \mathbf{0}. \tag{5.29}$$

To determine the CRB when the FIM is singular, constraints have to be added to regularize the estimation problem. Note that as the calibration parameters are complex, one complex constraint corresponds to two real constraints. It is shown in [54], [55] that a choice of constraints such that their linearized version fills up the null space of the FIM result in the lowest CRB, while not adding information in subspaces where the data provides information. One such choice of constraints, that we will refer to as the "norm and phase"

constraint, is the combination of the following two real constraints (where $\mathbf{f}$ is the true set of calibration parameters and $\hat{\mathbf{f}}$ is an estimate thereof): 1) norm constraint: $\|\hat{\mathbf{f}}\|^2 = \|\mathbf{f}\|^2$, 2) phase constraint: $\begin{bmatrix}-\text{Im}\{\mathbf{f}^T\} & \text{Re}\{\mathbf{f}^T\}\end{bmatrix} \begin{bmatrix}\text{Re}\{\hat{\mathbf{f}}\} \\ \text{Im}\{\hat{\mathbf{f}}\}\end{bmatrix} = 0$.

With these constraints, the (constrained) CRB corresponds to the pseudo inverse of the FIM. The CRB for the calibration parameters (after the nuisance parameters $\mathbf{h}$ have been eliminated) is then given by

$$\text{CRB}_{\mathbf{f}} = \sigma^2 \left(\boldsymbol{\mathcal{H}}^H \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{F}}}^{\perp} \boldsymbol{\mathcal{H}}\right)^{\dagger}, \tag{5.30}$$

where $\dagger$ corresponds to the Moore-Penrose pseudo inverse, $\boldsymbol{\mathcal{P}}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^H\mathbf{X})^{\dagger}\mathbf{X}^H$ and $\boldsymbol{\mathcal{P}}_{\mathbf{X}}^{\perp} = \mathbf{I} - \boldsymbol{\mathcal{P}}_{\mathbf{X}}$ are the projection operators on resp. the column space of matrix $\mathbf{X}$ and its orthogonal complement. Note that in some group calibration scenarios, $\boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{F}}$ can be singular (i.e, $\mathbf{h}$ could be not identifiable even if $\mathbf{f}$ is identifiable or known).

If the regularization constraint $\mathbf{e}_1^H\mathbf{f} = 1$ (see Section 5.1.2) is used instead, the corresponding CRB can be obtained as follows: let $\boldsymbol{\theta}$ to denote the overall vector of unknown parameters, i.e., $\boldsymbol{\theta} = \begin{bmatrix}\mathbf{f}^T & \mathbf{h}^T\end{bmatrix}^T$. Then, according to Theorem 1 in [55], the corresponding constrained CRB is

$$\text{CRB}_{\boldsymbol{\theta}} = \boldsymbol{\nu}_{\boldsymbol{\theta}}(\boldsymbol{\nu}_{\boldsymbol{\theta}}^T \mathbf{J}\, \boldsymbol{\nu}_{\boldsymbol{\theta}})^{\dagger}\boldsymbol{\nu}_{\boldsymbol{\theta}}^T. \tag{5.31}$$

where $\boldsymbol{\nu}_{\boldsymbol{\theta}}$ corresponds to an identity matrix without the first column (and hence its column space is the orthogonal complement of that of $\mathbf{e}_1$). The $\text{CRB}_{\mathbf{f}}$ would be the first $M \times M$ sub-block of $\text{CRB}_{\boldsymbol{\theta}}$. We will refer to this constraint as the first coefficient constraint.

Note that [28] also addresses the CRB for a scenario where transmission happens one antenna at a time. It considers transmit and receive side calibration parameters separately. Those become identifiable because a model is introduced for the internal propagation channel, with the mean taken as the line of sight (LoS) component (distance induced delay and attenuation) and complex Gaussian non-LoS (NLOS) components contributing to the variance. Using this model, the authors eliminate the nuisance parameters (channel) by replacing it with the known mean and the NLOS components added to the noise. The covariance of this channel is a scaled identity matrix with the scale factor roughly 60dB below the mean channel power. This approach would result in under estimation of the CRB, as mentioned in [28, Sec. III-E-2]. We show this behavior of the Bayesian CRB of [28] via simulations in Section 5.6.

5.4.2 Maximum Likelihood estimation

We now turn our attention to the design of an optimal estimator; from (5.26) we get the negative log-likelihood up to an additive constant, as

$$\frac{1}{\sigma^2}\left\|\mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{h}, \mathbf{P})\mathbf{f}\right\|^2 = \frac{1}{\sigma^2}\left\|\mathbf{y} - \boldsymbol{\mathcal{F}}(\mathbf{f}, \mathbf{P})\mathbf{h}\right\|^2 . \tag{5.32}$$

The maximum likelihood estimator of $(\mathbf{h}, \mathbf{f})$, obtained by minimizing (5.32), can be computed using alternating optimization on $\mathbf{h}$ and $\mathbf{f}$, which leads to a sequence of quadratic problems. As a result, for given $\mathbf{f}$, we find $\widehat{\mathbf{h}} = (\boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{F}})^{-1}\boldsymbol{\mathcal{F}}^H\mathbf{y}$ and for given $\mathbf{h}$, we find $\widehat{\mathbf{f}} = (\boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}})^{-1}\boldsymbol{\mathcal{H}}^H\mathbf{y}$. This leads to the Alternating Maximum Likelihood (AML) algorithm

(Algorithm 1) [52, 53] which iteratively maximizes the likelihood by alternating between the desired parameters $\mathbf{f}$ and the nuisance parameters $\mathbf{h}$ for the formulation (5.26)[4].

Algorithm 1 Alternating Maximum Likelihood (AML)

1: **Initialization:** Initialize $\widehat{\mathbf{f}}$ using existing calibration methods (e.g., the method in 5.2.2) or as a vector of all 1's.
2: **repeat**
3: Construct $\boldsymbol{\mathcal{F}}$ as in (5.27) using $\widehat{\mathbf{f}}$.
$\widehat{\mathbf{h}} = (\boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{F}})^{-1}\boldsymbol{\mathcal{F}}^H\tilde{\mathbf{y}}$
4: Construct $\boldsymbol{\mathcal{H}}$ as in (5.27) using $\widehat{\mathbf{h}}$.
$\widehat{\mathbf{f}} = (\boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}})^{-1}\boldsymbol{\mathcal{H}}^H\tilde{\mathbf{y}}$
5: **until** the difference on the calculated $\hat{\mathbf{f}}$ between two iterations is small enough.

5.4.3 Maximum Likelihood vs. Least-Squares

At first, it would seem that the ML and CRB formulations above are unrelated to the LS method introduced in Section 5.1 and used in most existing works. However, consider again the received signal in a pair (i, j) as in (5.25). Eliminating the common auxiliary channel $\boldsymbol{\mathcal{H}}_{i\to j}$, we get the elementary equation (5.2) for the LS method (5.9) or (5.10). From (5.25), we can rewrite (5.2) as

$$\begin{aligned}\left[\mathbf{I}\otimes(\mathbf{P}_j^T\mathbf{F}_j) \;\; -(\mathbf{F}_i\mathbf{P}_i)^T\otimes\mathbf{I}\right]\begin{bmatrix}\mathrm{vec}(\mathbf{Y}_{i\to j})\\ \mathrm{vec}(\mathbf{Y}_{j\to i}^T)\end{bmatrix}\\ =\left[\mathbf{I}\otimes(\mathbf{P}_j^T\mathbf{F}_j) \;\; -(\mathbf{F}_i\mathbf{P}_i)^T\otimes\mathbf{I}\right]\begin{bmatrix}\mathrm{vec}(\mathbf{N}_{i\to j})\\ \mathrm{vec}(\mathbf{N}_{j\to i}^T)\end{bmatrix}\end{aligned} \tag{5.33}$$

Thus, equivalently to (5.5), one obtains

$$\boldsymbol{\mathcal{Y}}(\mathbf{P})\mathbf{f} = \boldsymbol{\mathcal{F}}^{\perp H}\mathbf{y} = \widetilde{\mathbf{n}} \tag{5.34}$$

where

$$\boldsymbol{\mathcal{F}}^{\perp} = \begin{bmatrix}
\mathbf{I}\otimes(\mathbf{F}_2\mathbf{P}_2)^* & 0 & 0 & 0 & \dots\\
-(\mathbf{F}_1\mathbf{P}_1)^*\otimes\mathbf{I} & 0 & 0 & 0 & \dots\\
0 & \mathbf{I}\otimes(\mathbf{F}_3\mathbf{P}_3)^* & 0 & 0 & \dots\\
0 & -(\mathbf{F}_1\mathbf{P}_1)^*\otimes\mathbf{I} & 0 & 0 & \dots\\
0 & 0 & \mathbf{I}\otimes(\mathbf{F}_3\mathbf{P}_3)^* & 0 & \dots\\
0 & 0 & -(\mathbf{F}_2\mathbf{P}_2)^*\otimes\mathbf{I} & 0 & \dots\\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix} \tag{5.35}$$

such that the column space of $\boldsymbol{\mathcal{F}}^{\perp}$ corresponds to the orthogonal complement of the column space of $\boldsymbol{\mathcal{F}}$ (see Appendix 5.B) assuming that either $M_i \geq L_i$ or $L_i \geq M_i$ for all $1 \leq i \leq G$. Now, the ML criterion in (5.32) is separable in $\mathbf{f}$ and $\mathbf{h}$. Optimizing (5.32) w.r.t. $\mathbf{h}$ leads to $\mathbf{h} = (\boldsymbol{\mathcal{F}}^H\boldsymbol{\mathcal{F}})^{\dagger}\boldsymbol{\mathcal{F}}^H\mathbf{y}$ as mentioned earlier. Substituting this estimate for $\mathbf{h}$ into (5.32) yields a ML estimator $\hat{\mathbf{f}}$ minimizing

$$\mathbf{y}^H\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{F}}}^{\perp}\mathbf{y} = \mathbf{y}^H\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{F}}^{\perp}}\mathbf{y} = \mathbf{y}^H\boldsymbol{\mathcal{F}}^{\perp}(\boldsymbol{\mathcal{F}}^{\perp H}\boldsymbol{\mathcal{F}}^{\perp})^{\dagger}\boldsymbol{\mathcal{F}}^{\perp H}\mathbf{y} \tag{5.36}$$

[4]The method used in [28] to derive the ML estimator, although called "Expectation Maximization" in the original paper, actually corresponds to the AML scheme, but using 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.

where we used $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{F}}}^{\perp} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{F}}^{\perp}}$. This should be compared to the least-squares method which consists in minimizing $\|\boldsymbol{\mathcal{F}}^{\perp H}\mathbf{y}\|^2 = \|\boldsymbol{\mathcal{Y}}\mathbf{f}\|^2$ in (5.9) or (5.10). Hence (5.36) can be interpreted as an optimally weighted least-squares method since from (5.26) $\boldsymbol{\mathcal{F}}^{\perp H}\mathbf{y} = \boldsymbol{\mathcal{F}}^{\perp H}\mathbf{n} = \widetilde{\mathbf{n}}$ leads to colored noise with covariance matrix $\sigma^2\boldsymbol{\mathcal{F}}^{\perp H}\boldsymbol{\mathcal{F}}^{\perp}$. The compressed log-likelihood in (5.36) can now be optimized using a variety of iterative techniques such as Iterative Quadratic ML (IQML), Denoised IQML (DIQML) or Pseudo-Quadratic ML (PQML) [53], and initialized with the least-squares method. It is not clear though whether accounting for the optimal weighting in ML would lead to significant gains in performance. The weighting matrix (before inversion) $\boldsymbol{\mathcal{F}}^{\perp H}\boldsymbol{\mathcal{F}}^{\perp}$ is block diagonal with a square block corresponding to the pair of antenna groups (i, j) being of dimension L_iL_j. If, e.g., all $L_i = 1$, then $\boldsymbol{\mathcal{F}}^{\perp H}\boldsymbol{\mathcal{F}}^{\perp}$ is a diagonal matrix. If furthermore all $M_i = 1$ (groups of isolated antennas), all pilots are of equal magnitude, and if all calibration factors would be of equal magnitude, then $\boldsymbol{\mathcal{F}}^{\perp H}\boldsymbol{\mathcal{F}}^{\perp}$ would be just a multipe of identity and hence would not represent any weighting. We shall leave this topic for further exploration. In any case, the fact that the CRB derived above and the ML and least-squares methods are all based on the signal model (5.26) shows that in contrast to [28], the CRB above is the appropriate CRB for the estimation methods discussed here.

5.4.4 Calibration bias at low SNR

Whereas the CRB applies to unbiased estimators, at low SNR the estimators are biased which turns out to lead to MSE saturation. In the case of a norm constraint, $\|\widehat{\mathbf{f}}\|^2 = \|\mathbf{f}\|^2$ then due to the triangle inequality

$$\|\widehat{\mathbf{f}} - \mathbf{f}\| \leq \|\widehat{\mathbf{f}}\| + \|\mathbf{f}\| = 2\|\mathbf{f}\| \tag{5.37}$$

and hence MSE $= \mathrm{E}\,[\|\widehat{\mathbf{f}} - \mathbf{f}\|^2] \leq 4\|\mathbf{f}\|^2$. However, MSE saturation occurs also in the case of a linear constraint. We shall provide here only some brief arguments. For a linear constraint of the form $\mathbf{e}^H\widehat{\mathbf{f}} = c$, the least-squares method leads to

$$\widehat{\mathbf{f}} = \frac{c}{\mathbf{e}^H(\boldsymbol{\mathcal{Y}}^H\boldsymbol{\mathcal{Y}})^{-1}\mathbf{e}}(\boldsymbol{\mathcal{Y}}^H\boldsymbol{\mathcal{Y}})^{-1}\mathbf{e}\,. \tag{5.38}$$

As the SNR decreases, the noise part $\boldsymbol{\mathcal{N}}$ of $\boldsymbol{\mathcal{Y}}$ will eventually dominate $\boldsymbol{\mathcal{Y}}$. Hence $\widehat{\mathbf{f}} = \frac{c}{\mathbf{e}^H(\boldsymbol{\mathcal{N}}^H\boldsymbol{\mathcal{N}})^{-1}\mathbf{e}}(\boldsymbol{\mathcal{N}}^H\boldsymbol{\mathcal{N}})^{-1}\mathbf{e}$ in which the coefficients will tend to be decaying (similar to the linear phase property of linear prediction filters). To take a short-cut, consider replacing $\boldsymbol{\mathcal{N}}^H\boldsymbol{\mathcal{N}}$ by its mean $\mathrm{E}\,[\boldsymbol{\mathcal{N}}^H\boldsymbol{\mathcal{N}}] = c'\,\mathbf{I}$. Then we get $\widehat{\mathbf{f}} = \frac{c}{\mathbf{e}^H\mathbf{e}}\mathbf{e}$ which is clearly bounded. Hence $\widehat{\mathbf{f}}$ will be strongly biased with bounded MSE.

5.5 Non-coherent accumulation

5.5.1 Overview

We have assumed in Sections 5.1 and 5.2 that the channel is constant during the whole calibration process, which may become questionable if the number of antennas becomes very large since more time is then needed to accomplish the whole calibration process. As a consequence, it is possible that we cannot accumulate enough observations in the channel coherence time and bandwidth. Considering non-coherent accumulation is thus essential

in TDD reciprocity calibration for massive MIMO calibration. Calibration methods presented in Section 5.2, however, rely on the assumption that the channel stays unchanged since they assume that reciprocity holds during the calibration process. Note that it is possible to adapt the method of Rogalin et al. to the non-coherent context. Indeed, if the channel is stable during the whole calibration process, one can use the method detailed in Section 5.2.2, broadcasting pilots from each antenna in a round-robin manner when all other antennas are listening, thus M slots are needed to accomplish the whole process. If the the number of coherent slots is not large enough, a non-coherent way to accumulate observations can be performing bi-directional transmissions for each antenna pair independently, using therefore $M(M-1)$ slots. Some papers also implicitely use non-coherent accumulations; see for example [56] who derives a TLS estimator from such measurements.

Let us extend the signal model in Section 5.1 by allowing to accumulate measurements over several time slots beyond the channel coherence time. We assume that these are indexed by $1 \leq t \leq T$, so that T represents the number of non-coherent slots at disposal. Clearly, the "over the air" reciprocity equation $\mathbf{C}_{i\to j} = \mathbf{C}_{j\to i}^T$ holds only for measurements obtained during the same time slot. However, measurements related to several groups of antennas obtained during multiple non-coherent time slots can be successfully combined to perform joint calibration of the complete array, as shown next. Let us assume that, during a given non-coherent slot t, a subset $\mathcal{G}(t)$ of the groups forming the partition of the array transmit training signals; we require that $\mathcal{G}(t)$ has at least two elements. When group A_i, $i \in \mathcal{G}(t)$ is transmitting, the received signal at group A_j, $j \in \mathcal{G}(t)$, $j \neq i$ is written as $\mathbf{Y}_{j\to i,t} = \mathbf{R}_j \mathbf{C}_{i\to j,t} \mathbf{T}_i \mathbf{P}_{i,t} + \mathbf{N}_{j,t}$, and $\mathbf{Y}_{i\to j,t}$ is defined similarly. (5.4) then becomes

$$(\mathbf{Y}_{j\to i,t}^T * \mathbf{P}_{i,t}^T)\mathbf{f}_i - (\mathbf{P}_{j,t}^T * \mathbf{Y}_{i\to j,t}^T)\mathbf{f}_j = \widetilde{\mathbf{n}}_{ij,t}. \tag{5.39}$$

Stacking these equations similarly to (5.5), but with respect to the $i, j \in \mathcal{G}(t)$, gives $\boldsymbol{\mathcal{Y}}_t(\mathbf{P}_t)\mathbf{f} = \widetilde{\mathbf{n}}_t$ for each time slot t.

5.5.2 LS Estimation

The LS estimator of the calibration matrix is thus, taking into account all accumulated observations,

$$\begin{aligned} \hat{\mathbf{f}} &= \arg\min_{\mathbf{f}} \sum_{t=1}^{T} \sum_{\substack{i,j\in\mathcal{G}(t)\\ i\neq j}} \left\| (\mathbf{Y}_{j\to i,t}^T * \mathbf{P}_{i,t}^T)\mathbf{f}_i - (\mathbf{P}_{j,t}^T * \mathbf{Y}_{i\to j,t}^T)\mathbf{f}_j \right\|^2 \\ &= \arg\min_{\mathbf{f}} \|\boldsymbol{\mathcal{Y}}(\mathbf{P})\mathbf{f}\|^2 \end{aligned} \tag{5.40}$$

where the minimum is taken either under the constraint $f_1 = 1$ or $\|\mathbf{f}\| = 1$ and $\boldsymbol{\mathcal{Y}}(\mathbf{P}) = [\boldsymbol{\mathcal{Y}}_1(\mathbf{P}_1)^T, \ldots, \boldsymbol{\mathcal{Y}}_T(\mathbf{P}_T)^T]^T$. Therefore, the approach of (5.40) are very similar to (5.9) and (5.10). This shows that calibration using a joint estimator based on non-coherent measurements can be readily implemented by making sure that the measurements $\mathbf{Y}_{j\to i,t}$ and $\mathbf{Y}_{i\to j,t}$ appearing in each term of the sum above have been obtained during the same coherence interval. Note also that this approach allows to collect multiple measurements between the same pair (i, j) of antenna groups, hence providing a way to increase the accuracy (by averaging over multiple noise realizations) and robustness (by minimizing the effect of a single catastrophic realization of the internal channel which could yield a rank-deficient set of linear equations for a given t) of the estimator.

5.5.3 Optimal grouping

Statement similar to Section 5.3 can be made for non-coherent group-based fast calibration. The maximization proposed in Section 5.3 is still valid in this context leading to an optimal number of groups equal to the number of coherent slots $G = K$. Therefore, since $\frac{1}{2}K(K-1)$ independent rows in $\boldsymbol{\mathcal{Y}}(\mathbf{P})$ are accumulated per non-coherent slot, if we fix the number of antennas to be equal to M, the number of non-coherent slots T should satisfy $\frac{T}{2}K(K-1) \geq M-1$ in order to calibrate all antenna elements. Note that the total number of calibrated antennas, equal to $\frac{T}{2}K(K-1)+1$ is linear in T and quadratic in K, which confirms that is more valuable to perform coherent measurements in order to speed up the calibration process. However, non-coherent accumulations allow to perform sparse measurements, profiting from the fact that the calibration parameters are stable over time. This makes the calibration process interleaved with the normal data transmission or reception, leaving it consuming vanishing resources.

5.6 Numerical Validation

In this section, we assess numerically the performance of the various calibration algorithms and also compare them against their CRB. We first simulate the performance of the proposed group-based fast calibration method (introduced in section 5.3 as an innovation enabled by the general calibration framework). We use the mean square error (MSE) as the performance evaluation metric and CRB as the benchmark. The Tx and Rx calibration parameters for the base station antennas are assumed to have random phase uniformly distributed over $[-\pi, \pi]$ and a uniformly distributed amplitude variation in the range $[1-\delta, 1+\delta]$ where δ is chosen such that the standard deviation of the squared magnitude is 0.1. For a fair comparison of MSE across different schemes, the number of channel uses have to be kept constant. Hence, we compare the fast calibration method against the Avalanche scheme proposed in [29]. Note that the Argos and the method from Rogalin et al. in [26] are not fast algorithms and need channel uses on the order of M, so they cannot be compared with the fast calibration method. The number of antennas that transmit at each time instant is shown in Table 5.1. FC-I corresponds to a fast calibration scheme where the antenna grouping is exactly the same as for Avalanche. However, we also try a more equally partitioned grouping of antennas in FC-II. The pilots used for transmission are unit magnitude with random phase uniform in $[-\pi, \pi]$. The channel between all the BS antennas is assumed to be i.i.d Rayleigh fading.

Table 5.1: Number of antennas transmitting at each channel use.

Scheme	Antennas transmitting per channel use. $M = 64$											
Avalanche	1	1	2	3	4	5	6	7	8	9	10	8
FC-I	1	1	2	3	4	5	6	7	8	9	10	8
FC- II	5	5	5	5	5	5	5	5	6	6	6	6

The performance of these schemes is depicted in Fig. 5.5 for $M = 64$. From Section 5.3, it can be seen that the minimal number of channel uses required for calibration here is 12. The performance is averaged over 1000 channel realizations and calibration parameters. The MSE computation for fast calibration is done after satisfying the norm and phase constraints mentioned in Section 5.4. In Fig. 5.5, the performance of our proposed fast

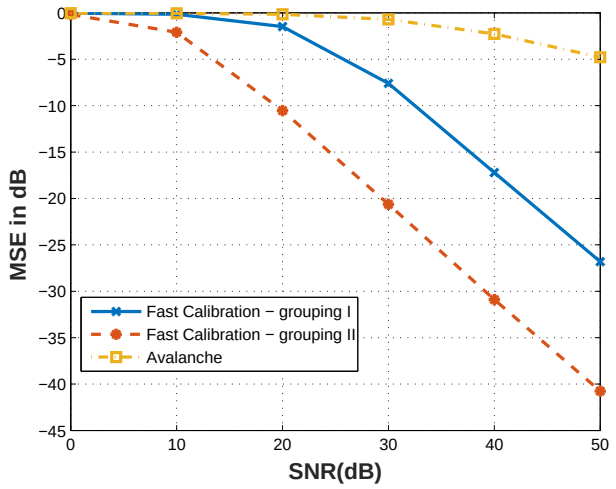


Figure 5.5: Comparison of Fast calibration with Avalanche scheme ($M = 64$ and the number of channel use is 12). The curves are averaged across 1000 channel realizations.

calibration is far superior to that of the Avalanche scheme as it does not have the issue of error propagation and fully makes use of the bi-directional measurements. In addition, its performance improves when the group sizes are allocated more equitably. Intuitively, the overall estimation performance of the fast calibration would be limited by the largest group size and hence it is reasonable to use a grouping scheme that tries to minimize the size of the largest antenna group.

Fig. 5.6 depicts the MSE for the LS estimator under the fast calibration scheme with the CRB derived as in Section 5.4 for a given channel realization and $M = 64$. The antenna grouping is as in FC-II mentioned in Table 5.1. The MSE computation is done after satisfying the norm and phase constraints mentioned in 5.4. Note that in contrast to the performance shown in Fig. 5.5, Fig. 5.6 gives the plot for a single channel realization. We observe that the fast calibration scheme almost overlap the CRB, which implies that the estimation technique needs no further improvement for the given $\mathbf{F}$.

Fig. 5.7 compares the MSE across different SNR for the Argos and calibration schemes from Rogalin in al. in [26] with the CRB derived as in 5.4 for a given channel realization and $M = 16$. The MSE computation is done after satisfying the "norm and phase" constraints mentioned in 5.4.

Fig. 5.8 compares the AML method in [28] to the CRB derived as in Section 5.4 for a given channel realization and $M = 16$ with only one antenna in each group. The MSE computation assumes the first calibration parameter to be known. Also plotted is the CRB as given in [28] (termed CRB Viera in the figure) assuming the channel estimate is fully known (the mean is known and the variance is negligible). In this plot, we use the CRB equation in (5.31) that is derived under the constraint that the first calibration parameter is known. We note that our CRB and MSE performance overlap. This demonstrates the accuracy of our approach to the CRB derivation.

So far, we have focused on an i.i.d internal channel and we have seen in Fig. 5.5 that size of the transmission groups is an important parameter that determines the MSE of the calibration parameter estimates. We now consider a more realistic scenario where the internal channel is based on the geometry of the BS antenna array and make some

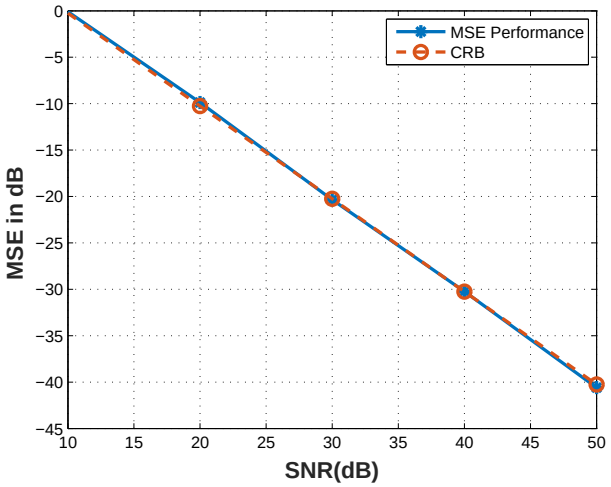


Figure 5.6: Comparison of Fast calibration with CRB ($M = 64$, the number of channel use is 12). The plots are generated over one realization of an i.i.d Rayleigh channel.

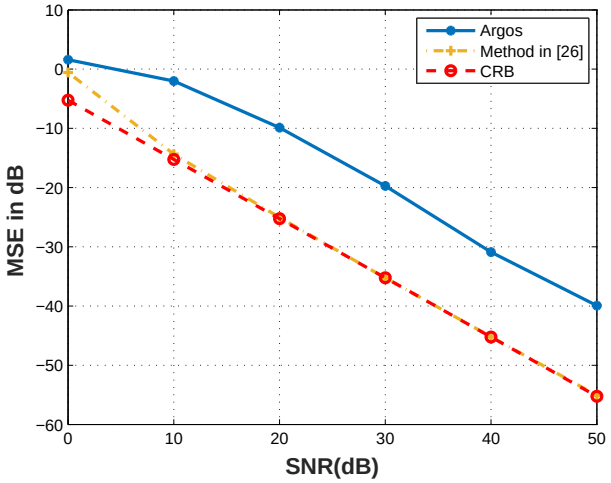


Figure 5.7: Comparison of Argos and the method of successive single antenna transmission and joint estimation with CRB ($M = 16$ and the number of channel uses is 16). The plots are generated over one realization of an i.i.d Rayleigh channel.

observations on the choice of the antennas to form a group. We consider an array of 64 antennas arranged as in Fig. 5.9. The path loss between any two antennas is a function of the distance $d_{i\to j}$ between the two antennas, $(4\pi\frac{d_{i\to j}}{\lambda})^2$, where λ is the wavelength of the received signal. In the simulations, the distance between adjacent antennas, d, is chosen as $\frac{\lambda}{2}$. Such a model was also observed experimentally in [28]. The phase of the channel between any two antennas is modeled in two different ways, due to non-availability of a true model: 1) random phase, 2) phase dependent on the delay between the two antennas. i.e, $2\pi\frac{d_{i\to j}}{\lambda}$.

Fig. 5.10 compares the MSE and the CRB under the two different assumptions of channel phase, when the antennas transmit one at a time (i.e, no grouping) and shows that the choice of phase has no impact when only one antenna transmits at a time (in this scenario,

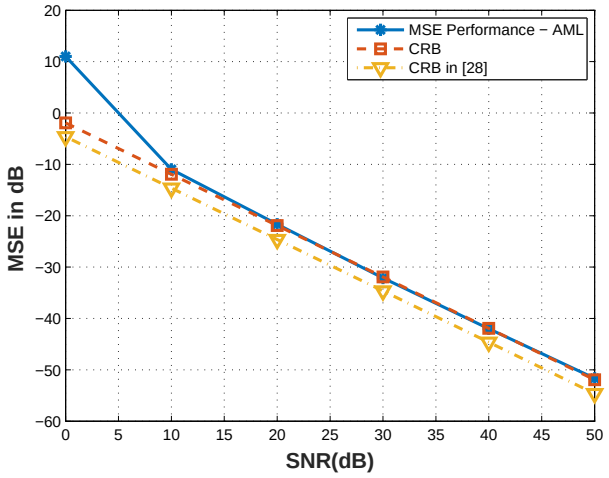


Figure 5.8: Comparison of Alternating ML with CRB for antenna grouping of size unity and $M = 16$.

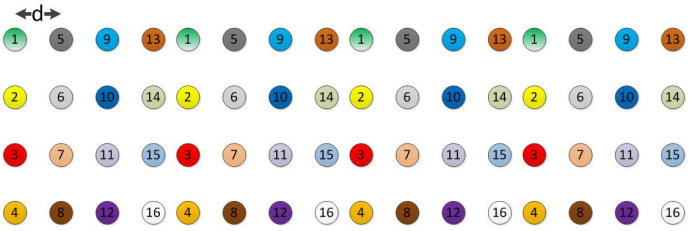


Figure 5.9: 64 Antennas arranged as a 4×16 grid.

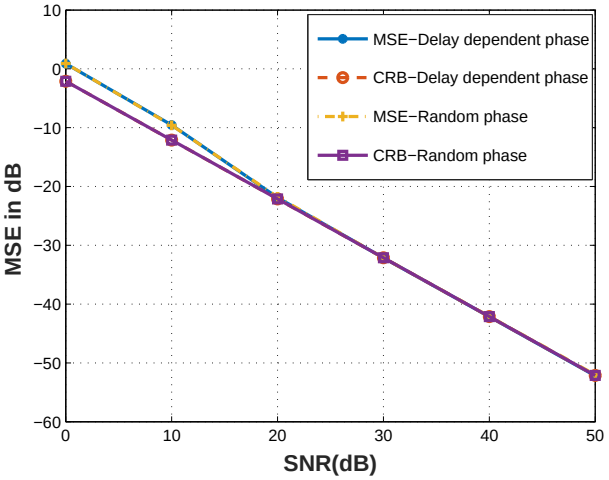


Figure 5.10: MSE and CRB with delay dependent and random assumptions for the channel phase for an antenna transmit group size of 1 ($M = 16$ and number of channel uses is 16).

the SNR is defined as the signal to noise ratio observed at the receive antenna nearest to the transmitter).

We next consider a scenario where antennas transmit in groups of 4 each. Note that this is not the fastest grouping possible, but the example is used for the sake of illustration.

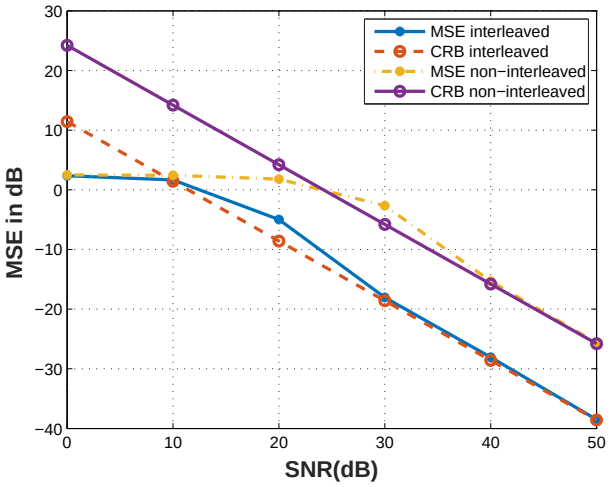


Figure 5.11: Interleaved and non-interleaved MSE and CRB for random phase assumptions for the channel for an antenna transmit group size of 4 ($M = 64$ and the number of channel uses is 16).

We consider two different choices to form the antenna group: 1) Interleaved grouping corresponding to selecting antennas with the same numbers as in Fig. 5.9 into one group, 2) Non-interleaved grouping corresponding to selecting antennas in each column into one group.

Fig. 5.11 shows that in the case of randomized phase for the channel between antenna elements, interleaving of the antennas used in a group result in performance gains of about 10dB. Intuitively, the interleaving of the antennas ensures that when this group transmits, the received signal to the rest of the antennas appear as independent as possible. Also note that at regions where the MSE begins to saturate, the CRB is higher than the MSE and is no longer applicable, due to reasons explained in section 5.4.4. Fig. 5.12 shows the performance in the case of delay dependent phase for the internal channel. In this scenario, interleaving is absolutely essential to ensure identifiability in the fast calibration estimator. In fact, for the non-interleaved case, the Fischer matrix, apart from the singularity that occurs due to the scale factor ambiguity, is very ill-conditioned, leading to a CRB that falls out of the range of Fig. 5.12. The corresponding estimation MSE remains bounded though as discussed in Section 5.4.4.

5.7 Summary

In this work we presented a general calibration framework which unifies the existing calibration schemes. We show that different antenna partition leads to different calibration methods. By deriving the calibration parameter estimators under this general framework, we point out that most existing calibration methods are based on LS estimation.

The unified framework opens up new possibilities on calibration and as an example, we propose a family of fast calibration schemes based on antenna grouping. The number of slots needed for the whole calibration process is in the order of a square root of the antenna array size rather than scaling linearly with number of antennas. In fact it can be

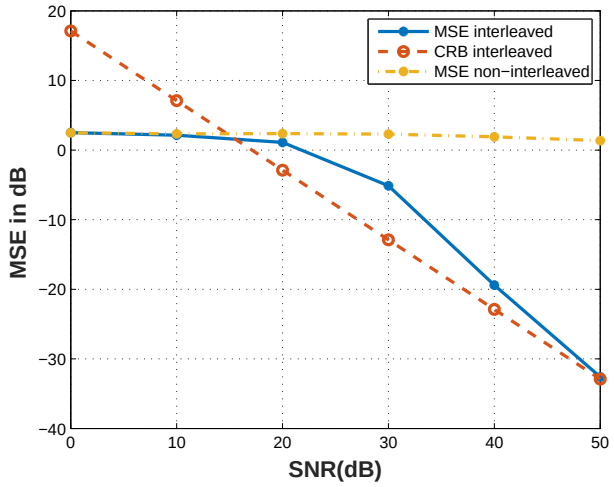


Figure 5.12: Interleaved and non-interleaved MSE and CRB for delay dependent phase assumptions for the channel for an antenna transmit group size of 4 ($M = 64$ and the number of channel uses is 16).

as fast as the existing Avalanche calibration method, but avoids the severe error propagation problem and greatly outperforms its counterpart, as has been shown by simulation results. We also come up with a simple and elegant CRB formulation for the estimation of the calibration parameters. As the group calibration reduces to the existing calibration methods for groups of size unity, the CRB can be used to evaluate the existing state of the art calibration methods as well.

Moreover, we also differentiate the notion of coherent and non-coherent accumulations for calibration observations. We illustrate that, it is possible to perform calibration measurement using slots sparsely distributed on the time axis, which makes the calibration process consume a vanishing fraction of time resources and will not interrupt the ongoing data service.

In the simulation part, we compared the performance of different calibration methods with the CRB as the benchmark. We also show that interleaved grouping outperforms non-interleaved grouping.

Appendices

5.A Optimal grouping

Lemma 1. *Fix $K \geq 1$. Let us define an optimal grouping as the solution $G^*, L_1^*, \ldots, L_{G^*}^*$ of the maximization*

$$\max_{\sum_{i=1}^{G} L_i = K} \sum_{i<j} L_i L_j. \tag{5.41}$$

Then the optimal grouping corresponds to the case $L_1^* = \cdots = L_{G^*}^* = 1$ with $G^* = K$. The number of calibrated antennas is then equal to $\frac{1}{2}K(K-1)+1$.

Proof. Since the variables $L_1, \ldots, L_G, G$ are discrete and $\sum_{i<j} L_j L_i$ is upper bounded by K^2, (5.41) admits at least one solution. Let $\mathbf{L} = (L_1, \ldots, L_G)$ be such a solution. We reason by contradiction: suppose that there exists j such that $L_j > 1$. Without loss of generality, we can suppose that $L_G > 1$. Then, we can break up group G and add one group which contains a single antenna, i.e. let us consider $\mathbf{L}' = (L_1, \ldots, L_G - 1, 1)$. In that case, it holds $\sum_{i=1}^{G} L_i = \sum_{i=1}^{G+1} L'_i = K$ and

$$\begin{aligned} & \sum_{j=2}^{G+1} \sum_{i=1}^{j-1} L'_j L'_i \\ = & \sum_{j=2}^{G-1} \sum_{i=1}^{j-1} L'_j L'_i + (L'_G + L'_{G+1}) \sum_{i=1}^{G-1} L'_j L'_i + L'_G L'_{G+1} \\ = & \sum_{j=2}^{G} \sum_{i=1}^{j-1} L_j L_i + L'_G > \sum_{j=2}^{G} \sum_{i=1}^{j-1} L_j L_i \end{aligned}$$

which contradicts the fact that $\mathbf{L}$ is solution to (5.41). We conclude therefore that $L_j = 1$ for any j and $G^* = K$. □

5.B Construction of $\boldsymbol{\mathcal{F}}^{\perp}$

We show in the following that the column space of $\boldsymbol{\mathcal{F}}^{\perp}$ defined by (5.35) spans the orthogonal complement of the column space of $\boldsymbol{\mathcal{F}}$ assuming that $\mathbf{P}_i$ is full rank for all i and that either $L_i \geq M_i$ or $M_i \geq L_i$ for all i.

Proof. First, using $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$, it holds

$$\underbrace{\begin{bmatrix} \mathbf{I}_{L_i} \otimes \mathbf{P}_j^T \mathbf{F}_j & -\mathbf{P}_i^T \mathbf{F}_i \otimes \mathbf{I}_{L_j} \end{bmatrix}}_{L_i L_j \times (L_i M_j + L_j M_i)} \underbrace{\begin{bmatrix} \mathbf{P}_i^T \mathbf{F}_i \otimes \mathbf{I}_{M_j} \\ \mathbf{I}_{M_i} \otimes \mathbf{P}_j^T \mathbf{F}_j \end{bmatrix}}_{(L_i M_j + L_j M_i) \times M_i M_j} = \mathbf{0} \,. \tag{5.42}$$

Then, the row space of the left matrix of (5.42) is orthogonal to the column space of the right matrix. As $\boldsymbol{\mathcal{F}}$ in (5.27) and $\boldsymbol{\mathcal{F}}^{\perp H}$ are block diagonal with blocks of the form of (5.42), it suffices then to prove that the following matrix $\mathbf{M}$ has full column rank, i.e. $L_i M_j + L_j M_i$, which is then also its row rank

$$\mathbf{M} := \begin{pmatrix} \mathbf{I}_{L_i} \otimes \mathbf{P}_j^T \mathbf{F}_j & -\mathbf{P}_i^T \mathbf{F}_i \otimes \mathbf{I}_{L_j} \\ (\mathbf{F}_i \mathbf{P}_i)^* \otimes \mathbf{I}_{M_j} & \mathbf{I}_{M_i} \otimes (\mathbf{F}_j \mathbf{P}_j)^* \end{pmatrix}. \tag{5.43}$$

Denote $\mathbf{A}_i := \mathbf{P}_i^T \mathbf{F}_i \in \mathbb{C}^{L_i \times M_i}$ and $\mathbf{A}_j := \mathbf{P}_j^T \mathbf{F}_j \in \mathbb{C}^{L_j \times M_j}$. Then, by assumption, it holds that either $\operatorname{rank}(\mathbf{A}_i) = M_i$ and $\operatorname{rank}(\mathbf{A}_j) = M_j$ or $\operatorname{rank}(\mathbf{A}_i) = L_i$ and $\operatorname{rank}(\mathbf{A}_j) = L_j$. Let $\mathbf{x} = [\mathbf{x}_1^T \; \mathbf{x}_2^T]^T$ be such that $\mathbf{Mx} = 0$ and show that $\mathbf{x} = 0$. Since $\mathbf{Mx} = 0$, it holds

$$\begin{cases} (\mathbf{I}_{L_i} \otimes \mathbf{A}_j)\mathbf{x}_1 - (\mathbf{A}_i \otimes \mathbf{I}_{L_j})\mathbf{x}_2 = 0 \\ (\mathbf{A}_i^H \otimes \mathbf{I}_{M_j})\mathbf{x}_1 + (\mathbf{I}_{M_i} \otimes \mathbf{A}_j)\mathbf{x}_2 = 0. \end{cases}$$

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be matrices such that $\operatorname{vec}(\mathbf{X}_1) = \mathbf{x}_1$ and $\operatorname{vec}(\mathbf{X}_2) = \mathbf{x}_2$. Then

$$\begin{cases} \mathbf{A}_j \mathbf{X}_1 - \mathbf{X}_2 \mathbf{A}_i^T = 0 \\ \mathbf{X}_1 \mathbf{A}_i^* + \mathbf{A}_j^H \mathbf{X}_2 = 0. \end{cases}$$

Multiplying the first equation by $\mathbf{A}_j^H$ and the second by $\mathbf{A}_i^T$, and summing up the two, we get $\mathbf{A}_j^H\mathbf{A}_j\mathbf{X}_1 + \mathbf{X}_1(\mathbf{A}_i\mathbf{A}_i^H)^* = 0$ which is a Sylvester's equation which admits a unique solution if $\mathbf{A}_j^H\mathbf{A}_j$ and $-(\mathbf{A}_i\mathbf{A}_i^H)^*$ have no common eigenvalues.

On the other hand, the eigenvalues of $\mathbf{A}_j^H\mathbf{A}_j$ and $\mathbf{A}_i\mathbf{A}_i^H$ are real positive so common eigenvalues of $\mathbf{A}_j^H\mathbf{A}_j$ and $-(\mathbf{A}_i\mathbf{A}_i^H)^*$ can only be 0. However, this does not occur since by the assumptions either $\mathbf{A}_j^H\mathbf{A}_j$ or $\mathbf{A}_i\mathbf{A}_i^H$ is full rank. We can then conclude that $\mathbf{X}_1 = 0$, i.e. $\mathbf{x}_1 = 0$. We prove similarly that $\mathbf{x}_2 = 0$ which ends the proof. □

Chapter 6

Calibration for Hybrid Beamforming Structure

To bring massive MIMO to practice, it is essential to reduce the cost of building up such complex systems. Among the most promising solutions, hybrid analog and digital beamforming structure has achieved great attention. By introducing phase shifters and reducing the number of expensive components on digital and RF chains, such as DACs/ADCs as well as signal mixers, hybrid beamforming structure opens up possibilities to build relatively low cost massive MIMO systems.

A common way of enabling hybrid beamforming is to pre-define a set of fixed beams in the DL on which pilots are transmitted to a UE who then simply selects the best beam and then sends the index back to the BS, who will use it directly for data transmission [57,58]. Such systems have also been specified for LTE-Advanced Pro, in the so-called FD-MIMO technology [35], but are clearly suboptimal compared to the the case where full CSIT is available [59]. Under the assumption of full CSIT, a hybrid massive MIMO system can achieve the same performance of any fully digital beamforming scheme, as long as the number of RF chains is at least twice the number of data schemes [60]. However, acquisition of CSIT in a hybrid massive MIMO system is a non-trivial matter, both for FDD and TDD systems.

The problem was studied in the mmWave band in [61], where the channel can be considered to have only a few number of dominant rays because of the poor scattering nature of the channel. While this method works out well for mmWave, it can hardly be generalized to an arbitrary channel, especially when hybrid beamforming massive MIMO systems are used in a sub-6GHz band.

Theoretically, if the hybrid analog digital beamforming system is under TDD mode, channel reciprocity can be used for CSIT acquisition. However, regardless the variety of existing calibration methods for fully digital systems, none of them can be directly used in a hybrid beamforming structure. This is the main reason why TDD reciprocity based methods have been left behind in hybrid beamforming massive MIMO systems [58]. In this chapter, we introduce an equivalent hybrid model, which allows us to perform TDD reciprocity calibration. We show that by addressing the CSIT acquisition problem from a different aspect, we can avoid beam training or selection and achieve near perfect CSIT without any assumptions on the channel.

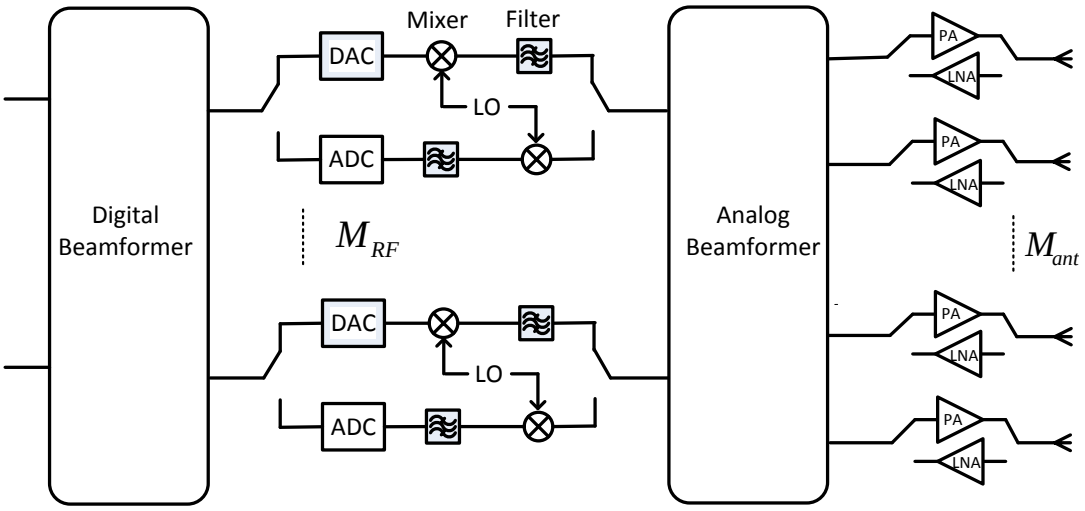


Figure 6.1: Structure of a TDD hybrid beamforming transceiver, where both the transmit and receive paths are shown. The transceiver can dynamically change the connexion of different switches on the two sides of the analog beamformer to set itself to the transmitting or receiving mode.

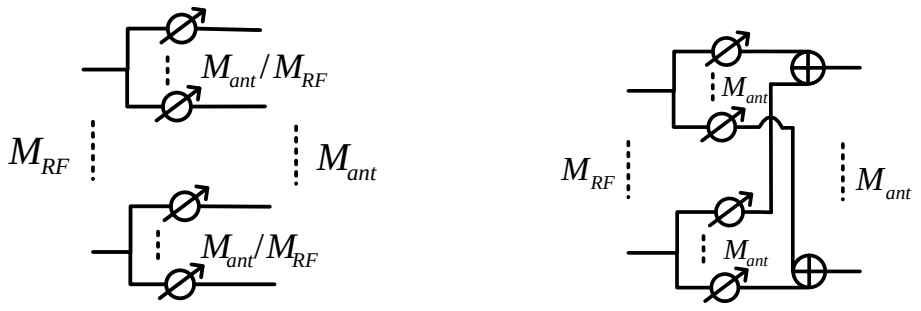
6.1 Hybrid structure

The structure of a TDD hybrid beamforming transceiver is shown in Fig. 6.1 [62] where the digital beamformer is connected to M_{RF} RF chains, which then go through an analog beamforming network and are connected with power amplifiers (PA)/low noise amplifier (LNA) and M_{ant} antennas. Note that it is also possible to place PAs and LNAs in the RF chains before the analog beamformer so that the number of amplifiers are less. However, in that case, each amplifier needs more power since it amplifies signal for multiple antennas. Additionally, in the transmission mode, the insertion loss of analog precoder working in the high power region makes the transceiver less efficient in terms of power consumption. In reception mode, the fact of having phase shifters before LNAs also results in a higher noise figure in the receiver. It is thus a better choice to have PAs and LNAs close to antennas. To this reason, we stick our study in this chapter to the structure in Fig. 6.1. The discussion in this chapter, however, can also be applied to the case where the PAs/LNAs are placed before the analog beamformer.

The analog beamformer is interpreted as analog precoder and combiner in the transmit and receive path, respectively. Two types of architecture can be found in literature [58,63]:

- **Subarray architecture:** Each RF chain is connected to M_{ant}/M_{RF} phase shifter as shown in Fig. 6.2a. Such a structure can be found in [57, 64–66]
- **Fully connected architecture:** M_{ant} phase shifter are connected to each RF chain. Phase shifters with the same index are then summed up to be connected to the corresponding antenna, as shown in Fig. 6.2b. This structure can be found in [61, 67–69].

Since the BS is not fully digital, assigning orthogonal pilots to different antennas for channel estimation per antenna can not be used. Additionally, even assuming that we can



(a) Subarray architecture.

(b) Fully connected.

Figure 6.2: Two types of analog beamforming structure.

have perfect channel estimation for all antennas at the UE, it is unfeasible to feedback this information back to the BS, because in a massive MIMO system, the UL overhead will be so heavy that at the time the BS gets the whole CSIT, the information has already outdated.

In order to address this problem, we are going to make use of TDD channel reciprocity for CSIT acquisition. We firstly show how this is possible for "subarray architecture" by enable reciprocity calibration. We then provide some ideas to calibrate a fully connected hybrid beamforming architecture.

6.2 Calibration for subarray architecture

6.2.1 System model

Consider a sub-array hybrid beamforming system as shown in Fig. 6.3, where node A with M_{ant}^A antennas communicates M_s data streams to node B with M_{ant}^B antennas. M_{RF}^A and M_{RF}^B are the number of RF chains, such that $M_s \leq M_{RF}^A \leq M_{ant}^A$ and $M_s \leq M_{RF}^B \leq M_{ant}^B$. In Fig. 6.3, we use $\mathbf{V}_{BB}^A \in \mathbb{C}^{M_{RF}^A \times M_s}$ and $\mathbf{W}_{BB}^B \in \mathbb{C}^{M_s \times M_{RF}^B}$ to represent the base-band digital beamforming matrix at node A and B, respectively. $\mathbf{V}_{RF}^A \in \mathbb{C}^{M_{ant}^A \times M_{RF}^A}$ and $\mathbf{W}_{RF}^B \in \mathbb{C}^{M_{RF}^B \times M_{ant}^B}$ are the analog beamforming precoders and combiners. We use $\mathbf{T}_1^A \in \mathbb{C}^{M_{RF}^A \times M_{RF}^A}$, $\mathbf{T}_2^A \in \mathbb{C}^{M_{ant}^A \times M_{ant}^A}$, $\mathbf{R}_1^B \in \mathbb{C}^{M_{RF}^B \times M_{RF}^B}$ and $\mathbf{R}_2^B \in \mathbb{C}^{M_{ant}^B \times M_{ant}^B}$ to represent the transfer functions of the corresponding hardwares. The diagonal elements of $\mathbf{T}_1^A$ and $\mathbf{R}_1^B$ capture the hardware characteristics of the M_{RF}^A and M_{RF}^B RF chains including the DACs/ADCs, signal mixers and some other components around, whereas, their off-diagonal elements represent the RF crosstalk. Similarly the diagonal of $\mathbf{T}_2^A$ and $\mathbf{R}_2^B$ are used to represent the properties of power amplifiers as well as some surrounding components after phase shifter on each branch and their off-diagonal elements represent RF crosstalk and antenna mutual coupling [39]. If we transmit a signal $\mathbf{s}$ through a channel $\mathbf{C} \in \mathbb{C}^{M_{ant}^B \times M_{ant}^A}$, at the output of the digital combiner of node B, we have

$$\mathbf{y} = \mathbf{W}_{BB}^B \mathbf{R}_1^{UE} \mathbf{W}_{RF}^B \mathbf{R}_2^B \mathbf{C} \mathbf{T}_2^A \mathbf{V}_{RF}^A \mathbf{T}_1^A \mathbf{V}_{BB}^A \mathbf{s} + \mathbf{n}, \tag{6.1}$$

where $\mathbf{y}$ is the $M_s \times 1$ received signal vector and $\mathbf{n} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$ is the noise vector.

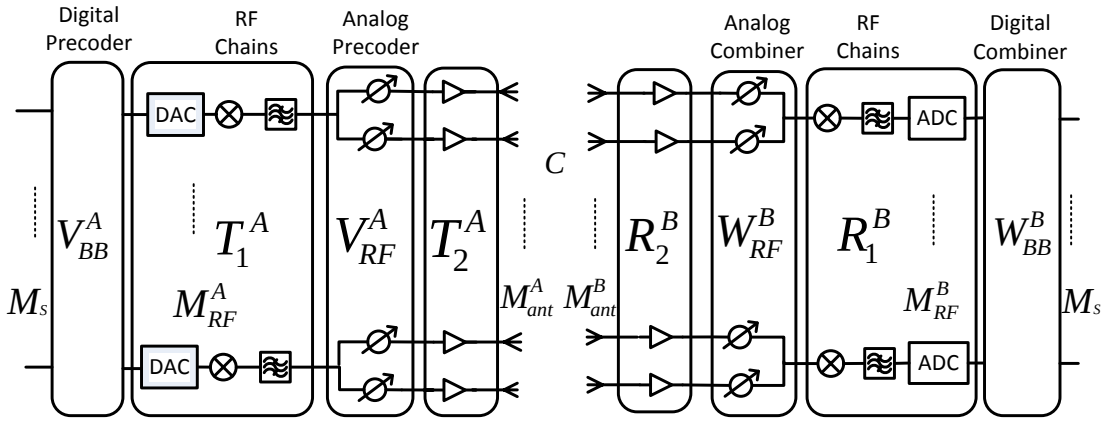


Figure 6.3: Hybrid beamforming system where node A is transmitting M_s data schemes to node B. The switches at node A are connected to the transmit path whereas those at node B are connected to the receive path.

In a TDD system, the physical channel is reciprocal within the channel coherence time, i.e., in the reverse transmission, the physical channel from B to A can be represented by $\mathbf{C}^T$.

6.2.2 Equivalent system model

In order to calibrate the hybrid beamformng TDD systems, let us firstly introduce an equivalent system model which simplifies the signal model in (6.1), where we observe the hardware blocks are mixed up with digital and analog beamforming matrices. Note that $\mathbf{T}_1^A$ and $\mathbf{R}_1^B$ can be assumed diagonal, since proper RF circuit design usually ensures very small RF crosstalk (i.e., RF leakage from one RF chain to the others) with regard to the diagonal values. Since $\mathbf{V}_{RF}^A$ and $\mathbf{W}_{RF}^B$, representing the analog beamformers for each RF chain, have block diagonal structures, the matrix multiplication is commutative if we introduce a Kronecker product such as $\mathbf{V}_{RF}^A\mathbf{T}_1^A = (\mathbf{T}_1^A \otimes \mathbf{I}_A)\mathbf{V}_{RF}^A$ and $\mathbf{R}_1^B\mathbf{W}_{RF}^B = \mathbf{W}_{RF}^B(\mathbf{R}_1^B \otimes \mathbf{I}_B)$, where $\mathbf{I}_A$ and $\mathbf{I}_B$ are identity matrices of size M_{ant}^A/M_{RF}^A and M_{ant}^B/M_{RF}^B, respectively. The signal model in (6.1) thus has an equivalent representation as

$$\mathbf{y} = \underbrace{\mathbf{W}_{BB}^B\mathbf{W}_{RF}^B}_{\mathbf{W}_B}\underbrace{(\mathbf{R}_1^B \otimes \mathbf{I}_B)\mathbf{R}_2^B}_{\mathbf{R}_B}\,\mathbf{C}\,\underbrace{\mathbf{T}_2^A(\mathbf{T}_1^A \otimes \mathbf{I}_A)}_{\mathbf{T}_A}\underbrace{\mathbf{V}_{RF}^A\mathbf{V}_{BB}^A}_{\mathbf{V}_A}\mathbf{s} + \mathbf{n}, \tag{6.2}$$

where we group up the digital and analog transmit and receive beamforming matrices into $\mathbf{V}_A$ and $\mathbf{W}_B$. The hardware transfer functions are also put together and become $\mathbf{T}_A$ and $\mathbf{R}_B$.

An intuitive understanding of this alternative representation on the transmit part is shown in Fig. 6.4, where we

1. replace all shared hardware components (mixers, filters) on RF chain by its replicas on each branch with phase shifters;
2. change the order of hardware components such that all components in $\mathbf{T}_A$ go to the front end near the antennas.

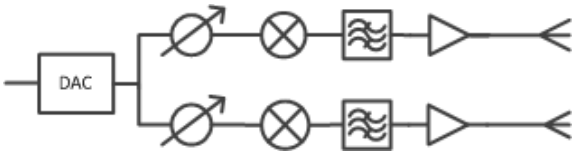


Figure 6.4: Equivalent hybrid structure where shared hardware components (mixers, filters) on RF chain are copied on each branch with phase shifters. The hardware components are then re-ordered such that all components in $\mathbf{T}_A$ go to the front end near the antennas.

Note that this equivalent model is general for different hardware implementation, i.e., no matter how hardware impairments are distributed on the hybrid structure, we can always use these two steps to create an equivalent system model. For example, if there's any hardware impairment within the phase shifter or in DAC, they can also be extracted out and put into $\mathbf{T}_A$ using the same methodology.

6.2.3 Effective channel estimation

Consider sending pilots ($\mathbf{s}_A = \mathbf{p}_A$) from A to B using K transmit precoders combined with L different receive combiners, we can totally accumulate KL measurements:

$$\underbrace{[\mathbf{y}_{B,l,k}]}_{\mathbf{Y}_B} = \underbrace{[\mathbf{W}_{B,1}^T, \ldots, \mathbf{W}_{B,L}^T]^T}_{\tilde{\mathbf{W}}_B} \mathbf{H}_{A\to B} \underbrace{[\mathbf{V}_{A,1}\mathbf{p}_{A,1}, \ldots, \mathbf{V}_{A,K}\mathbf{p}_{A,K}]}_{\tilde{\mathbf{P}}_A} + \underbrace{[\mathbf{n}_{l,k}]}_{\mathbf{N}}. \tag{6.3}$$

where $\mathbf{y}_{B,l,k}$ is the block element of $\mathbf{Y}_B$ on the l^{th} row and k^{th} column. $\tilde{\mathbf{W}}_B$ and $\tilde{\mathbf{P}}_A$ are matrices of size $M_s L \times M_{ant}^B$ and $M_{ant}^A \times K$, respectively. To obtain the channel estimation, we vectorize the receive vector as

$$\mathrm{vec}(\mathbf{Y}_B) = \underbrace{\tilde{\mathbf{P}}_A^T \otimes \tilde{\mathbf{W}}_B}_{\mathbf{D}} \cdot \mathrm{vec}(\mathbf{H}_{A\to B}) + \mathrm{vec}(\mathbf{N}), \tag{6.4}$$

where we define $\mathbf{D} = \tilde{\mathbf{P}}_A^T \otimes \tilde{\mathbf{W}}_B$. The LS channel estimator is

$$\mathrm{vec}(\mathbf{H}_{A\to B}) = (\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H \cdot \mathrm{vec}(\mathbf{Y}_B). \tag{6.5}$$

In order to guarantee that the estimation problem is over determined, we should have $\mathrm{rank}(\mathbf{D}) \geq M_{ant}^A \times M_{ant}^B$, where $\mathrm{rank}(\mathbf{D}) = \mathrm{rank}(\tilde{\mathbf{P}}_A^T)\mathrm{rank}(\tilde{\mathbf{W}}_B)$ according to Kronecker product's property on matrix rank. Noting that $\mathrm{rank}(\tilde{\mathbf{P}}_A^T) \leq \min(M_{ant}^t, K)$ and $\mathrm{rank}(\tilde{\mathbf{W}}_B) \leq \min(M_s L, M_{ant}^B)$, thus, in order to meet the sufficient condition of over determination on the estimation problem, we should have $K \geq M_{ant}^A$ and $L \geq M_{ant}^B/M_s$.

Since the objective here is to estimate the effective channel, digital precoder and combiner are not necessarily needed, i.e. pilots for channel estimation can be inserted after the digital precoder. In this case $M_s = M_{RF}^A$ and $L \geq M_{ant}^B/M_{RF}^B$. Additionally, in a multi-carrier system, where, for example, OFDM modulation is used, it is possible to allocate different carriers to the pilots of different RF chains. Assuming β the number of frequency multiplexing factor on transmit RF chains, the number of the needed transmit precoder $K \geq M_{ant}^A/\beta$.

The effective channel estimation can be used to obtain UL channel estimation but will also be served to estimate calibration matrices as will be presented hereafter.

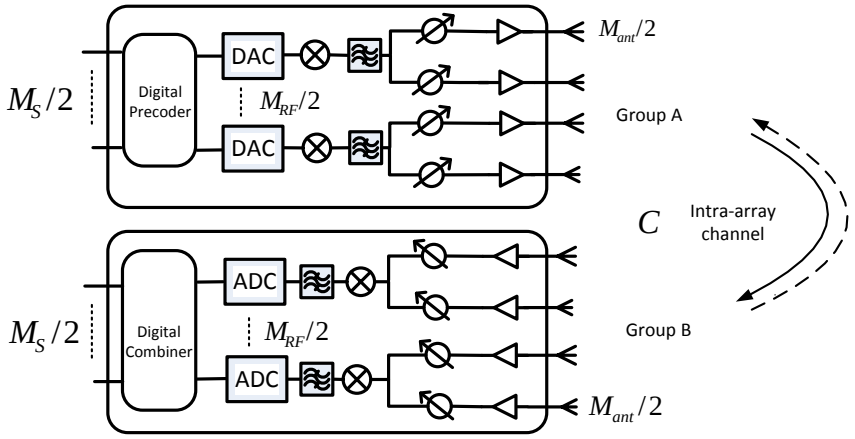


Figure 6.5: Internal calibration where the whole antenna array is partitioned into group A and group B. We then perform intra-array measurement between the two groups.

6.2.4 Internal reciprocity calibration

Internal calibration means that the pilot-based channel estimation happens between different antennas of the same transceiver. Assuming a transceiver with M_{ant} antennas, which can be partitioned into two groups A and B, e.g., $A = \{1, 2, ..., \frac{M_{ant}}{2}\}$ and $B = \{\frac{M_{ant}}{2} + 1, ..., M_{ant}\}$, as shown in Fig. 6.5. When the antennas in group A are connected to the transmit path of $\frac{M_{RF}}{2}$ RF chains, the antennas in group B are connected to the receive path of the rest $\frac{M_{RF}}{2}$ RF chains. We firstly perform an intra-array transmission from A to B, and within the channel coherence time, we switch the roles of group A and B in order to transmit signal from B to A. The bi-directional received signals are given by

$$\begin{cases} \mathbf{y}_{A\to B} = \mathbf{W}_B\mathbf{R}_B\mathbf{C}\mathbf{T}_A\mathbf{V}_A\mathbf{p}_A + \mathbf{n}_{A\to B}, \\ \mathbf{y}_{B\to A} = \mathbf{W}_A\mathbf{R}_A\mathbf{C}^T\mathbf{T}_B\mathbf{V}_B\mathbf{p}_B + \mathbf{n}_{B\to A}, \end{cases} \tag{6.6}$$

where $\mathbf{p}_A$ and $\mathbf{p}_B$ are the transmit pilots; $\mathbf{C}$ is the reciprocal intra-array channel and $\mathbf{n}_{A\to B}$ and $\mathbf{n}_{B\to A}$ are the noise components.

If we use $\mathbf{H}_{A\to B} = \mathbf{R}_B\mathbf{C}\mathbf{T}_A$ and $\mathbf{H}_{B\to A} = \mathbf{R}_A\mathbf{C}^T\mathbf{T}_B$ to represent the bi-directional channels between group A and B, including the physical channel in the air as well as transceiver's hardware, similar to (4.2), we have

$$\mathbf{H}^T_{A\to B}\mathbf{F}_B = \mathbf{F}_A{}^T\mathbf{H}_{B\to A}, \tag{6.7}$$

where $\mathbf{F}_A = \mathbf{R}_A^{-T}\mathbf{T}_A$ and $\mathbf{F}_B = \mathbf{R}_B^{-T}\mathbf{T}_B$ are the calibration matrices. As pointed out in Section. 4.1, calibration matrices are usually diagonal, thus, $\mathbf{F} = \text{diag}\{\mathbf{F}_A, \mathbf{F}_B\} = \text{diag}\{f_1, \ldots, f_{M_{ant}}\}$.

Internal reciprocity calibration consists in estimating $\mathbf{F}$ based on the intra-array channel measurement $\hat{\mathbf{H}}_{A\to B}$ and $\hat{\mathbf{H}}_{B\to A}$, without any involvement of other transceivers. Since the calibration coefficients stay quite stable during a relatively long time, once they are estimated, we can use them together with instantaneously estimated UL channel estimation to obtain CSIT.

Replacing the real channel in (6.7) by their estimates (minus estimation errors), (6.7) can be viewed as a special case of (5.2) of the general framework in Section 5.1, where $\mathbf{P}_i$ and $\mathbf{P}_j$ are identity matrices in this case, $\mathbf{Y}_{j\to i} = \hat{\mathbf{H}}_{B\to A}$, $\mathbf{Y}_{i\to j} = \hat{\mathbf{H}}_{A\to B}$. Thus, estimating $\mathbf{f}$ (the diagonal of $\mathbf{F}$) is to solve (6.8) by assuming $f_1 = 1$ or under the unit norm constraint:

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\boldsymbol{\mathcal{Y}}(\mathbf{P})\mathbf{f}\|^2, \tag{6.8}$$

where

$$\boldsymbol{\mathcal{Y}}(\mathbf{P}) = \left[\hat{\mathbf{H}}^T_{B\to A} * \mathbf{I} \quad -\mathbf{I} * \hat{\mathbf{H}}^T_{A\to B}\right]. \tag{6.9}$$

The solution is given by (5.9) and (5.10).

Although the general framework in Chapter 5 helps us quickly solve (6.8), we provide also a classical approach to address the problem, which helps us explain some simulation results in Section. 6.4. Let us denote the antenna index in group A and B by i and j, respectively, since $\mathbf{F}$ is a diagonal matrix, from (6.7), we have

$$\begin{gathered} f_j h_{i\to j} = f_i h_{j\to i}, \\ \text{where, } i \in \{1, 2, ..., \frac{M_{ant}}{2}\}, j \in \{\frac{M_{ant}}{2} + 1, ..., M_{ant}\}. \end{gathered} \tag{6.10}$$

The problem then becomes very similar to that in [26]. Let us use J to denote the cost function of a LS estimation problem such as

$$J(f_1, f_2, ..., f_{ant}) = \sum_{i\in A, j\in B} |f_j h_{i\to j} - f_i h_{j\to i}|^2. \tag{6.11}$$

Estimating the calibration coefficients concerns in minimizing J subject to a $\|\mathbf{f}\|^2 = 1$or $f_1 = 1$. We adopt in our method the the unit norm constraint, the Lagrangian function of the constrained LS problem is given by

$$L(\mathbf{f}, \lambda) = J(\mathbf{f}) - \lambda(\|\mathbf{f}\|^2 - 1), \tag{6.12}$$

where λ is the Lagrangian multiplier. By setting the partial derivatives of $L(\mathbf{f}, \lambda)$ with regard to f_i^* and f_j^* to zeros, respectively, where f_i^* and f_i are treated as if they were independent variable [50], we obtain

$$\begin{cases} \dfrac{\partial L(\mathbf{f}, \lambda)}{\partial f_i^*} = \Sigma_{j\in B}(f_i|h_{j\to i}|^2 - f_j h^*_{j\to i} h_{i\to j}) - \lambda f_i = 0, \\ \dfrac{\partial L(\mathbf{f}, \lambda)}{\partial f_j^*} = \Sigma_{i\in A}(f_j|h_{i\to j}|^2 - f_i h^*_{i\to j} h_{j\to i}) - \lambda f_j = 0. \end{cases} \tag{6.13}$$

The matrix representation of (6.13) is $\mathbf{Qf} = \lambda\mathbf{f}$, where $\mathbf{Q} \in \mathbb{C}^{M_{ant}\times M_{ant}}$ has its element on the i-th row and u-th column as

$$Q_{i,u} = \begin{cases} \Sigma_{j\in B}|h_{j\to i}|^2 & \text{for } u = i, \\ -h^*_{u\to i} h_{i\to u} & \text{for } u \in B, \end{cases} \tag{6.14}$$

and its element on the j-th row and u-th column given by

$$Q_{j,u} = \begin{cases} \Sigma_{i\in A}|h_{i\to j}|^2 & \text{for } u = j, \\ -h^*_{u\to j} h_{j\to u} & \text{for } u \in A, \end{cases} \tag{6.15}$$

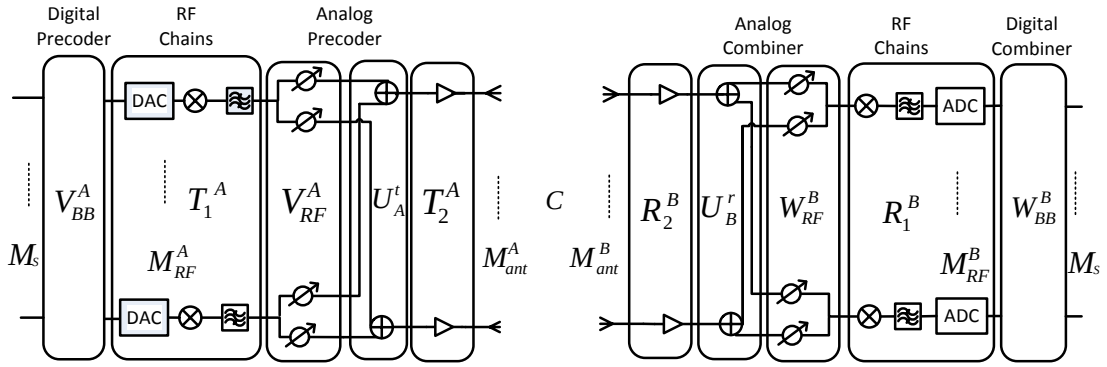


Figure 6.6: Hybrid beamforming system where both the A and B have full connected architecture as the analog beamformer

with all other elements being 0. The solution is given by the eigenvector of $\mathbf{Q}$ corresponding to its eigenvalue with the smallest magnitude, $V_{min}(\mathbf{Q})$.

Note that the general framework under the unit norm constraint gives the solution as $V_{min}(\boldsymbol{\mathcal{Y}}(\mathbf{P})^H\boldsymbol{\mathcal{Y}}(\mathbf{P}))$ in (5.10). If we expand $\boldsymbol{\mathcal{Y}}(\mathbf{P})^H\boldsymbol{\mathcal{Y}}(\mathbf{P})$ in its element form, we observe that $\mathbf{Q}$ and $\boldsymbol{\mathcal{Y}}(\mathbf{P})^H\boldsymbol{\mathcal{Y}}(\mathbf{P})$ are totally the same. Thus the solutions from these two approaches are equivalent.

Note also that when the calibration coefficients are estimated, we can use them together with instantaneously estimated UL channel for DL CSIT acquisition. The effective UL channel estimation can be obtained using the method in Section 6.2.3.

6.3 Calibration for fully connected structure

Until now, we have concentrated on reciprocity calibration for subarray structure. In this section, we give some ideas on how to calibrate a fully connected architecture for CSIT acquisition. Consider a system with A and B both using fully connected hybrid beamforming structure as in Fig. 6.6. We use $\mathbf{U}_A^t \in \mathbb{C}^{M_{ant}^A \times M_{ant}^A M_{RF}^A}$ and $\mathbf{U}_B^r \in \mathbb{C}^{M_{RF}^B M_{ant}^B \times M_{ant}^B}$ to denote the summation array between amplifiers and phase shifters. The signal model (6.1) can be written as

$$\mathbf{y} = \mathbf{W}_{BB}^B \mathbf{R}_1^B \mathbf{W}_{RF}^B \mathbf{U}_B^r \mathbf{R}_2^B \mathbf{C} \mathbf{T}_2^A \mathbf{U}_A^t \mathbf{V}_{RF}^A \mathbf{T}_1^A \mathbf{V}_{BB}^A \mathbf{s} + \mathbf{n}, \tag{6.16}$$

An example of the summation array $\mathbf{U}_A^t$ for $M_{ant}^A = 4$ and $M_{RF}^A = 2$ (i.e. 8 phase shifters) has the following structure:

$$\mathbf{U}_A^t = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.17}$$

As $\mathbf{U}_A^t$ can be viewed as a block row vector composed of M_{RF}^A identity matrix $\mathbf{I}_{M_{ant}^A}$, i.e. $\mathbf{U}_A^t = \begin{bmatrix} \mathbf{I}_{M_{ant}^A} & \mathbf{I}_{M_{ant}^A} & \cdots & \mathbf{I}_{M_{ant}^A} \end{bmatrix}$, we can use a Kronecker product to commute $\mathbf{T}_2^A \mathbf{U}_A^t$

such as $\mathbf{T}_2^A\mathbf{U}_A^t = \mathbf{U}_A^t(\mathbf{I}_{M_{RF}^A} \otimes \mathbf{T}_2^A)$. This is equivalent to move the replicas of the PAs (as well as other components) connected to transmit antennas onto each branch before the summation operation. A similar approach can be adopted for node B, we can thus get an equivalent system model of (6.16) as

$$\begin{aligned}\mathbf{y} = &\underbrace{\mathbf{W}_{BB}^B\mathbf{W}_{RF}^B}_{\mathbf{W}_B}\underbrace{(\mathbf{R}_1^B \otimes \mathbf{I}_{M_{ant}^B})(\mathbf{I}_{M_{RF}^B} \otimes \mathbf{R}_2^B)}_{\mathbf{R}_B} \cdot \underbrace{\mathbf{U}_B^r\mathbf{C}\mathbf{U}_A^t}_{\tilde{\mathbf{C}}} \cdot \\ &\underbrace{(\mathbf{I}_{M_{RF}^A} \otimes \mathbf{T}_2^A)(\mathbf{T}_1^A \otimes \mathbf{I}_{M_{ant}^A})}_{\mathbf{T}_A}\underbrace{\mathbf{V}_{RF}^A\mathbf{V}_{BB}^A}_{\mathbf{V}_A}\mathbf{s} + \mathbf{n},\end{aligned} \tag{6.18}$$

where $\mathbf{I}_{M_{ant}^A}$ and $\mathbf{I}_{M_{RF}^B}$ are identity matrices of size M_{ant}^A and M_{ant}^B, respectively. If we consider $\mathbf{U}_B^r\mathbf{C}\mathbf{U}_A^t$ as a composite propagation channel $\tilde{\mathbf{C}}$, the equivalent signal model is similar to (6.2).

When the system is in reverse transmission, the switches at the node A are connected to the receive path whereas those at the node B are connected to the transmit path. Thus, the reverse composite channel can be written as $\mathbf{U}_A^r\mathbf{C}^T\mathbf{U}_B^t$, which can be easily verified as $\tilde{\mathbf{C}}^T$, implying that the channel reciprocity is maintained. Note that if there exist some hardware impairment in the summation operation $\mathbf{U}_A^t$ or $\mathbf{U}_B^r$, we can represent $\mathbf{U}_A^t$ and $\mathbf{U}_B^r$ as $\mathrm{E}_A^t\mathbf{U}_{A,0}^t$ or $\mathbf{U}_{B,0}^r\mathrm{E}_B^r$ where $\mathbf{U}_{A,0}^t$ and$\mathbf{U}_{B,0}^r$ are the ideal summation matrices as in (6.17), E_A^t and E_B^r are impairment matrices which can be absorbed into $\mathbf{T}_2^A$ or $\mathbf{R}_2^B$.

For a fully connected architecture, internal reciprocity calibration is not feasible since it is not possible to partition the whole antenna array into transmit and receive antenna groups. To enable TDD reciprocity calibration for this structure, a reference UE with a good enough channel should be selected to assist the BS to calibrate, such as [23] proposed for a fully digital system. In this case, the bi-directional transmission no longer happens between two partitioned antenna groups A and B but is between the BS and the UE. The selected reference UE needs to feed back its measured DL channel to the BS during the calibration procedure. Methods in Section 6.2.4 can still be used to estimate the calibration matrices for both BS and UE. Note that although UE feedback is heavy, the calibration does not have to be done very frequently, thus, such a method is still feasible.

Another possible way is to use a dedicated device at the BS to assist the antenna array for calibration, e.g., using a reference antenna as in [25]. Using this method, DL channel measurements feedback from UE can be avoided, but a dedicated digital chain needs to be allocated to the assistant device, introducing an extra cost.

6.4 Simulation results

As a proof-of-concept, we perform simulation of an internal calibration for a sub-array hybrid transceiver with 64 antennas and 8 RF chains. To the extent of our knowledge, signal mixers and amplifiers are the main source of hardware asymmetry. For different RF chains, signal mixers introduce random phases when multiplying the baseband signal with the carrier, whereas the gain imbalance between different amplifiers can cause their output signal having different amplitudes. Apart from these two main factors, other components can also have some minor impacts, e.g., the non-accuracy in the phase shifter can add a further random factor to the phase. In this simulation, we capture the main

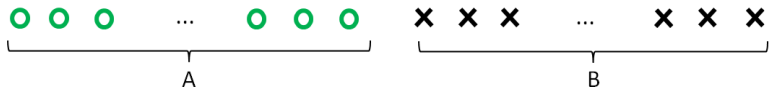
effects of these hardware properties introduced by signal mixers and amplifiers, though the calibration method is not limited to this simplified case. We assume that the random phase introduced by the signal mixers in $\mathbf{T}_1$ and $\mathbf{R}_1$ are uniformly distributed between $-\pi$ and π whereas the amplitude difference in $\mathbf{T}_2$ and $\mathbf{R}_2$ are independent variables uniformly distributed between $[1-\epsilon\ 1+\epsilon]$, with ϵ chosen such that the standard deviation of the squared-magnitude is 0.1.

The intra-array channel model between antenna elements strongly depends on the antenna arrangement in the array, antenna installation, as well as the frequency band. In the simulation, we focus on a sub-6GHz scenario and adopt the experiment based intra-array radio channel in [28], where the physical channel $c_{i,j}$ between two antenna elements i and j in the same planar antenna array is modeled as

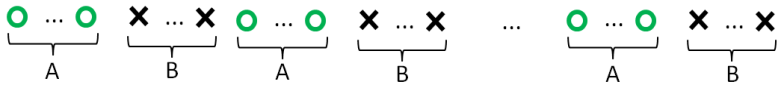
$$c_{i,j} = |\bar{c}_{i,j}|\exp(j2\pi\phi_{i,j}) + \tilde{c}_{i,j}. \tag{6.19}$$

where $\bar{c}_{i,j}$ is the near field path [1] between two antenna elements and $\tilde{c}_{i,j}$ absorbs all other multi-path contributions due to reflections from obstacles around the antenna array. For simplicity reasons, we assume the 64 antennas follows a co-polarized linear arrangement with an antenna space of half of the wavelength. According to the measurements in [28], the magnitude for two half-wavelength spaced antennas are -15dB and at each distance increase of half of the wavelength, $|\bar{c}_{i,j}|$ decreases by 3.5dB. $\phi_{i,j}$ is modeled as uniformly distributed in $[0, 1[$ since a clear dependence with distance was not found. The multi-path components by an i.i.d zero-mean circularly symmetric complex Gaussian random variable with variance $\sigma^2 = 0.001$.

For the internal calibration, different antenna partition strategies are possible, where the optimal solution is yet to be discovered. In our simulation, we chose two different antenna partition scenarios: "two sides partition" and "interleaved partition" as shown in Fig. 6.7. The "two sides partition" separate the whole antenna array to group A and B on the left and right sides whereas the "interleaved partition" assigns every 8 antennas to A and B alternatively.



(a) Two sides partition



(b) Interleaved Partition

Figure 6.7: Two partitions. The above figure shows the "two sides partition" where group A and B contain 32 antennas on the left and right sides of the linear antenna array, respectively. The bottom figure illustrates the "interleaved partition" where every 8 antennas are assigned to group A and B, alternatively.

In the first simulation, we would like to verify the feasibility to calibrate a hybrid beamforming transceiver using internal calibration. For this purpose, we use the "two sides

[1] This term is called "antenna mutual coupling" in [28], which is slightly different from the classical mutual coupling defined in [39] where two nearby antennas are both transmitting or receiving. We thus call this term "near field path" describing the main signal propagation from one antenna to its neighbor element.

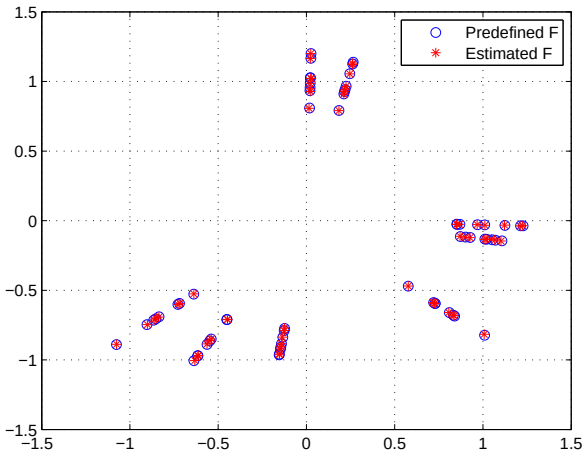


Figure 6.8: Estimated calibration matrix vs. real calibration matrix. The blue circles are predefined calibration coefficients and the red stars are estimated values after elimination of the complex scalar ambiguity.

partition" scenario and assume no noise in the bi-directional transmission between group A and B. We use 8 i.i.d Gaussian random variables as pilots after the baseband digital beamforming and only apply analog precoding whose weights have a unit amplitude, with their phases uniformly distributed in $[-\pi\ \pi[$. Using $K = 32$ and $L = 5$ such randomly generated transmit and receive analog beam weights to accumulate 160 measurements[2] and applying the method in Section 6.2.4 on the accumulated signal, we can obtain the estimated calibration coefficients. For the purpose of illustration, we eliminate the complex scalar ambiguity, the results are shown in Fig. 6.8.

We observe that the calibration matrix are partitioned in 8 groups, corresponding to 8 RF chains each with its own signal mixer. On each angle, elements with different amplitudes mainly correspond to the gain imbalance of independent amplifiers on each branch. We also observe that the estimated calibration parameters perfectly match the predefined values, implying that we can recover the coefficients using the proposed method. In a practical system, as no real value of $\mathbf{F}$ is known, all estimated coefficients have an ambiguity up to a common complex scalar value as explained in Section 6.2.4.

In the next simulation, we study the calibration performance with regard to the number of intra-array channel measurements. Since the measurements are within the antenna array, noise from both transmitter and receiver hardware can impact the received signal quality. For antennas near each other, the main noise source comes from the transmit signal, usually measured in error vector magnitude (EVM). Assuming a transmitter with an EVM of -20dB, the SNR of the transmit signal is 40dB. For antennas far away from each other, noise at the receive is the main limitation factor. Assuming that the system bandwidth is 5MHz, the thermal noise at room temperature would be -107dBm at the receiving antenna. Using a radio chain with a noise figure (NF) of 10dB and a total receive gain equaling to 0dB, the noise received in the digital domain would be around -97dBm. We assume a 0dBm transmission power per antenna and use the intra-array channel model

[2]Note that in a practical multi-carrier system, the channel estimation on different RF chains can be performed on different frequencies as explained in Section 6.2.3, the needed K can then be much less.

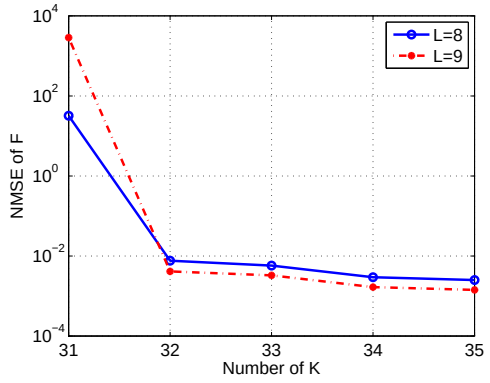


Figure 6.9: MSE of estimated calibration matrix vs. the number of K and L in the "two sides partition scenario". Both Tx and Rx noise are considered.

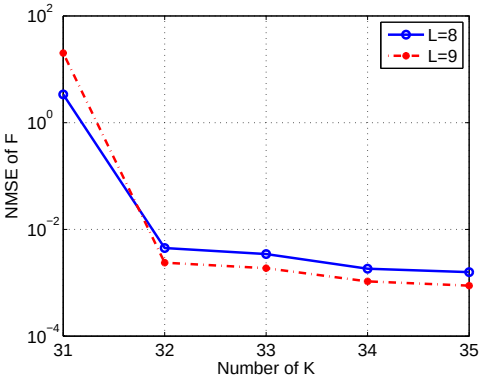


Figure 6.10: MSE of estimated calibration vs. the number of K and L in the "interleaved partition scenario". Both Tx and Rx noise are considered.

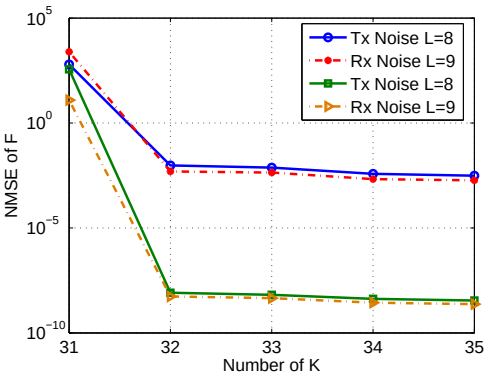


Figure 6.11: MSE of estimated calibration vs. the number of K and L in the "two sides partition scenario". Tx and Rx noise are simulated independently.

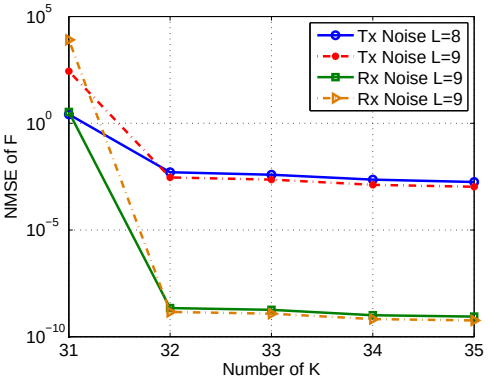


Figure 6.12: MSE of estimated calibration matrix vs. the number of K and L in the "interleaved partition scenario". Tx and Rx noise are simulated independently.

as in (6.19) with different K and L values, the calibrated coefficients are measured in its mean square error (MSE) normalized by the square of the norm of $\mathbf{F}$, such as

$$\text{NMSE}_{\mathbf{F}} = \frac{\|\hat{\mathbf{F}} - \mathbf{F}\|^2}{\|\mathbf{F}\|^2}. \tag{6.20}$$

The result using a unit norm after satisfying the "norm and phase" constraint is shown in Fig. 6.9 and Fig. 6.10 for "two sides partition" and "interleaved partition", respectively. We observe in both cases that, when $K < 32$, the estimation of $\mathbf{F}$ can not converge, since the intra-array channel estimation problem is under-determined, as explained in Section 6.2.3. As long as $K \geq 32$ and $L \geq 8$, it is possible to estimate $\mathbf{F}$ to an accuracy with an NMSE below 10^{-2}. "interleaved partition" has a better performance than "two sides partition" when the minimum K and L requirements are met. This can be explained by the fact that the received signals in "interleaved partition" have more balanced amplitudes than in "two sides partition", where, the bi-directional transmission between far away antenna elements

have very little impact on the estimation of $\mathbf{F}$ since the received signal are small. Note that different sets of transmit and receive analog precoding weights can lead to different performance in the estimation of $\mathbf{F}$, with the best set left to be discovered in future work. In our simulation, we randomly choose a set of weights and use it for both "two sides partition" and "interleaved partition". For comparison purpose, the set of weights for given K and L values (e.g $K = 32, L = 8$) is a subset for the weights used when K and L are bigger (e.g $K = 33, L = 9$).

Since we simulate the intra-array transmission, both the transmit and receive noise have been taken into account. In order to understand the impact from the two noise source, let us simulate for them independently under both antenna partition scenarios. Fig. 6.11 and Fig. 6.12 illustrate the NMSE of $\mathbf{F}$ with independently considered noise for "two sides partition" and "interleaved partition", respectively. It is obvious that, in both cases, the noise at the transmit side is dominant and limits the accuracy of the estimated $\mathbf{F}$ whereas if only the receiver's thermal noise is considered, NMSE of $\mathbf{F}$ approaches machine precision. In fact, if we look back at (6.11), it is the errors present in the bi-directional channel estimation h_i and h_j with the highest amplitudes (i.e. internal channels between nearby antenna elements) that dominate the cost function. For a receiving antenna near the transmitting element, the received transmit noise is much higher than the thermal noise generated at the receiving antenna itself.

When the system has accomplished internal calibration, it can use the estimated calibration matrix together with the instantaneously estimated UL channel to assess the DL CSIT in order to create a beam for data transmission. The accuracy of the DL CSIT depends on both the UL CSI and the estimated calibration matrices. In order to study the impact of both factors, we assume a simple scenario where a subarray hybrid structure BS performs beamforming towards a single antenna UE. In this case, the DL channel $\mathbf{h}_{BS\to UE}^T$ (we use transpose since the DL channel is a row vector) can be estimated by $\hat{f}_{UE}^{-1}\hat{\mathbf{h}}_{UE\to BS}^T\hat{\mathbf{F}}_{BS}$, where $\hat{\mathbf{h}}_{UE\to BS}$ is the estimated UL channel. $\hat{\mathbf{h}}_{UE\to BS} = \mathbf{h}_{UE\to BS} + \Delta\mathbf{h}_{UE\to BS}$, where $\Delta\mathbf{h}_{UE\to BS}$ is the UL channel estimation error, $\mathbf{h}_{UE\to BS} = \mathbf{R}_{BS}\mathbf{c}t_{UE}$, with the UL physical channel vector $\mathbf{c}$ modeled as a standard Rayleigh fading channel. $\hat{f}_{UE}^{-1}$ and $\hat{\mathbf{F}}_{BS}$ are the calibration coefficients for both UE and BS. Let us use $\mathbf{F} = f_{UE}\mathbf{F}_{BS}$ to denote the calibration matrix including coefficients from both UE and BS. Its estimation $\hat{\mathbf{F}}$ can be represented by $\hat{\mathbf{F}} = \mathbf{F} + \Delta\mathbf{F}$ with $\Delta\mathbf{F}$ denoting the estimation error. The estimation errors in $\Delta\mathbf{h}_{UE}$ and $\Delta\mathbf{F}$ are assumed to be i.i.d Gaussian random variables with zero mean and $\sigma_{n,UL}^2$, $\sigma_{\mathbf{F}}^2$ as their variance, respectively. $\mathrm{NMSE}_{\mathbf{F}}$ can be calculated as $N_{ant}^{BS}\sigma_{\mathbf{F}}^2/\|\mathbf{F}\|^2$. Without considering the complex scalar ambiguity, which does not harm the finally created beam, we can calculate the NMSE of the DL CSI as

$$\begin{aligned}
\mathrm{NMSE}_{DL} =& \frac{1}{N_{ant}^{BS}}\mathbb{E}\left[\|\hat{\mathbf{h}}_{UE\to BS}^T\hat{\mathbf{F}} - \mathbf{h}_{BS\to UE}^T\|^2\right] \\
=& \frac{1}{N_{ant}^{BS}}\mathbb{E}\left[\|\mathbf{h}_{UE\to BS}^T\Delta\mathbf{F} + \Delta\mathbf{h}_{UE\to BS}^T\hat{\mathbf{F}}\|^2\right] \\
=& \frac{1}{N_{ant}^{BS}}\mathrm{Tr}\left\{\Delta\mathbf{F}^H\mathbf{\Omega}^*\Delta\mathbf{F} + \sigma_{n,UL}^2\hat{\mathbf{F}}^H\hat{\mathbf{F}}\right\}
\end{aligned} \tag{6.21}$$

where $\mathbf{\Omega}$ is the covariance matrix of the UL channel, i.e. $\mathbf{\Omega} = \mathbb{E}[\mathbf{h}_{UE\to BS}\mathbf{h}_{UE\to BS}^H]$.

The NMSE of the calibrated CSIT on different $\mathrm{NMSE}_{\mathbf{F}}$ and NMSE_{UL}[3] is shown in

[3] $\mathrm{NMSE}_{UL} = \frac{1}{N_{ant}^{BS}}\mathbb{E}\left[\|\Delta\mathbf{h}_{UE\to BS}\|^2\right] = \sigma_{n,UL}^2$.

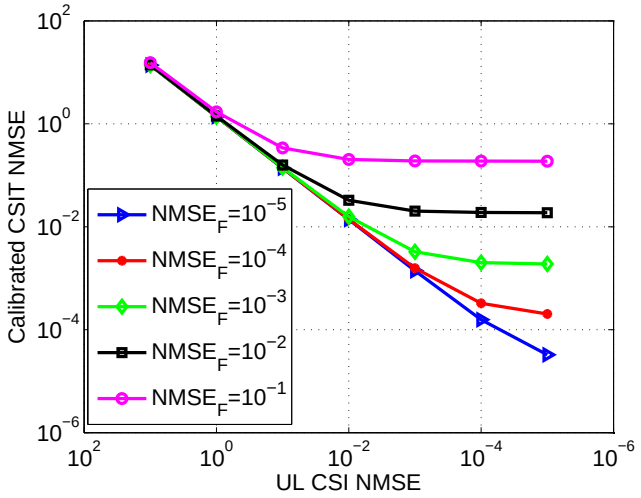


Figure 6.13: The accuracy of acquired CSIT as a function of the accuracy of the reciprocity calibration matrix and instantaneously measured UL CSI.

Fig. 6.13. We observe that when the accuracy of the UL CSI is low, it is the main limiting factor on the calibrated DL CSIT. As the UL CSI accuracy increases, the accuracy on $\hat{\mathbf{F}}$ begins to influence the DL CSIT. In a calibrated system where $\text{NMSE}_{\mathbf{F}} = 10^{-2}$ and $\text{NMSE}_{UL} = 10^{-2}$, it is possible to have almost perfect DL CSIT with an NMSE under 10^{-1}.

6.5 Summary

We presented a CSIT acquisition method based on reciprocity calibration in a TDD hybrid beamforming massive MIMO system. Compared to state-of-the-art methods which assume a certain structure in the channel such as the limited scattering property validated only in mmWave, this method can be used for all frequency bands and arbitrary channels. Once the TDD system is calibrated, accurate CSIT can be directly obtained from the reverse channel estimation, without any beam training or selection. It thus offers a new way to operate hybrid analog and digital beamforming systems.

Chapter 7

Calibration Parameter Tracking Allowing Fast Correction of Sudden Variations

The existing calibration schemes, as well as our previously proposed methods, are mainly for the objective of tuning a totally uncalibrated system. For a BS continuously providing data services, such calibration procedures needs to be repetitively performed at a certain pace. Although, in theory, the calibration parameters stay quite stable during the time, as the number of antennas becomes large, the time interval between two consecutive calibrations tends to be shorter if we would like to ensure the whole system to be always under calibrated status. At each calibration, resources need to be allocated to estimate the hardware asymmetry, the normal function of data service would be unavoidably influenced or even interrupted. It is thus beneficial to come up with a calibration parameter tracking algorithm, which can fully make use of the previous calibration information to render the current calibration light and easy to do. Such a tracking scheme can be considered as a system health monitoring mechanism which observes the calibration status during the time, and once detects a sudden variation, can quickly correct the changed parameter.

State of the art work on this topic can be found in [48,70], where under the assumption that the majority of the calibration parameters remain unchanged, the vector representing the coefficients' difference between two consecutive calibrations enjoys sparsity. The author thus suggests to use compressive sensing techniques to recover the changed parameter. By assuming a certain number of changed parameters, it is possible to perform bi-directional transmission with the assistance of a UE or a reference antenna using a set of virtual antenna beams on the order of $\mathcal{O}(\log(M))$, where M represents the number of antennas, to accumulate a few observations and recover the changed parameter. This method allows less usage of resources for consecutive calibrations. However, the error recovery is based on the assumption that the number of changed parameters are predefined, and as long as the number of real changed parameters are larger than the predefined one, the recovery can not be achieved. Moreover, as the channel is embedded into the vector representing the parameter change, this method is not robust enough against noise and small variations for scenarios where channels between BS and UE have very different amplitudes, especially in distributed massive MIMO cases.

We propose, in this chapter, a calibration parameter tracking method, which evolves the

estimated parameters together with the real calibration coefficients[1]. The method is suitable for both co-located and distributed massive MIMO and can be easily adapted to almost all existing initial calibration processes. It can avoid the whole calibration process if all calibration parameters stay unchanged during two consecutive calibrations, which can happen quite often. When some parameters changed, we use a binary tree search algorithm to detect the changed parameter and correct them rather than re-calibrate the whole system.

The remainder of this chapter is organized as follows: Section 7.1 presents two types of changes in calibration parameters. Section 7.2 describes our proposed calibration parameter tracking algorithm. Section 7.3 illustrates the simulation results and we draw the conclusion in Section 7.4.

7.1 Evolution of calibration parameters

Calibration parameters represent the RF hardware properties and are independent of the physical channel in the air. Some measurement results in [25] show that these parameters stay quite stable during the time. In order to observe how calibration parameters evolve with the time, we also carried out a measurement on a SISO TDD system built up with two cable-synchronized ExpressMIMO2 cards on EURECOM's OpenAirInterface platform. The whole measurement lasts 62 hours and every 2 minutes we activated the Argos calibration parameter estimation algorithm. The result of the experiment is illustrated in Fig. 7.1, where the full and dotted line represent the deviation of magnitude and angle respectively. The angle deviation is calculated as the difference with the mean angle value normalized by π, same as in [25]. We observe that the calibration parameter stays quite stable during a relatively long time with very small deviation and then a sudden sporadic change can appear in the angle and the a new stable state can be established. Note that such sporadic changes have been observed several times after repetition of the same experiment.

The observations in Fig. 7.1 indicates two types of variations:

- **Smooth variations** which usually varies over several minutes and especially due to temperature changes. Experimental measurements on the relationship of hardware parameters and temperature can also be found in [71].

- **Sporadic sudden variations** that may be caused by hardware glitches. In reality, a practical system will always experience phenomenons such as power surge or interruption, as well as lost of synchronization because of various reasons. Such "incidents" happening in hardware can cause sporadic sudden variations as observed in Fig. 7.1.

The object of this work is to propose a calibration tracking scheme addressing these two types of variation.

[1]This chapter is a joint work with Alexis Decunringe and Maxime Guillaud from Huawei Technologies (Paris)

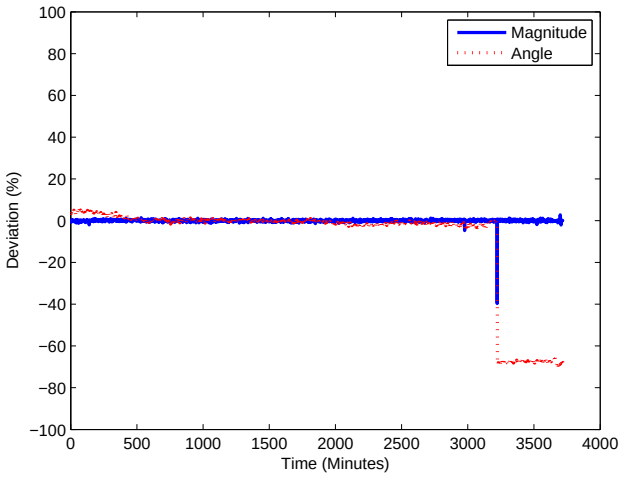


Figure 7.1: Deviation of calibration parameter during 62 hours.

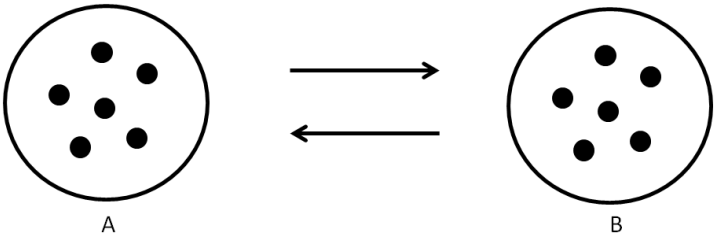


Figure 7.2: Illustration of the division of the antenna array into two groups A and B of antenna elements.

7.2 Calibration parameter tracking

Let us consider a massive MIMO system with M antennas and model the transmit and receive RF front-ends on the given carrier using $\mathbf{T} = \text{diag}\{t_1, t_2, \ldots, t_M\}$ and $\mathbf{R} = \text{diag}\{r_1, r_2, \ldots, r_M\}$ respectively. Our proposed calibration tracking scheme lies in two principles:

- **Bi-directional measurements between groups of antenna elements:** we partition the whole antenna array into two groups, denoted by A and B, each with N_A and N_B antennas, as in Fig. 7.2. Bi-directional measurements between these two subgroups are then transmitted over the air:
  - Antennas of group A jointly send a pilot; all antennas of group B are listening.
  - Antennas of group B jointly send a pilot; all antennas of group A are listening.
- **Two modes of calibration:** a normal mode where bi-directional measurements are exchanged between the two subgroups of antennas (in a coherent or a non-coherent manner) and an urgent mode performed when a sudden variation in the calibration parameters is detected. Dedicated signals are then sent in order to correct incriminated calibration parameters.

In the sequel, we detail the calibration tracking algorithm for the two modes in a top-down manner. Note that signal exchanges between BS antenna elements, as well as the

tracking algorithms related to these signals are independent from the mode framework. In particular, we should be able to use the signal exchanges proposed in normal mode without ever switching to urgent mode. In that case, the hierarchic two mode structure does not exist and the system always stays at the normal mode.

7.2.1 Inter-operation between two calibration modes

Let us firstly look at how the two calibration modes inter-operate with each other:

- **Normal mode:** bi-directional measurements between A and B are performed at different time slots, as in Fig. 7.3. At each time slot, an estimation of calibration coefficients is performed if enough bi-directional measurements have been accumulated and a metric η is calculated from the received signals and the previously estimated calibration parameters. We then compare η with a threshold χ_0. If this metric is too high, the system switches to urgent mode, as in Fig. 7.4.
- **Urgent mode:** in this mode, a dedicated sequence of signals is sent and calibration parameters are corrected after L_u bi-directional measurements have been exchanged, where L_u depends on how many measurements are required to identify the uncalibrated antenna(s). After L_u bi-directional measurements, the system goes back to the normal mode, as shown in Fig. 7.4.

The objective of the normal mode is to 1) evolving the estimation of calibrated parameters in line with its smooth evolution; 2) detect any loss of calibration within the system as early as possible and in a spectrally efficient manner (requiring few pilot signal transmissions). This mode however does not always enable to identify the (one or multiple) uncalibrated antennas. The objective of the urgent mode is, after a detection of loss of calibration, we identify as quickly as possible the (one or several) uncalibrated antenna(s), and to re-calibrate them.

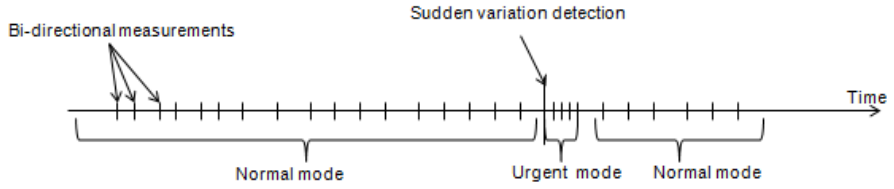


Figure 7.3: Illustration of two calibration modes.

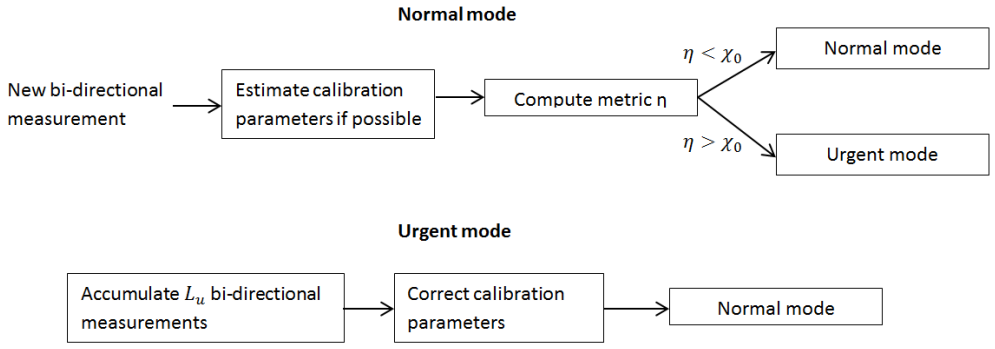


Figure 7.4: Transition between the normal and urgent mode.

7.2.2 Normal mode

In normal mode, bi-directional measurements are iteratively exchanged between two antenna groups A and B. The channel may be considered as constant during a given time interval where we can accumulate L_c coherent consecutive measurements. We then perform non-coherent accumulation for T_{NC} such intervals as described in Section. 5.5 and illustrated in Fig. 7.5.

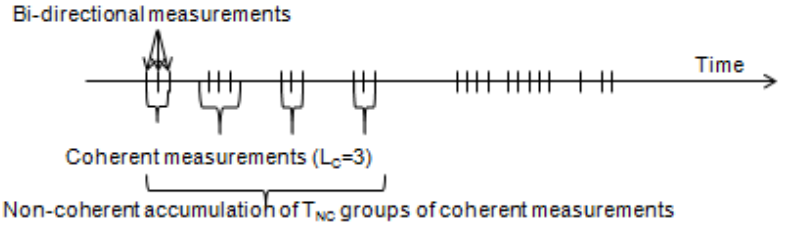


Figure 7.5: Non-coherent accumulation of groups of coherent signals.

If enough accumulations have been exchanged in the past, i.e. $T_{NC} \geq \frac{M}{L_c^2}$, an estimation of calibration parameters is performed.

Consider $\mathbf{P}_{A,t}$ and $\mathbf{P}_{B,t}$ as the pilot matrices of group A and B with size $M_A \times L_c$ and $M_B \times L_c$, respectively, transmitted during non-coherent slot t, where $t \in \{1, 2, \ldots, T_{NC}\}$ with T_{NC} representing the total non-coherent time slots. For each coherent time slot l, where $l \in \{1, 2, \ldots, L_c\}$, A and B exchange the l^{th} column of the their pilot matrices $\mathbf{P}_{A,t}$ and $\mathbf{P}_{B,t}$. The received matrices during time slot t are denoted as $\mathbf{Y}_{B\to A,t}$ and $\mathbf{Y}_{A\to B,t}$ with size $M_A \times L_c$ and $M_B \times L_c$, respectively. We use $\mathbf{F}_{A,t}$ and $\mathbf{F}_{B,t}$ to denote the calibration matrices at time slot t, i.e. the diagonal calibration matrix $\mathbf{F}_t$ is composed of $\mathbf{F}_{A,t}$ and $\mathbf{F}_{B,t}$, such as

$$\mathbf{F}_t = \begin{bmatrix} \mathbf{F}_{A,t} & 0 \\ 0 & \mathbf{F}_{B,t}. \end{bmatrix} \tag{7.1}$$

Same as the observation under the general calibration framework in (5.2), the reciprocity property can be expressed by

$$\mathbf{P}_{A,t}^T \mathbf{F}_{A,t}^T \mathbf{Y}_{B\to A,t} = \mathbf{Y}_{A\to B,t}^T \mathbf{F}_{B,t} \mathbf{P}_{B,t} + \tilde{\mathbf{N}}_{A,B}, \tag{7.2}$$

with $\tilde{\mathbf{N}}_{A,B}$ representing some noise sourcing from the received signal. Assuming the calibration matrix stays unchanged during T_{NC} non-coherent time slot, we can thus temporarily drop the time index on $\mathbf{F}_A$ and $\mathbf{F}_B$. Similar as(5.40), the estimated calibration parameters at time slot t can be obtained by solving the following LS problem

$$\begin{aligned} \hat{\mathbf{f}}_t &= \arg\min_{\mathbf{f}_A, \mathbf{f}_B} \sum_{t=1}^{T_{NC}} \left\| (\mathbf{Y}_{B\to A,t}^T * \mathbf{P}_{A,t}^T)\mathbf{f}_A - (\mathbf{P}_{B,t}^T * \mathbf{Y}_{A\to B,t}^T)\mathbf{f}_B \right\|^2, \\ &= \arg\min_{\mathbf{f}} \|\boldsymbol{\mathcal{Y}}(\mathbf{P})\mathbf{f}\|^2. \end{aligned} \tag{7.3}$$

where $\mathbf{f}_t$, $\mathbf{f}_A$, $\mathbf{f}_B$ are the diagonal vector of $\mathbf{F}_t$, $\mathbf{F}_A$ and $\mathbf{F}_B$, respectively, $\boldsymbol{\mathcal{Y}}(\mathbf{P}) = [\boldsymbol{\mathcal{Y}}_1(\mathbf{P}_1)^T, \ldots, \boldsymbol{\mathcal{Y}}_{T_{NC}}(\mathbf{P}_{T_{NC}})^T]^T$ with $\boldsymbol{\mathcal{Y}}_t(\mathbf{P}_t)^T = [\mathbf{Y}_{B\to A,t}^T * \mathbf{P}_{A,t}^T \ - \mathbf{P}_{B,t}^T * \mathbf{Y}_{A\to B,t}^T], t \in \{1, 2, \ldots, T_{NC}\}$. The minimization is calculated subject to $\|\mathbf{f}\| = 1$ or$f_1 = 1$. As the cost function in (7.3) takes into account the measurement of a quite long time duration, when $\mathbf{f}$ experiences

some smooth variation as described in section 7.1, the estimated calibration matrix can gradually update its value, thus following the smooth evolution of the real $\mathbf{f}$.

When A and B exchange pilots, $\mathbf{P}_{A,t}$ and $\mathbf{P}_{B,t}$ can be constructed as follows. Each element on the i^{th} row and j^{th} column, denoted as $p_{A_{i,j},t}$ and $p_{B_{i,j},t}$ can be randomly chosen with a constant amplitude p_0 and a phase, noted by $\theta_{i,j}$ or $\phi_{i,j}$ uniformly distributed in $]-\pi\ \pi]$, given as

$$\begin{aligned} p_{A_{i,j},t} &= p_0 e^{i\theta_{i,j}} \quad \text{where } 1 \leq i \leq M_A,\ 1 \leq j \leq L_c, \\ p_{B_{i,j},t} &= p_0 e^{i\phi_{i,j}} \quad \text{where } 1 \leq i \leq M_B,\ 1 \leq j \leq L_c \end{aligned} \tag{7.4}$$

After each non-coherent time slot t, we should judge if the $\mathbf{f}$ has experienced a sporadic sudden variation. This can be done by calculating the LS metric $\|\boldsymbol{\mathcal{Y}}_t(\mathbf{P_t})\mathbf{f}_{t-1}\|^2$, where we use the pilots and accumulated observations at time t with the previously estimated $\mathbf{f}$ at $t-1$. Let us first ignore the noise, if we assume $\mathbf{f}$ stays unchanged, i.e., $\mathbf{f}_t = \mathbf{f}_{t-1}$, then $\|\boldsymbol{\mathcal{Y}}_t(\mathbf{P}_t)\mathbf{f}_{t-1}\|^2 = 0$, otherwise, the changes in $\mathbf{f}$ will lead this metric to be non-zero. Thus evaluating whether the metric is zero is a basic criterion to detect if $\mathbf{f}$ has changed. When noise is considered, we should define a threshold, such that when the calculated metric is larger than it, the urgent mode is triggered. Assume $\mathbf{f}_t = \mathbf{f}_{t-1}$, from (5.34), we have

$$\boldsymbol{\mathcal{Y}}_t(\mathbf{P}_t)\mathbf{f}_{t-1} = \boldsymbol{\mathcal{F}}_{t-1}^{\perp H}\mathbf{y}_t = \widetilde{\mathbf{n}} \tag{7.5}$$

where

$$\boldsymbol{\mathcal{F}}_{t-1}^{\perp} = \begin{bmatrix} \mathbf{I} \otimes (\mathbf{F}_{B,t-1}\mathbf{P}_{B,t})^* \\ -(\mathbf{F}_{A,t-1}\mathbf{P}_{A,t})^* \otimes \mathbf{I} \end{bmatrix}. \tag{7.6}$$

$\widetilde{\mathbf{n}}$ is a colored noise of covariance matrix $\sigma^2\boldsymbol{\mathcal{F}}_{t-1}^{\perp H}\boldsymbol{\mathcal{F}}_{t-1}^{\perp}$ as explained in Section 5.4.3. By whitening and normalizing the noise as we've done in (5.36), we can obtain a weighted metric such as

$$\eta_t = \frac{2}{\sigma^2}\mathbf{y}_t^H\boldsymbol{\mathcal{F}}_{t-1}^{\perp}(\boldsymbol{\mathcal{F}}_{t-1}^{\perp H}\boldsymbol{\mathcal{F}}_{t-1}^{\perp})^{\dagger}\boldsymbol{\mathcal{F}}_{t-1}^{\perp H}\mathbf{y}_t = \frac{2}{\sigma^2}\mathbf{y}_t^H\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{F}}_{t-1}}^{\perp}\mathbf{y}_t = \frac{2}{\sigma^2}\|\mathbf{n}\|^2. \tag{7.7}$$

η_t follows a $\boldsymbol{\chi^2}$ distribution with $2M$ degrees of freedom.

Detection of the changes in calibration parameters can be considered as a hypothesis testing problem. Considering the null hypothesis (H_0) : "calibration parameters stay the same" and the alternative hypothesis (H_1) : "some of the parameters changed", we can choose the threshold such that the type I error, noted by α, is fixed at a certain value:

$$\Pr(\eta_t > \chi_0 | H_0) = \alpha. \tag{7.8}$$

Note that in our case, in order to have a small type II error (some parameters indeed changed but the algorithm fails to detect them), α should not be too small.

7.2.3 Urgent mode

In case of $\eta_t > \chi_0$, the system is switched to the urgent mode, there is a high probability that the system is uncalibrated. To re-calibrate the system, there are two possibilities. The basic solution is to accumulate $L_u \geq \sqrt{M}$ coherent consecutive bi-directional measurements, similar as detailed in the normal mode. We then estimate the calibration parameters using the same equations as in (7.3), with $T_{NC} = 1$. After this estimation, the system goes back to the normal mode. This method is suitable to cases where the channel remains constant over the L_u measurements.

A better solution would be quickly find the uncalibrated antennas and correct them. This solution can be further divided into a two step procedure:

- A binary tree search process involving $L_b \leq L_u$ bi-directional measurements in order to identify uncalibrated antennas and compute corrected calibration parameters (detailed hereafter),
- $L_u - L_b$ further bi-directional measurements exchanged between two groups of A and B. If possible, an estimation of calibration parameters is performed after each bi-directional measurement as detailed in normal mode. The system then goes back to normal mode after a total number of L_u bi-directional measurements, as shown in Fig. 7.6.

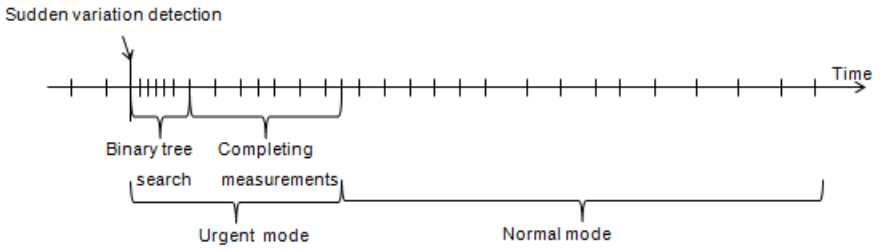


Figure 7.6: Illustration of the two steps solution in urgent mode.

The binary tree search is performed as illustrated in Fig. 7.7 (with the threshold optimized off-line):

- During the normal mode, at each non-coherent time slot, a metric η (the subscript t is dropped for simplicity) is calculated based on the coherent bi-directional measurement between A and B as in (7.7). If $\eta > \chi_0$, the urgent mode is triggered and we start the binary tree search. Consider this step as the initializations step of the whole binary tree search method and denote A and B as A_0^1 and B_0^1, where A_k^d represent the antenna group A in the d^{th} division of the k^{th} iteration, detailed hereafter.
- Each group A_0^1 and B_0^1 is further divided into two subgroups. Group A_0^1 is divided into two subgroups A_1^1 and B_1^1. $L_{b,1}^1$ coherent bi-directional channel measurements are performed between these two subgroups and a corresponding metric η_1^1 is computed. If $\eta_1^1 < \chi_1^1$, where χ_1^1 is the corresponding threshold, we do not divide A_1^1 any further. Otherwise, A_1^1 and B_1^1 will then also be divided into two subgroups each.
- In the same way, B_0^1 is divided into two subgroups A_1^2 and B_1^2 and $L_{b,1}^2$ coherent bi-directional measurements are exchanged between them leading to the computation of a criterion η_1^2. The same decision for a further division of A_1^2 and B_1^2 is performed.
- The process is iterated until each remaining group contains only one element. During the whole process, $k \in \{0, 1, 2, \cdots, \log_2 M - 1\}$, $m \in \{1, \cdots, 2^k\}$.

Note that for each iteration k and division d, we can chose a different threshold χ_k^d.

The complexity of the binary tree search algorithm can be measured by the number of bi-directional transmission needed, which goes up with the increase of the number of changed parameters. In case where only one parameter has experienced an sporadic sudden change, $2\log_2(M) - 1$ bi-directional transmissions are needed. If all parameters lost the calibrated

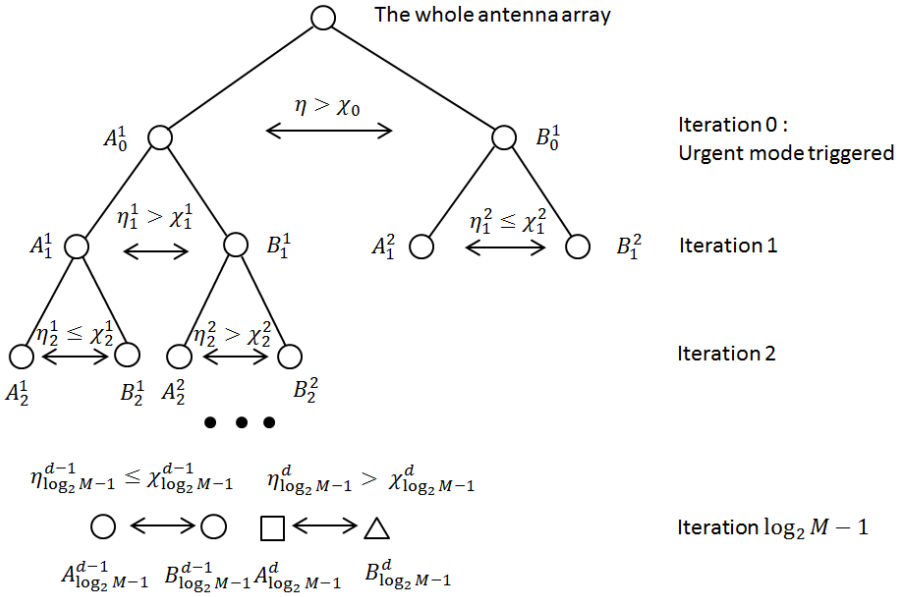


Figure 7.7: Illustration of binary tree search.

status, then $M^2/8 + M/4$ bi-directional transmissions are needed. For the number of changed parameters in between, the complexity depend on how they distribute in the groups we classify. For example, if two parameters changed, we need $2\log_2(M) - 1$ bi-direction transmission in the luckiest case and $4\log_2(M) - 5$ in the worst case.

7.3 Simulation results

In this section, we present the simulation results of our calibration parameter tracking method in a 64 co-located antenna array. The diagonal elements in $\mathbf{T}$ and $\mathbf{R}$ are modeled as i.i.d. random variables, with uniformly distributed phase between $[-\pi \ \ \pi]$ and independent magnitude uniformly distributed on $[1-\epsilon \ \ 1+\epsilon]$, with ϵ chosen such that the standard deviation of the squared-magnitudes is 0.1, as in [51]. A simplified standard Rayleigh channel $\mathcal{CN}(0, \mathbf{I})$ with the same SNR is assumed between any two antenna elements. Using the same SNR can be justified by the fact that the during the reciprocity calibration, it is mainly the noise at the transmitter side which limits the calibration performance, as we have observed in Section. 6.4.

In the simulation, we use a Fourier pilot with a unit norm ($p_0 = 1$) and phases of its elements uniformly distributed in $[-\pi \ \ \pi]$. The number of coherent accumulation L_c is 1. During two consecutive non-coherent accumulation at time t_0 and t_1, we assume that only one parameter suffers a random phase change whereas all other parameters stay unchanged (smooth variations are ignored). Thus if the algorithm successfully detects that the systems is out of calibration, it switches to the urgent mode at time t_1. The type I error α is fixed during the the whole binary tree searching process, according to which we can calculate a dynamic threshold χ_k^d. We perform two simulations with $\alpha = 0.1$ and $\alpha = 0.4$, respectively, and the results are shown in Fig. 7.8-7.10.

We observe that when α is set to be larger, the probability of not successfully detecting the changed parameter (type II error) is smaller whereas the number of stable parameters being

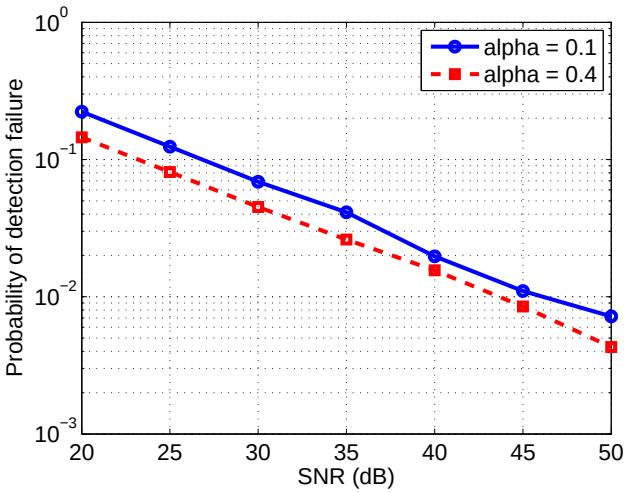


Figure 7.8: Probability of failing to detect the changed parameter (type II error).

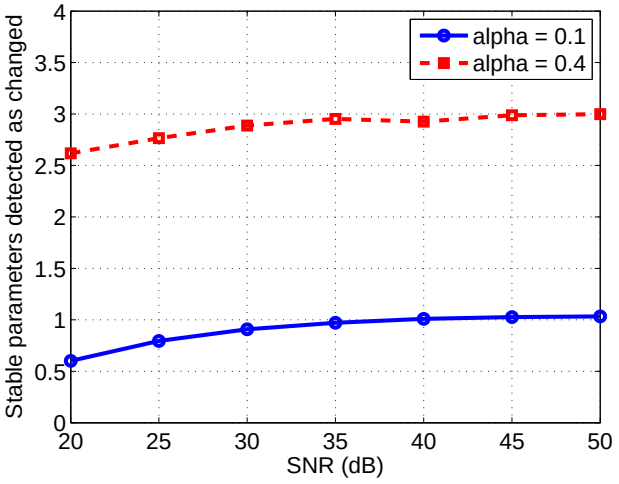


Figure 7.9: Number of stable calibration parameters detected as changed (type I error).

wrongly detected as changed parameters (type I error) are higher. This is reasonable since α represents the probability of type I error. The threshold is bigger when α is smaller, thus less bi-directional transmissions are needed to accomplish the bi-directional tree search. With an α set to 0.4, at the SNR of 40dB (SNR between two antenna element in the same array), we can find the changed parameter with a success probability above 98%. The number of needed bi-directional transmission is 28. In practice, we should choose the α according to an acceptable type II error.

7.4 Summary

We propose an easy and fast reciprocity calibration parameter tracking method which can unleash massive MIMO TDD systems from the repetitive resource consuming calibration process. The tracking method, in its normal mode alone, is more spectrally efficient than methods in [25,26] (i.e. requiring less pilot exchange) to maintain the system under a cal-

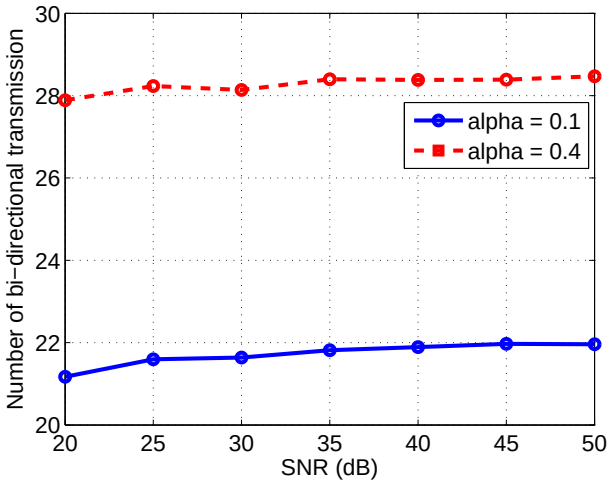


Figure 7.10: Number of bi-directional transmission used to detect the random phase change.

ibrated status, by introducing coherent and non-coherent accumulation of bi-directional measurements between two antenna groups. With the inter-operation between normal and urgent mode, the calibration tracking scheme makes the system more robust against hardware glitches by allowing fast detection and correction of sporadic sudden variations of calibration parameters. If all parameters stay unchanged, which is often the case in practice, the tracking algorithm can make the system avoid the whole process of re-calibration.

Chapter 8

Reciprocity Calibration Coefficients Measurements

A widely adopted assumption in literature, e.g., in [21, 22, 25, 28], is that the crosstalk between different RF chains and the antenna mutual coupling effect can be ignored. This assumption was previously mentioned in Section. 4.1 in this thesis and has greatly helped us to simplify the calibration of calibration matrix. However studies on hardware impairment show that these effects do exit. For the former, the authors in [72] provided a RF crosstalk model for two interacting chains and studied its impact on the MIMO performance whereas for the latter, authors in [19] had an intensive study and showed the relationship between the antenna mutual coupling matrix and its impedance matrix; besides authors in [41] show that the transmit and receive antenna mutual coupling matrices are not reciprocal in most cases. It is thus doubtful in the research community whether the relative calibration matrix could really be assumed to be diagonal and the performance degradation arising therefrom is unknown.

To address this problem, in this chapter, we carry out measurements of the calibration matrix on EURECOM's OpenAirInterface platform and compare the beamforming performance based on different CSIT acquisition methods. We estimate the full matrix, including the RF chain crosstalk and antenna mutual coupling effect, which has never been done before. The real-world results from experiment do not only verify the assumption on the calibration matrix structure but also provide a direct insight on various other phenomenon on the transceiver hardware and can thus be a useful support for TDD reciprocity hardware impairment modeling and relevant theoretical analysis.

8.1 System model

In the measurement, we consider a MISO system as illustrated in Fig. 8.1. Node A and B are equipped with M_A antennas and 1 antenna, respectively. The forward and reverse links between A and B are represented by $\mathbf{h}_{A\to B}^T$ (transpose since the channel from A to B is a row vector) and $\mathbf{h}_{B\to A}$. The relationship between them is given by

$$\mathbf{h}_{A\to B}^T = \mathbf{h}_{B\to A}^T \mathbf{F}, \tag{8.1}$$

where the calibration matrix $\mathbf{F} = f_B^{-1}\mathbf{F}_A = \frac{r_B}{t_B}\boldsymbol{R}_A^{-T}\boldsymbol{T}_A$ includes all the hardware properties on both sides .

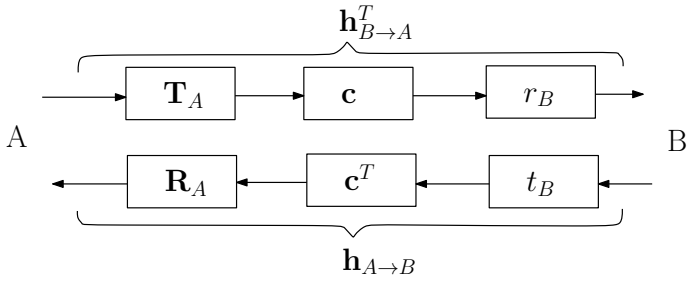


Figure 8.1: Reciprocity Model of an $M_A \times 1$ MISO system.

8.2 Full estimation of the calibration matrix

In order to verify the diagonal assumption on the calibration matrix, a method able to estimate the full calibration matrix, including both diagonal and off-diagonal elements is needed. This can be accomplished by our unified framework in Chapter 5, where for the sake of simplicity, we explored it under the diagonal assumption. Recalling (5.2) in the form of two groups A and B under the given MISO system:

$$\mathbf{P}_A^T\mathbf{F}_A^T\mathbf{Y}_{B\to A} - \mathbf{y}_{A\to B}f_B\mathbf{p}_B^T = \widetilde{\mathbf{N}}_{A,B}, \tag{8.2}$$

as $\mathbf{F}_A$ is no longer considered as diagonal, (5.4) becomes

$$(\mathbf{Y}_{B\to A}^T \otimes \mathbf{P}_A^T)\mathrm{vec}(\mathbf{F}_A^T) - (\mathbf{p}_B \otimes \mathbf{y}_{A\to B})f_B = \widetilde{\mathbf{n}}_{AB}. \tag{8.3}$$

The full estimation of $\mathrm{vec}(\mathbf{F}_A)$ is given by

$$\mathrm{vec}(\hat{\mathbf{F}}_A^T) = (\mathbf{Q}_A^H\mathbf{Q}_A)^{-1}\mathbf{Q}_A^H\mathbf{q}_B f_B \tag{8.4}$$

where $\mathbf{Q}_A = \mathbf{Y}_{B\to A}^T \otimes \mathbf{P}_A^T$ and $\mathbf{q}_B = \mathbf{p}_B \otimes \mathbf{y}_{A\to B}$. By giving f_B a predefined value, e.g., 1, we can estimate the full matrix of $\mathbf{F}_A$ w.r.t f_B.

However, in the place of using the above LS estimator, we adopt in this measurement campaign the TLS introduced in [21, 22], which also allows for the estimation of full $\mathbf{F}$[1]. This method performs the estimation of $\mathbf{F}$ directly on the bi-directional channel estimation and takes into account the errors on both directions. Let us consider K pairs of such estimation vectors organized in matrices

$$\begin{cases} \tilde{\mathbf{H}}_{A\to B} = \left[\hat{\mathbf{h}}^1_{A\to B}, \hat{\mathbf{h}}^2_{A\to B}, \ldots, \hat{\mathbf{h}}^K_{A\to B}\right]^T, \\ \tilde{\mathbf{H}}_{B\to A} = \left[\hat{\mathbf{h}}^1_{B\to A}, \hat{\mathbf{h}}^2_{B\to A}, \ldots, \hat{\mathbf{h}}^K_{B\to A}\right]^T, \end{cases} \tag{8.5}$$

where $K > M_A^2$. The TLS problem for $\mathbf{F}$ estimation is given by

$$\begin{aligned} \hat{\boldsymbol{F}} = & \underset{\Delta\mathbf{H}_{B\to A}, \Delta\mathbf{H}_{A\to B}, \boldsymbol{F}}{\arg\min} \|\Delta\mathbf{H}_{B\to A}\|_{\mathrm{F}}^2 + \|\Delta\mathbf{H}_{A\to B}\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & \tilde{\mathbf{H}}_{A\to B} + \Delta\mathbf{H}_{A\to B} = (\tilde{\mathbf{H}}_{B\to A} + \Delta\mathbf{H}_{B\to A})\boldsymbol{F}, \end{aligned} \tag{8.6}$$

where $\Delta\mathbf{H}_{A\to B}$ and $\Delta\mathbf{H}_{B\to A}$ are the corrections applied to the estimated values and $\|\cdot\|_{\mathrm{F}}$ is Frobenius norm.

[1]The main reason for using TLS method is that the measurement campaign is performed before the discovery of the general framework. However, without going into detail, we point out that the TLS approach corresponds to a ML estimation in Section 5.4.2 (with the pilots being identity matrices), considering the fact that it minimizes the sum of estimation error on all bi-directional channel estimations.

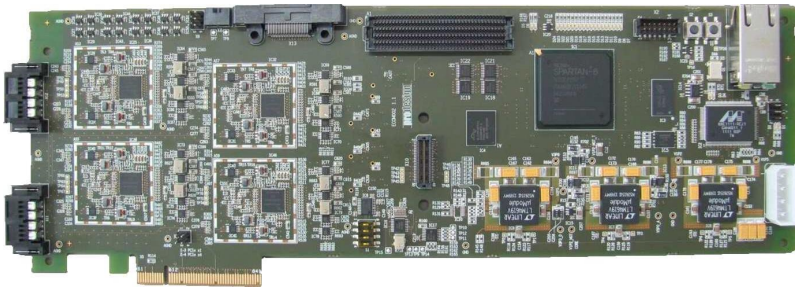


Figure 8.2: ExpressMIMO2 board.

A classical method for solving the TLS problem is based on singular value decomposition (SVD) [73]. Let $\boldsymbol{D} = \begin{bmatrix}\tilde{\mathbf{H}}_{B\to A} & \tilde{\mathbf{H}}_{B\to A}\end{bmatrix}$, the SVD algorithm in complex domain gives

$$\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H, \tag{8.7}$$

where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, ..., \sigma_{2M_A})$ is composed of the singular values of $\boldsymbol{D}$ and $\sigma_1 \geq ... \geq \sigma_{2M_A}$. Write $\boldsymbol{V}$ in a block matrix representation as

$$\boldsymbol{V} = \begin{bmatrix}\boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22}\end{bmatrix} \tag{8.8}$$

with $\boldsymbol{V}_{ij}$ $(i, j = 1, 2)$ being $M_A \times M_A$ matrices. The sufficient and necessary condition for the existence of a TLS solution is that $\boldsymbol{V}_{22}$ is non-singular. In addition, if and only if $\sigma_{M_A} \neq \sigma_{M_A+1}$ the unique solution is given by

$$\hat{\boldsymbol{F}}_{opt} = -\boldsymbol{V_{12}}\boldsymbol{V_{22}}^{-1}. \tag{8.9}$$

This solution allows an estimation of the full calibration matrix $\mathbf{F}$ and enables our validation using experiment in this chapter on the widely used assumption that $\mathbf{F}$ is diagonal.

For comparison purpose, if we assume the calibration matrix is diagonal, we decompose (8.6) into M_A independent SISO TLS equations.

8.3 Measurement setup

The measurement is carried out using the open-source hardware and software development platform OpenAirInterface with ExpressMIMO2 boards as illustrated in Fig. 8.2. This board is built around a low-cost Spartan-6 150LXT FPGA with native PCIexpress on the FPGA fabric, which is coupled with 4 high-performance LTE RF ASICs on-board, manufactured by Lime Micro Systems (LMS6002D). The chosen RF technology covers a very large part of the available spectrum from 300MHz to 3.8GHz with a programmable bandwidth up to 28 MHz. The board can be used together with OpenAirInterface's software defined radio (SDR) OpenAir4G Modem implementing the 3GPP LTE Rel. 10 standard and running in real-time on common x86 Linux machines. For the measurements in this work, we however used the non real-time mode by simply sending and receiving frames. All the measurements were taken indoors in a controlled laboratory environment.

In the experiment, we used an LTE-like OFDM waveform for the transmission. Each OFDM symbol consists of 512 carriers, out of which 300 are filled with random QPSK

symbols and the rest are set to zero. An extended cyclic prefix (ECP) of 128 samples is added to each OFDM symbol after the 512-point inverse fast Fourier transform (IFFT). The sampling rate is 7.68M samples per second, resulting in an effective bandwidth of 4.5MHz. Ten subframes each with 12 ECP-OFDM symbols compose the TDD OFDM frame which is split into the bi-directinal transmission between A and B. The number of subframes used for A to B and B to A transmission is determined by the number of antennas at node A.

We carried our measurements for 3 configurations: 2×1, 4×1 and 15×1 MISO. For 2×1 and 4×1 MISO, we use two ExpressMIMO2 boards, acting as node A and B, respectively. The two boards are connected with cables for both frame and frequency synchronization. We activated multiple RF chains (2 or 4) at node A whereas only one RF chain of node B was used. The antennas at node A spaced by a quarter of the wavelength had fixed positions whereas we moved around the antenna of node B to create different channels. In order to make the TLS solution converge, channel measurements for different locations of B are preferred to be uncorrelated, for which reason, we randomly chose 45 different locations for node B in the laboratory. When one antenna of node A is on transmission, other antennas of the same side keep silent so that an orthogonality in the time domain is achieved. On each chosen location of B, 10 such TDD OFDM frames are sent to have a better estimation result of the calibration matrix. The carrier frequency used in the experiment is 1.9 GHz whereas the transmission power is of around 10dBm. Both transmit and receive gains on all the RF chains are set to 10dB. The receive noise figure is around 10dB. Depending on the location of B, this gives a SNR of up to 40 dB.

For 15×1 MISO, we use 4 ExpressMIMO2 synchronized by Ettus Research's Octo-clock [74]. The first card is used as the master card. It generates a 61.44MHz clock signal and a pulse per frame (every 10ms) signal which is amplified by the Octo-clock to synchronize the other 3 cards. The antenna connected to the last chain of the last card is used as node B whereas others act as node A. The pilot for each antenna on side A occupies one subframe duration and every other sub-carrier, thus the first 8 subframes are used for the 15 antennas on node A to transmit its pilots to node B. The antenna at side B uses the whole band of the last two subframes for B to A transmission.

8.4 Experiment results

We first perform the measurements for 2×1 and 4×1 MISO systems both using full estimation and diagonal estimation. The results are shown by Fig. 8.3-8.6 in which each arc is composed of 300 elements covering the whole bandwidth from the first carrier ν_1 to the last carrier ν_{300}. The blue dots are the diagonal elements and other colors are off-diagonal elements both indicated by f_{ij} $(i, j = 1, 2, 3, 4)$ representing the value on the i^{th} row and j^{th} column in $\boldsymbol{F}$. We observe that the diagonal estimations in Fig. 8.4 and 8.6 are very similar to the corresponding elements in Fig. 8.3 and 8.5. Diagonal elements are at least 30dB larger than off-diagonal elements. We also observe that the amplitude of diagonal elements have different values between 0.8 and 1.4, which is a result of the RF gain imbalance, knowing that in the perfect case, they should all be 1 under the given configuration. It is also worth noting that the estimation of $\boldsymbol{F}$ is carried out independently for different carriers and the smoothness of the amplitudes over the whole bandwidth implies that efficient pilot design on certain carriers is possible, i.e. in practice, the calibration does not have to be done for all carriers. Moreover, the phases of the

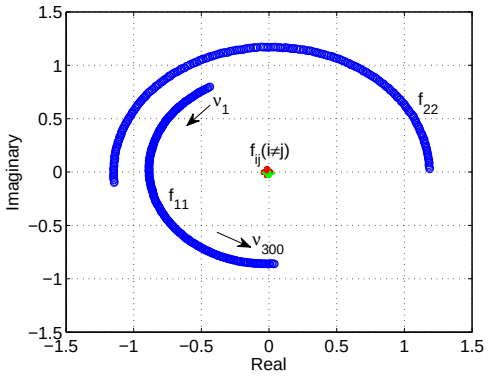


Figure 8.3: Full estimation of $\boldsymbol{F}$ in a 2×1 MISO.

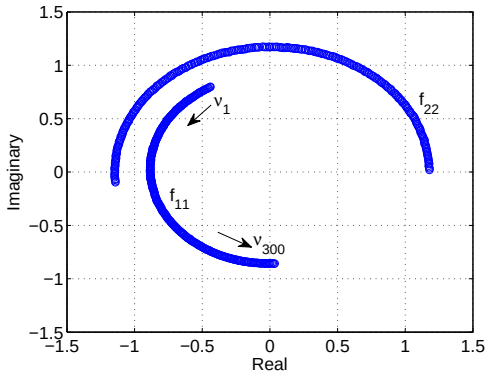


Figure 8.4: Diagonal estimation of $\boldsymbol{F}$ in a 2×1 MISO.

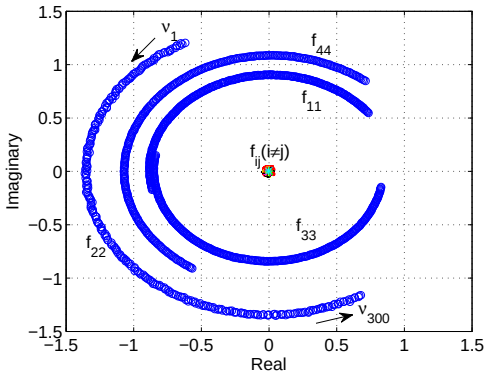


Figure 8.5: Full estimation of $\boldsymbol{F}$ in a 4×1 MISO.

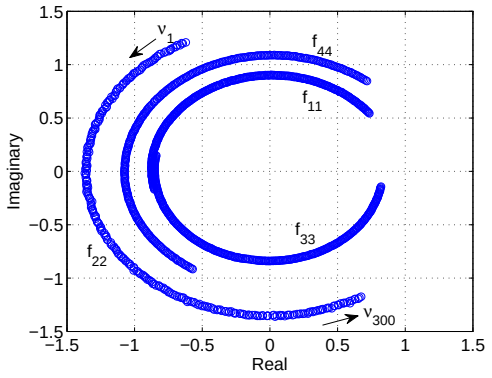


Figure 8.6: Diagonal estimation of $\boldsymbol{F}$ in a 4×1 MISO.

elements, taking the first carrier ν_1 as an example, change randomly at each reset of the card, however its evolution as a function of the frequency can be explained by the signal propagation delay on the boards. Let us consider a SISO system model in Fig. 8.7 where the delay effect is separated from other factors. The delays in blocks t_A, r_A, t_B and r_B are noted by τ_{t_A}, τ_{r_A}, τ_{t_B}, τ_{r_B} respectively and t_{A_0}, r_{A_0}, t_{B_0} and r_{B_0} are blocks without delay. The calibration matrix can be represented as

$$\begin{aligned} f &= \frac{r_B t_A}{t_B r_A} = \frac{r_{B_0} t_{A_0} e^{-j2\pi\nu\tau_{r_B}} e^{-j2\pi\nu\tau_{t_A}}}{t_{B_0} r_{A_0} e^{-j2\pi\nu\tau_{t_B}} e^{-j2\pi\nu\tau_{r_A}}} \\ &= f_0 e^{-j2\pi\nu[(\tau_{t_A}+\tau_{r_B})-(\tau_{t_B}+\tau_{r_A})]} = f_0 e^{-j2\pi\nu\Delta\tau} \end{aligned} \tag{8.10}$$

where f_0 is the SISO calibration coefficient without delay and $\Delta\tau = (\tau_{t_A} + \tau_{r_B}) - (\tau_{t_B} + \tau_{r_A})$ is the time delay difference between the transmission from A to B and that from B to A. In our experiment, the total phase spans for different chains over the whole bandwidth are between 0.95π and 1.1π, which correspond to delay differences between $0.1\mu s$ and $0.12\mu s$. The delay can be introduced by hardwares on the RF chains, such as filters, whereas in our experiment, it is mainly to due to the delay of the daisy chain used for synchronization purpose. If two cards are perfectly synchronized, the span of the arc should be much shorter, or even become a point for a 5MHz band.

For the 15×1 MISO experiment, a full $\mathbf{F}$ estimation needs many more independent locations of B, thus is not easy to obtain. For this reason, we only carried out the diagonal

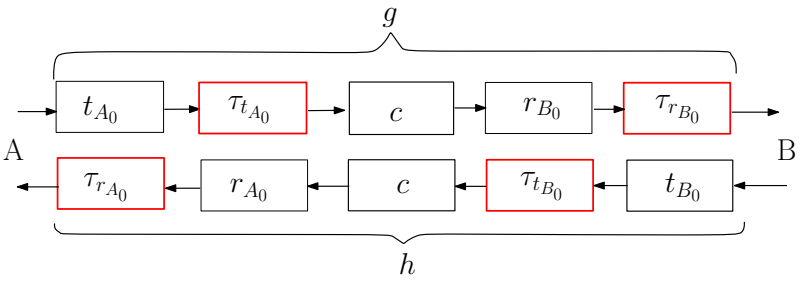


Figure 8.7: Reciprocity model with delay.

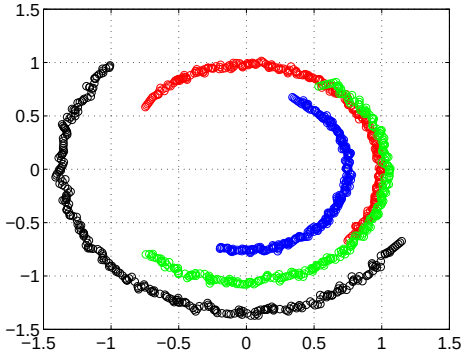


Figure 8.8: Calibration parameters for antenna 1-4 in a 15×1 MISO (card 1).

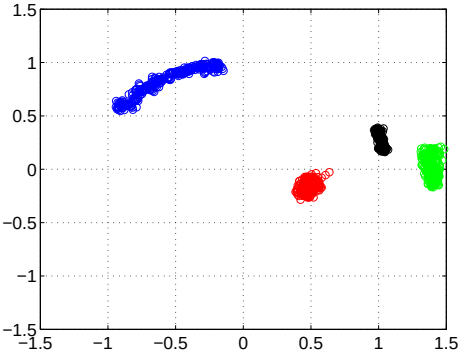


Figure 8.9: Calibration parameters for antenna 5-8 in a 15×1 MISO (card 2).

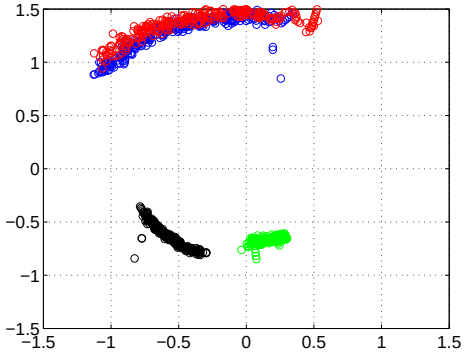


Figure 8.10: Calibration parameters for antenna 9-12 in a 15×1 MISO (card 3).

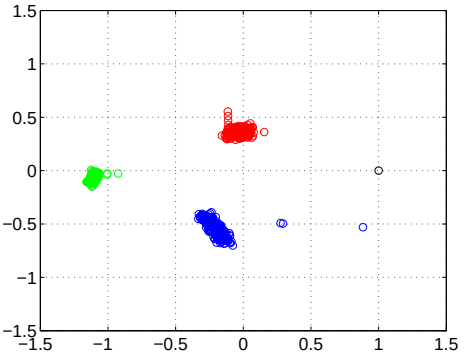


Figure 8.11: Calibration parameters for antenna 13-15 in a 15×1 MISO (card 4). The black point at $(1, 0)$ corresponds to node B.

estimation. The result is shown in Fig. 8.8-8.11, where the black point on card 4 is node B, whose calibration coefficient is defined as 1. We observe that for card 4, as the RF chains are on the same board as node B, the calibration coefficients on different frequencies tend to become a point. Most coefficients on card 2 and 3 also have very short frequency span since the output of the Octo-clock is quite synchronized. For card 1, since it acts as the master card and external clock provider, as the Octo-clock introduces a delay between its input and output, we observe similar arcs for different frequencies as in Fig. 8.3-8.6. In this measurement, since we perform the estimation directly on the raw data with out having

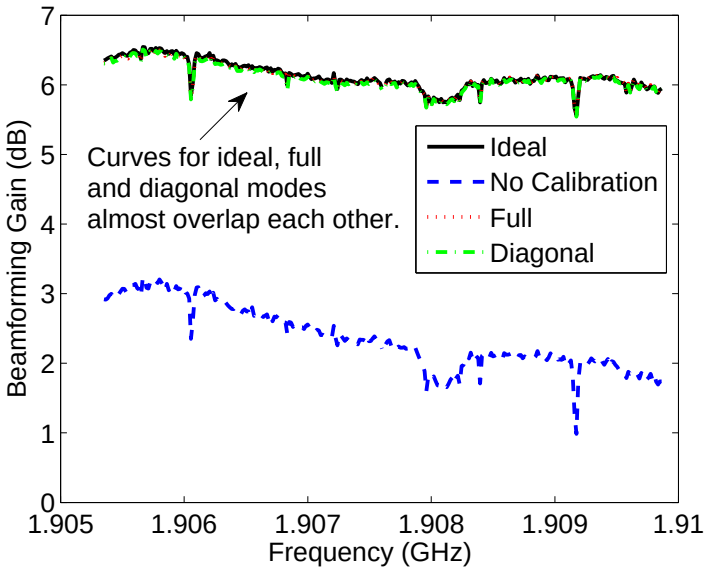


Figure 8.12: Beamforming gain of a 4×1 MISO system with regard to a SISO system under different assumptions (SNR averaged over 28 random locations).

selected the best channel measurements, also because of the interpolation and the fact that less channel accumulation are used for the estimation, the quality of the $\mathbf{F}$ estimation is not as good as in Fig. 8.3-8.6.

8.5 Beamforming performance

When the calibration matrix is obtained in the initializing phase, it can be used in the transmission phase to assess the CSIT based on the B to A measurement so that the feedback of the channel information is avoided. In this section we adopt the MRT beamforming to compare the beamforming performance under different CSIT acquisition methods. Let us consider the signal received by B as

$$y = \mathbf{h}_{A\to B}^T \boldsymbol{s} + n \tag{8.11}$$

MRT beamforming consists in precoding the transmitted symbol x by the normalized conjugate channel vector as

$$\boldsymbol{s} = \frac{(\hat{\mathbf{h}}_{A\to B}^T)^H}{\|\hat{\mathbf{h}}_{A\to B}\|} x = \frac{\hat{\mathbf{h}}_{A\to B}^*}{\|\hat{\mathbf{h}}_{A\to B}\|} x \tag{8.12}$$

We compare the beamforming SNR noted by γ for a randomly chosen location of B under 4 different assumptions.

- Ideal
  In this case, we assume node A knows $\hat{\mathbf{h}}_{A\to B}$ measured by node B. The beamforming SNR is given by
  $$\gamma_{\text{ideal}} = \frac{\|\mathbf{h}_{A\to B}^T \hat{\mathbf{h}}_{A\to B}^*\|^2}{\|\hat{\mathbf{h}}_{A\to B}\|^2} \frac{\sigma_x^2}{\sigma_n^2} \tag{8.13}$$

- No calibration
  Under this assumption, the transceiver hardware is considered totally reciprocal and $\hat{\mathbf{h}}_{B\to A}$ is considered to be equal to $\hat{\mathbf{h}}_{A\to B}$, thus no calibration is needed. The SNR is

$$\gamma_{\text{no calib}} = \frac{\|\mathbf{h}_{A\to B}^T \hat{\mathbf{h}}_{B\to A}^*\|^2}{\|\hat{\mathbf{h}}_{B\to A}\|^2} \frac{\sigma_x^2}{\sigma_n^2} \tag{8.14}$$

- Diagonal $\boldsymbol{F}$ estimation
  The RF chain crosstalk and the antenna mutual coupling are ignored and the calibration matrix is assumed to be diagonal. $\boldsymbol{F}$, noted by $\hat{\boldsymbol{F}}_d$ here, is thus estimated by solving 4 independent SISO TLS problems. The SNR is given by

$$\gamma_{\text{diag}} = \frac{\|\mathbf{h}_{A\to B}^T (\hat{\mathbf{h}}_{B\to A}^T \hat{\boldsymbol{F}}_d)^H\|^2}{\|\hat{\mathbf{h}}_{B\to A}^T \hat{\boldsymbol{F}}_d\|^2} \frac{\sigma_x^2}{\sigma_n^2}. \tag{8.15}$$

- Full $\boldsymbol{F}$ estimation
  Taking into account the RF chain crosstalk and the antenna mutual coupling, we estimate the complete calibration matrix $\boldsymbol{F}$. The SNR is

$$\gamma_{\text{full}} = \frac{\|\mathbf{h}_{A\to B}^T (\hat{\mathbf{h}}_{B\to A}^T \hat{\boldsymbol{F}})^H\|^2}{\|\hat{\mathbf{h}}_{B\to A}^T \hat{\boldsymbol{F}}\|^2} \frac{\sigma_x^2}{\sigma_n^2} \tag{8.16}$$

We randomly choose 28 new locations for node B in the 4×1 MISO system and let node A transmit data after MRT precoding under these 4 assumptions. We then spatially average the measured SNR for them and compare with that of a SISO system, where only one RF chain in node A is activated, thus obtain the beamforming gain as illustrated in Fig. 8.12. Note that, this experiment was conducted independently rather than using the data having been collected for the estimation of $\boldsymbol{F}$. We observe that the beamforming gains of both diagonal estimation and full estimation are very similar to that of the ideal case, being around 6dB, which means that the channel reciprocity is fully achieved using relative calibration and ignoring the off-diagonal elements in $\boldsymbol{F}$ is reasonable in a small scale MISO system. When no calibration is used for TDD system, there is some beamforming performance degradation. In our 4×1 MISO system, the average beamforming gain without channel calibration is around 2dB, thus having more than 3dB loss with regard to calibration modes.

8.6 Summary

In this chapter, we presented the calibration experiment setup and the real-world measurement results, which give an insight on the hardware impairment. We also studied the beamforming performance under different CSIT acquisition methods for a small scale MISO system and the results indicate that the diagonal assumption of the calibration matrix in [21], [22] is reasonable. However, it is unclear that whether increasing the number of antennas will make the impact of RF crosstalk and the antenna mutual coupling more severe, although most literature assumes that the calibration matrix is still diagonal [25, 28]. It is of high interest in the future work to scale up the experiment to see whether there would be visible performance degradation if the calibration matrix is still assumed to be diagonal in a massive MIMO case.

Chapter 9

How Accurately Should We Calibrate?

Although various calibration methods were provided, little attention has been given to the calibration accuracy needed. From a system design point of view, this topic is essential since it determines how much resources should be used to do the calibration. Consider a practical massive MIMO TDD system, two different working phases are normally necessary: the calibration phase which is dedicated to estimating calibration coefficients and the beamforming phase when we perform beamforming based on the CSIT inferred from these coefficients and the instantly measured UL CSI. Existing literature studies the impact of the calibration accuracy on the CSIT assuming that the UL channel estimation is perfect, which is not realistic in practice. Simulations of this type can be found in [27] with no closed-form solution provided. [75] gives a beamforming performance analysis of the TDD calibration, but the study was limited to Argos method and ZF precoding where the UL channel estimation was again assumed to be perfect.

In this chapter, we provide a general closed-form analysis on the DL CSI accuracy, taking into account the impact from both the accuracy of the calibration coefficients and the instantaneous UL channel estimation. We show that when the UL channel estimation is poor, the efforts to improve the calibration coefficients are in vain. Moreover, we will simulate the impact of both factors on the final beamforming performance using conjugate and ZF precoding. We will show that in high DL SNR region, ZF is more sensitive to the inaccuracy in the estimation of calibration coefficients and UL channel. We provide a method to determine the accuracy level that the calibration coefficients should achieve to guarantee a certain level of beamforming performance.

9.1 Calibration accuracy

In this section we assume that the calibration matrix $\mathbf{F}$ has been already estimated, we are now in the beamforming phase where we apply $\hat{\mathbf{F}}$ to the instantaneously estimated UL channel $\hat{\mathbf{h}}_{B\to A}$ to infer the CSIT $\mathbf{h}_{A\to B}$. The accuracy of the CSIT obtained from such a calibration process depends on three factors: 1) the accuracy of UL channel measurement $\hat{\mathbf{h}}_{B\to A}$, which depends on the channel quality and channel estimator; 2) the accuracy of the estimated relative calibration matrix $\hat{\mathbf{F}}$, which depends on the calibration method

and the number of resources used to do the calibration and 3) the reciprocity level of the UL/DL physical channel in the air, which is determined by the UL/DL switch time and the channel coherence time. In our analysis, we would not consider the third factor and assume a perfect reciprocity for the UL/DL physical channel in the air. We perform the theoretical analysis to understand how the first two factors influence the calibrated CSIT accuracy.

9.1.1 UL channel estimation error

Let us firstly stick to the MISO scenario in Fig. 8.1. We assume a perfect estimation on the relative calibration matrix, i.e., $\hat{\mathbf{F}} = \mathbf{F}$, and study the influence of the UL channel measurement's quality on the accuracy of CSIT. The signal model for the UL channel estimation at time instant t is given by

$$\mathbf{y}_{A,t} = \mathbf{h}_{B\to A} s_{B,t} + \mathbf{n}_{A,t} \tag{9.1}$$

where $s_{B,t}$ is the transmitted pilot, $\mathbf{y}_{A,t} \in \mathbb{C}^{M\times 1}$ is the received signal at BS, and the noise $\mathbf{n}_{A,t}$ is a vector of circularly-symmetric complex Gaussian random variables following $\mathcal{CN}(0, \sigma_{n,A}^2\mathbf{I})$. Assume that L_B symbols are used for UL channel estimation and stack the transmission for $t = 1, 2, \cdots, L_B$, we have

$$\mathbf{Y}_A = \mathbf{h}_{B\to A}\mathbf{s}_B^T + \mathbf{N}_A \tag{9.2}$$

where $\mathbf{Y}_A \in \mathbb{C}^{M\times T_B}$, $\mathbf{s}_B^T \in \mathbb{C}^{1\times L_B}$, $\mathbf{N}_A \in \mathbb{C}^{M\times L_B}$ are obtained by arranging the corresponding vectors in columns. We adopt the LS estimator as

$$\hat{\mathbf{h}}_{B\to A} = \mathbf{Y}_A \frac{\mathbf{s}_B^*}{\|\mathbf{s}_B\|^2}. \tag{9.3}$$

Using the estimated channel and the estimation error $\Delta\mathbf{h}_{B\to A}$, Eq. (8.1) can be rewritten as

$$\mathbf{h}_{A\to B}^T = (\hat{\mathbf{h}}_{B\to A}^T - \Delta\mathbf{h}_{B\to A}^T)\mathbf{F} \tag{9.4}$$

As LS estimators are linear, $\Delta\mathbf{h}_{B\to A}$ remain circularly-symmetric Gaussian vector. Given that the normalized transmitted symbols on different time slots are i.i.d variables with unit power, i.e., $\mathbb{E}[s_{B,t_1}s_{B,t_2}] = \delta(t_1 - t_2)$ where δ is the Kronecker delta, the variance (with regard to transmitted pilot and noise) of the channel estimators is

$$\mathbb{E}_{\mathbf{s}_B,\mathbf{N}_A}\left[\Delta\mathbf{h}_{B\to A}\Delta\mathbf{h}_{B\to A}^H\right] = \frac{\sigma_{n,A}^2}{L_B}\mathbf{I}. \tag{9.5}$$

The MSE of the calibrated channel normalized by the number of BS antennas is

$$\begin{aligned}
\text{NMSE} =& \frac{1}{M}\mathbb{E}_{\mathbf{s}_B,\mathbf{N}_A}\left[\|\mathbf{h}_{A\to B} - \mathbf{F}^T\hat{\mathbf{h}}_{B\to A}\|^2\right] \\
=& \frac{1}{M}\mathbb{E}_{\mathbf{s}_B,\mathbf{N}_A}\left[\text{Tr}\left\{(\mathbf{F}^T\Delta\mathbf{h}_{B\to A})(\mathbf{F}^T\Delta\mathbf{h}_{B\to A})^H\right\}\right] \\
=& \text{Tr}\left\{\mathbf{F}^T\mathbb{E}_{\mathbf{S}_A,\mathbf{N}_A}\left[\Delta\mathbf{h}_{B\to A}\Delta\mathbf{h}_{B\to A}^H\right]\mathbf{F}^*\right\} \\
=& \frac{\sigma_{n,A}^2}{ML_B}\text{Tr}\left\{\mathbf{F}^T\mathbf{F}^*\right\}
\end{aligned} \tag{9.6}$$

where $\text{Tr}\{\cdot\}$ is the trace of a matrix.

9.1.2 Relative calibration matrix estimation error

Now let us additionally consider the second factor, the accuracy of $\hat{\mathbf{F}}$ and study its impact on the calibrated CSIT's accuracy. The error of $\hat{\mathbf{F}}$ stems from two aspects: 1) approximation error which comes from the simplification on $\mathbf{F}$, e.g., assuming $\mathbf{F}$ is diagonal; 2) estimation error on $\mathbf{F}$, which can be caused by the bi-directional channel estimation inaccuracy during the calibration phase and the variation of $\mathbf{F}$. Indeed, even with perfect instantaneous bi-directional channel estimations, the small variation of the hardware circuits can cause an imperfect calibration matrix estimation in two ways: on the one hand, the real $\mathbf{F}$ during the beamforming phase varies from that in the calibration phase; on the other hand, $\mathbf{F}$ estimation is usually carried out in a certain time interval during which $\mathbf{F}$ is slightly varying as well, thus the obtained $\hat{\mathbf{F}}$ is an average value. Considering $\Delta\mathbf{F}$ as the error of $\mathbf{F}$, i.e., $\hat{\mathbf{F}} = \mathbf{F} + \Delta\mathbf{F}$, and note $\mathbf{V}$ the covariance matrix of the channel from B to A, i.e., $\mathbf{V} = \mathbb{E}\left[\mathbf{h}_{B\to A}\mathbf{h}_{B\to A}^H\right]$, we can represent the normalized MSE of calibrated CSIT by

$$
\begin{aligned}
\text{NMSE} =& \frac{1}{M}\mathbb{E}_{\mathbf{h}_{B\to A},\mathbf{s}_B,\mathbf{N}_A}\left[\|\hat{\mathbf{F}}^T\hat{\mathbf{h}}_{B\to A} - \mathbf{h}_{A\to B}\|^2\right] \\
=& \frac{1}{M}\mathbb{E}_{\mathbf{h}_{B\to A},\mathbf{s}_B,\mathbf{N}_A}\left[\|\hat{\mathbf{F}}^T\Delta\mathbf{h}_{B\to A} + \Delta\mathbf{F}^T\mathbf{h}_{B\to A}\|^2\right] \\
=& \frac{1}{M}\text{Tr}\left\{\hat{\mathbf{F}}^T\mathbb{E}_{\mathbf{s}_B,\mathbf{N}_A}\left[\Delta\mathbf{h}_{B\to A}\Delta\mathbf{h}_{B\to A}^H\right]\hat{\mathbf{F}}^* + \Delta\mathbf{F}^T\mathbb{E}_{\mathbf{h}_{B\to A}}\left[\mathbf{h}_{B\to A}\mathbf{h}_{B\to A}^H\right]\Delta\mathbf{F}^*\right\} \\
=& \frac{1}{M}\text{Tr}\left\{\frac{\sigma_{n,A}^2}{L_B}\hat{\mathbf{F}}^T\hat{\mathbf{F}}^* + \Delta\mathbf{F}^T\mathbf{V}\Delta\mathbf{F}^*\right\} \\
=& \frac{1}{M}\text{Tr}\left\{\frac{\sigma_{n,A}^2}{L_B}(\mathbf{F}+\Delta\mathbf{F})^T(\mathbf{F}+\Delta\mathbf{F})^* + \Delta\mathbf{F}^T\mathbf{V}\Delta\mathbf{F}^*\right\} \\
=& \frac{\sigma_{n,A}^2}{ML_B}\text{Tr}\left\{\mathbf{F}^T\mathbf{F}^*\right\} + \frac{1}{M}\text{Tr}\left\{\Delta\mathbf{F}^T\left(\mathbf{V} + \frac{\sigma_{n,A}^2}{L_B}\mathbf{I}\right)\Delta\mathbf{F}^*\right\} + \frac{\sigma_{n,A}^2}{ML_B}\text{Tr}\left\{\mathbf{F}^T\Delta\mathbf{F}^* + \Delta\mathbf{F}^T\mathbf{F}^*\right\}
\end{aligned}
\tag{9.7}
$$

The first term in Eq. (9.7) is the same as in Eq. (9.6), which is purely due to the UL channel estimation error, and the rest is the additional error brought in by considering the error on $\hat{\mathbf{F}}$. Note that if we assume $\mathbf{F}$ to be diagonal, then $\Delta\mathbf{F} = \Delta\mathbf{F}_d + \mathbf{F}_o$, where $\Delta\mathbf{F}_d$ represents the estimation error on the diagonal matrxi $\mathbf{F}_d$ and $\mathbf{F}_o$ is the approximation error by ignoring the off-diagonal elements.

9.2 Simulation results

In this section, we define models for $\mathbf{T}_A$, $\mathbf{R}_A$, t_B, r_B, based on which we calculate the calibration matrix $\mathbf{F}$. We also model the channel in the air $\mathbf{c}$ for a co-located massive MIMO system using a geometry based Rician channel. We illustrate how the calibration matrix inaccuracy and the error in the UL channel estimation impact the CSIT accuracy. Additionally we also perform simulations to view their final impact on beamforming performance. For these objectives, we use a BS operating at 2.6GHz with a 8×8 square antenna array whose elements are separated by half of the wavelength.

9.2.1 Hardware asymmetry model and channel model

For $\mathbf{T}_A$, $\mathbf{R}_A$, as antenna spacing in our antenna configuration is at least half of the wavelength, the antenna mutual coupling can be neglected [19] [42] and the off-diagonal elements can thus be assumed to be 0. The diagonal elements in $\mathbf{T}_A$, $\mathbf{R}_A$, as well as t_B and r_B, are modeled as i.i.d. random variables, with uniformly distributed phase between $[-\pi \ \ \pi]$ and independent magnitude uniformly distributed on $[1-\epsilon \ \ 1+\epsilon]$, with ϵ chosen such that the standard deviation of the squared-magnitudes is 0.1, as in [27] and [51]. Based on this model, we can easily obtain the calibration matrix $\mathbf{F}$ using $\mathbf{F} = \frac{r_B}{t_B}\mathbf{R}_A^{-T}\mathbf{T}_A$.

Moreover, for the channel in the air $\mathbf{c}$, we use a geometry based normalized Rician channel model as in [76] given by

$$\mathbf{c} = \sqrt{K}\mathbf{c}_{\text{LOS}} + \sqrt{1-K}\mathbf{c}_{\text{diffuse}}, \tag{9.8}$$

where $\mathbf{c}_{\text{LOS}}$ is the line-of-sight component, the elements of which have a unit amplitude and geometry based phase (i.e., the phase is calculated according to radio's incidence angle from UE, thus depends on the relative position of the UE and the antenna element in space); $\mathbf{c}_{\text{diffuse}}$ is the diffuse component corresponding to the standard i.i.d. Rayleigh distribution $\mathcal{CN}(0, \mathbf{I})$; and K is the linear Rician K factor.

9.2.2 Simulation results on the CSIT MSE

Let us first study the impact of calibration matrix accuracy and UL channel estimation on the MSE of CSIT. To obtain a general result, we don't specify the specific calibration method used, thus the elements in $\Delta\mathbf{F}_d$ are assumed to be i.i.d. circularly-symmetric Gaussian variables following $\mathcal{CN}\left(0, \sigma^2_{\Delta\mathbf{F}_d}\right)$. The $\mathbf{F}_d$ estimation quality can be evaluated by the normalized MSE defined as

$$\text{NMSE}_{\mathbf{F}_d} = \frac{\|\Delta\mathbf{F}_d\|_F^2}{\|\mathbf{F}_d\|_F^2} = \frac{M\sigma^2_{\Delta\mathbf{F}_d}}{\|\mathbf{F}_d\|_F^2} \tag{9.9}$$

where $\|\cdot\|_F$ represents the Frobenius norm, i.e., the norm of the diagonal vector in our case. For UL, we use $L_B = 10$ symbols as the pilots to estimate the UL channel. The K factor in the channel model (9.8) is 0, i.e., standard Rayleigh channel is used. Under this assumption, the covariance matrix of $\mathbf{h}_{B\to A}$ is $\mathbf{V} = |t_B|^2\mathbf{R}_A\mathbf{R}_A^H$.

The MSE of calibrated CSIT is shown in Fig. 9.1. We observe that the improvement of calibration matrix accuracy and UL channel estimation can both enhance the accuracy of CSIT. When the UL channel SNR is low, the curves for $\text{MSE}_{\mathbf{F}_d}$ from 0.01 to 10^{-5} almost overlap each other, meaning that the accuracy of UL channel estimation is limiting the calibrated CSIT accuracy and improving $\hat{\mathbf{F}}$ accuracy will be useless. On the other hand, when the UL channel SNR is sufficiently high, the accuracy on the calibration matrix become the limiting factors and all curves become flat. In this case, improving the UL channel estimation accuracy has no further contribution. Furthermore, when the accuracy of $\hat{\mathbf{F}}$ is poor, the corresponding calibration CSIT accuracy curve become flat at a relatively low SNR.

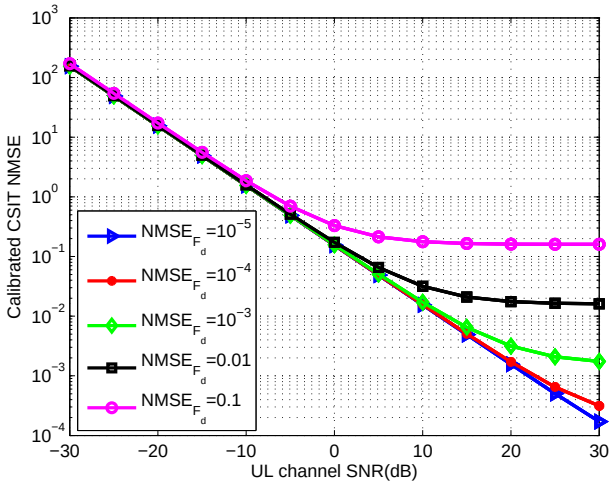


Figure 9.1: Calibrated CSIT normalized MSE as a function of the accuracy of UL estimation and the calibration matrix in a 64×1 MISO system ($L_B = 10$).

9.2.3 Simulation results on beamforming performance

The signal model for the i^{th} user in a MU-MIMO system is given by

$$y_i = \mathbf{h}_i^T \mathbf{w}_i x_i + \sum_{j \neq i} \mathbf{h}_i^T \mathbf{w}_j x_j + \mathbf{n}_i \tag{9.10}$$

where x_i and y_i are the transmitted and received signal for the i^{th} user. The transmission power is set to 1. $\mathbf{w}_i$ and $\mathbf{h}_i^T = r_{B,i} \mathbf{c}_i^T \mathbf{T}_A$ are the corresponding precoding weights and the channel from the BS to the i^{th} user respectively. We use conjugate and ZF beamforming in this simulation. For MRT beamforming, $\mathbf{w}_i = \hat{\mathbf{h}}_i^* / \|\hat{\mathbf{h}}_i\|$ with $\hat{\mathbf{h}}_i$ being the estimated DL channel, whereas for ZF, $\mathbf{w}_i = \hat{\mathbf{h}}_i^* (\hat{\mathbf{h}}_i^T \hat{\mathbf{h}}_i^*)^{-1} / \eta$, where η is the normalizing factor keeping the transmission power for each UE being 1. Note that the first term in (9.10) is the desired signal, the second term is the interference stemming from the transmission for other users and $\mathbf{n}_i$ is the circularly-symmetric complex Gaussian noise following $\mathcal{CN}(0, \sigma_{n,i}^2 \mathbf{I})$. The signal-to-interference-plus-noise ratio (SINR) for user i is given by

$$\mathrm{SINR}_i = \frac{\mathbb{E}\left[\|\mathbf{h}_i^T \mathbf{w}_i\|^2\right]}{\mathbb{E}\left[\|\sum_{j \neq i} \mathbf{h}_i^T \mathbf{w}_j\|^2\right] + \sigma_{n,i}^2}, \tag{9.11}$$

We use the same antenna array as defined in the beginning of this section and investigate a 64×8 MU-MIMO system. The K-factor in (9.8) is set to be 0.5. SINR loss with regard to a perfect CSIT will be used as the performance indicator, which is given by

$$\mathrm{SINR}_{\mathrm{loss},i} = \frac{\mathrm{SINR}_{\mathrm{ideal},i}}{\mathrm{SINR}_i}, \tag{9.12}$$

where $\mathrm{SINR}_{\mathrm{ideal},i}$ is also calculated using (9.11) but with $\mathbf{w}_i$ and $\mathbf{w}_j$ obtained with perfect channel estimation $\mathbf{h}_i$ and $\mathbf{h}_j$, rather than their estimated values.

Figs. 9.2–9.5 illustrate the conjugate and ZF beamforming SINR loss (in dB) due to joint impact of the inaccuracy in $\hat{\mathbf{F}}$ and UL channel estimation for both DL SNR = 20dB

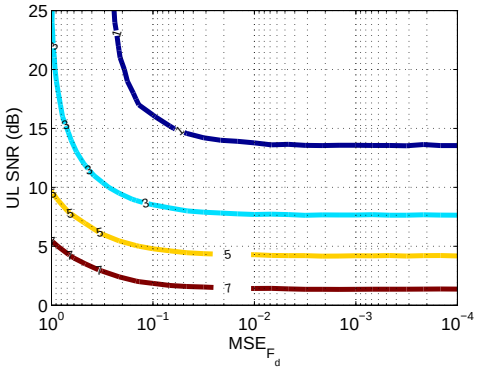


Figure 9.2: SINR loss (in dB) of MRT beamforming due to joint impact of the inaccuracy of both $\hat{\mathbf{F}}$ and UL channel estimation in a 64×8 system with DL SNR=0dB ($L_B = 10$).

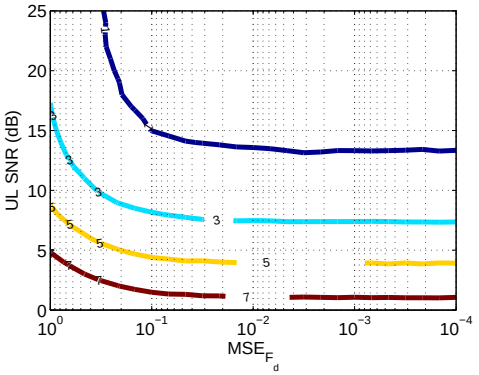


Figure 9.3: SINR loss (in dB) of MRT beamforming due to joint impact of the inaccuracy of both $\hat{\mathbf{F}}$ and UL channel estimation in a 64×8 system with DL SNR=20dB ($L_B = 10$).

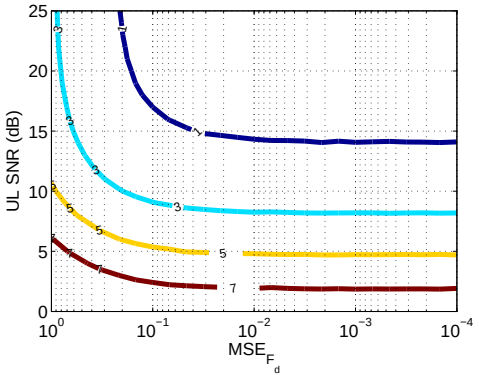


Figure 9.4: SINR loss (in dB) of ZF beamforming due to joint impact of the inaccuracy of both $\hat{\mathbf{F}}$ and UL channel estimation in a 64×8 system with DL SNR=0dB ($L_B = 10$).

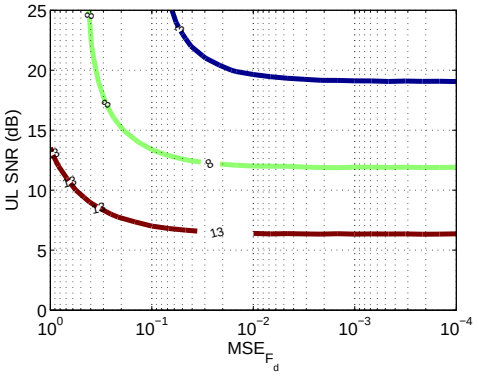


Figure 9.5: SINR loss (in dB) of ZF beamforming due to joint impact of the inaccuracy of both $\hat{\mathbf{F}}$ and UL channel estimation in a 64×8 system with DL SNR=20dB ($L_B = 10$).

($\sigma^2_{n,i} = 0.01$) and DL SNR = 0dB ($\sigma^2_{n,i} = 1$) cases. Different contours in these figures indicate certain values of SINR losses for corresponding $\hat{\mathbf{F}}$ accuracy and UL channel SNR.

We observe that when DL SNR is low (Fig. 9.2 and 9.4), the beamforming performance degradation due to TDD reciprocity calibration inaccuracy is similar for conjugate and ZF beamforming since in this case noise is dominant over interference, whereas when DL SNR is high (Fig. 9.3 and 9.5), ZF beamforming is much more sensitive to the calibration matrix and UL channel estimation inaccuracy. For the latter case where DL SNR=20dB, let's take an example in which $\text{MSE}_{\mathbf{F}_d} = 10^{-2}$ and UL SNR = 10dB, MRT beamforming has less than 3dB SINR loss whereas for ZF, this loss is above 8dB.

Furthermore the joint impact illustration on the SINR loss also offers a useful tool to determine the calibration matrix accuracy we need to achieve if we define an acceptable SINR loss value. For example, if the DL SNR=20dB, defining 1dB loss w.r.t the perfect channel estimation case in MRT beamforming when UL SNR = 15dB and $L_B = 10$ means

that $\text{MSE}_{\mathbf{F}_d}$ should be less than 0.1. Similarly for ZF beamforming, allowing 3dB SINR performance loss under the case where UL SNR = 20dB with $L_B = 10$ implies that $\text{MSE}_{\mathbf{F}_d}$ should be less than 0.02.

9.3 Summary

In this chapter, we address the problem on how accurately we should calibrate a TDD massive MIMO system. We perform theoretical analysis on the impact of calibration matrix and UL channel estimation on the CSIT accuracy. We observe that both of them can become a limiting factor, and the CSIT accuracy can be improved only when we allocate more resources on the limiting element. We also perform simulation to study the joint impact of these two factors on both conjugate and ZF beamforming performance. The study shows that ZF is more sensitive to inaccuracy in the calibration matrix and UL channel estimation, especially in high DL SNR region. At the same time, we provide a method to determine the accuracy level that the calibration matrix should achieve to guarantee a certain level of beamforming performance, which can be a useful tool for system design.

Chapter 10

Massive MIMO Prototyping on OpenAirInterface

Prototyping, as the bridge between theory and practice, is an essential step to bring the massive MIMO concept into reality. It helps to study the feasibility and limitation of this technology in a real world environment. Theoretical system performance under simplified assumptions in literature can be evaluated under real propagation channel condition. New challenges in system building will be identified, where the trade-off between the performance and implementation complexity has to be taken into account. Prototypes also lay the basis in accumulating measurements for massive MIMO channel characterization and modeling. In this chapter, we present the massive MIMO testbed based on the OpenAirInterface platform.

10.1 Introduction

Since the introduction of the massive MIMO concept, several testbeds have been announced, including the Argos testbed from Rice University, the LuMaMi testbed from Lund University, Bristol University's massive MIMO testbeds and many others from industry. Though these testbeds made great progress in pushing the state of the art on system implementation to a new level, most are limited to simple uncoded modulation. The integration to a real 3GPP standard system is not investigated.

At the same time, in recent years, the open source 4G platform OpenAirInterface [77] initialized by EURECOM has achieved great success from both academia and industry, with a strong demand from the fast growing community in evolving the platform towards 5G.

The need to go beyond current state of the art testbeds implementation combined with the demands from open source community has motivated us to build an LTE compliant massive MIMO testbed based on the OpenAirInterface platform. We use the totally software realized LTE protocol stack to drive a large antenna array in order to demonstrate the feasibility of direct use of massive MIMO in current 3GPP standards. It is a good starting point, from a point of view of system implementation in studying the possibility of smooth evolution from 4G to 5G. The fruit of this work will then be shared with the

open source community, allowing other members to push forward the frontier in massive MIMO research based on the work.

This testbed is an open source real-time massive MIMO testbed, with the 3GPP LTE protocols implemented from the physical layer to the network layer. It can inter-operate with commercial UEs using TMs 1-7. To overcome the challenges of accurate CSI acquisition, the system is built in TDD mode so that the inherent channel reciprocity can be exploited. With its 64 antenna array, the system is designed to simultaneously serve up to 4 UEs on the same frequency-time domain resource as defined in current 3GPP standard, although the currently implementation allows us to support only 1 UE using TM7.

The platform demonstrates that massive MIMO can be perfectly fit into LTE and its usage in 5G can be a smooth evolution starting from the current 4G standard. It provides a platform based on which engineers can innovate and test their concepts on the challenges in making massive MIMO more efficient, such as common channel beamforming, new reference signal design and the design of CSI feedback schemes.

The remainder of this chapter is organised as follows: Section. 10.2 represents the state of the art on the massive MIMO testbed development in academia and industry. Section. 10.3 gives an overview of the OpenAirInterface massive MIMO testbed. Section. 10.4 and . 10.5 describes the hardware and software of our testbed, respectively. In Section. 10.6, we present the achieved results and the corresponding dissemination. We then summarize in Section. 10.7.

10.2 State of the art

In this section, we briefly describe some existing massive MIMO testbeds in the world.

- **Argos Testbed from Rice University**

  Argos prototype [25], as shown in Fig. 10.1 is the world's first reported massive MIMO testbed. It is equipped with 64 antennas and capable of serving 15 UEs simultaneously. The system is built with 16 WARP [78] boards each with 4 RF chains. An external clock distribution module is used to provide the reference clocks for both frequency and time synchronization. The central control consists of a host PC using MATLAB to send data, beamforming weights and control commands to the WARP modules via an Ethernet connexion. Based on a BS internal calibration with regard to a reference antenna, the testbed exploits the TDD channel reciprocity to acquire CSIT. Additionally, a local normalized MRT precoding scheme with equal power per antenna is used to avoid massive channel information exchange between the radio boards and the central controller. Argos system achieves a 6.7 fold capacity gains while using 1/64 of the transmission power by scaling the BS antenna number from 1 to 64 and serving 15 terminals with MU-MIMO.

- **LuMaMi Testbed from Lund University**

  LuMaMi testbed [47, 79, 80], illustrated in Fig. 10.2 is the massive MIMO system from Lund University collaborated with National Instruments (NI), supporting 100 antennas with 20MHz bandwidth and can simultaneously serving 10 UEs. The system is composed of four 18-slot PCIe/PXIe chassis in a star topology where one acts as the master, taking the role as the central controller. The master chassis

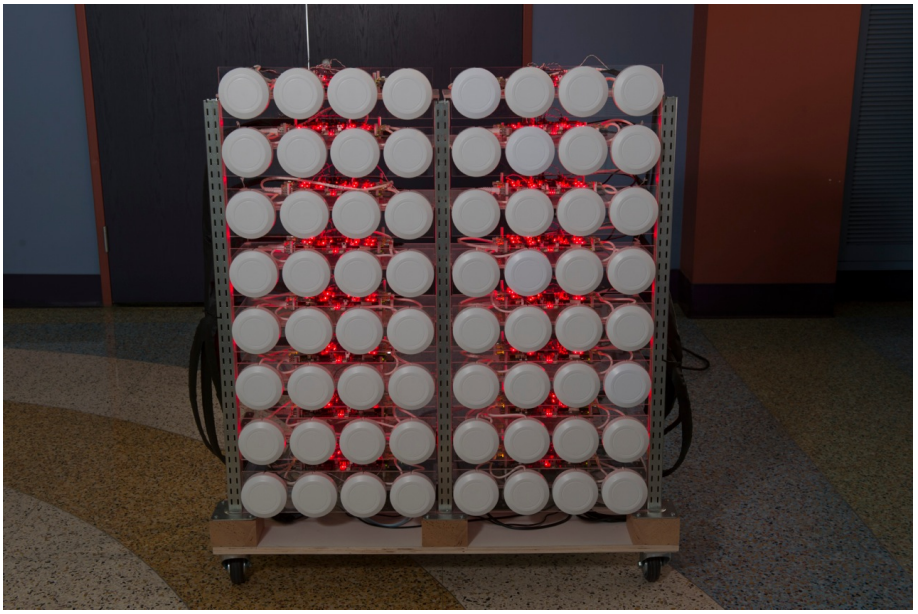


Figure 10.1: Rice University's Argos massive MIMO testbed.

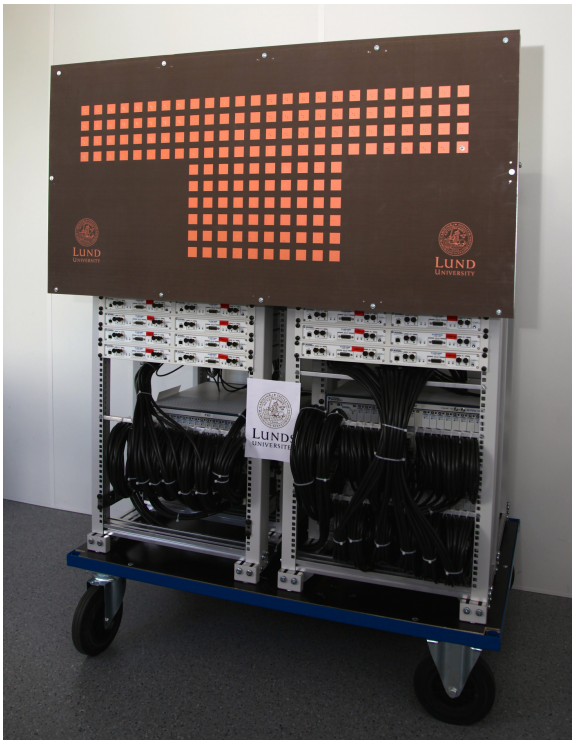


Figure 10.2: Lund University's LuMaMi massive MIMO testbed.

connects to the other three chassis through Gen 2 x8 PCI (MXIe) Express interface. 50 USRPs (NI 2943R/USRP-RIO) are then connected to these chassis using Gen 1 x4 MXIe cables for data transfer. Eight OctoClock modules are used to provide time and frequency synchronization for the USRPs. The master chassis embeds a x64 controller (NI PXIe-8135) which runs LabVIEW on a Windows 7 64-bit OS for radio configuration, system control, data storage, performance measurements, etc.

LuMaMi is also a TDD based system, using weighted LS calibration [27] (and later an ML estimator [28]) to compensate the hardware asymmetry.

- **Bristol University's massive MIMO testbed**

  The massive MIMO testbed being built at Bristol University [81], shown in Fig. 10.3 in collaboration with NI and Lund University follows very similar technology as the LuMaMi massive testbed. With 64 NI USRPs, this testbed is equipped with 128-antennas and aims to support 16 (or even 24) user streams over a 20MHz bandwidth. Different with LuMaMi, this testbed adopts a distributed architecture. This architecture overcomes the deployment challenges of co-located massive MIMO base stations and allows antennas to be deployed in segments. At the same time, it offers a greater spatial diversity. However, node synchronization and phase alignment becomes more challenging, which are currently under study.

Figure 10.3: Bristol's massive MIMO testbed.

Apart from those prototypes from academia, several testbeds have also been announced in industry, focusing on different aspects of system building. These testbeds include Huawei's massive MIMO testbed [82] which is quite similar as our work, Sumsung's prototype [83], CSIRO's Ngara prototype [84], and the testbed from Nutaq [85].

10.3 Testbed overview

We show in Fig. 10.4 the flexible and scalable TDD based OpenAirInterface massive MIMO system. It can support a large antenna array up to 64 elements with 5MHz bandwidth at the frequency of 2.6GHz. As defined in the current 3GPP standards, the system is designed to serve up to 4 users on the same time and frequency resource whereas the current implementation allows only 1 UE. The bandwidth limitation can be easily removed by upgrading the Field-Programmable Gate Array (FPGA) of ExpressMIMO2 RF platform or by using other RF platforms such as Ettus USRP B210, Ettus USRP x310

or LimeSDR. OpenAirInterface natively supports multiple RF platforms (USRP B210, USRP x310, LimeSDR) can be easily synchronized to support higher bandwidth massive MIMO platform using the same software architecture. The key parameters of the system are summarized in Table 10.1.

Figure 10.4: OpenAirInterface massive MIMO testbed.

Paramters	**Value**
Number of antennas	Up to 64
Center frequency	2.6GHz
Bandwidth	5MHz
Sampling Rate	7.68MS/s
FFT Size	512
Number of used subcarriers	300
Slot time	0.5ms
Maximum simultaneously served UEs	Currently 1, extendable

Table 10.1: Key parameters of OpenAirInterface massive MIMO testbed.

We show that massive MIMO can be smartly and perfectly fit into the current LTE standard. In fact, 3GPP has defined the notion of "Transmission Modes" (TMs) for different usage of MIMO in LTE, as described in Section 3.2. Among them, TM 7 is defined in Release 8, where an arbitrary number of physical antennas at base station can be used as a logical antenna port (port 5) to create a narrow beam for the targeted user. Release 9 extended TM 7 to TM 8, giving the possibility of transmitting a dual stream to a single or two users, whereas in release 10, this is further extended to TM 9, where up to 8 layers for a single user transmission and up to 4 layers for multiuser transmission is supported. Release 11 adds TM 10, similar to TM 9 with up to 8 layers transmission but the transmit antennas can be physically located on different base stations. In Release 13, no new transmission mode is defined, but CSI-RS has been extended to 16 ports [86]. In release 14 [87], the enhancement of Full-Dimension MIMO (special case of massive MIMO in 3GPP) for LTE has extended the CSI-RS to 32 ports with enhancement on CSI reports and support for providing higher robustness against CSI impairments.

The OpenAirInterface massive MIMO testbed relies on the implementation of TM 7-9 to

use the large number of antenna array, and can be easily extended beyond. Currently only TM 7 has been implemented. Commercial UEs supporting this transmission mode can be connected to the massive array base station, and can surf Internet through our massive MIMO testbed. We hope that with our community's support, we can extend OpenAirInterface towards future 3GPP releases thus implementing other transmission modes such as TM 8-10 and also integrate successfully with other supported RF platforms (USRP B210, USRP x310 and LimeSDR).

10.4 Hardware

This section presents the hardware architecture and components of the testbed.

10.4.1 Hardware architecture

The OpenAirInterface massive MIMO testbed mainly consists of 5 hardware components as shown in Fig. 10.5: a massive antenna array, 16 ExpressMIMO2 RF cards, a clock distribution module, 16-way PCIe backplane and a high-end Intel Xeon server running OpenAirInterface modem.

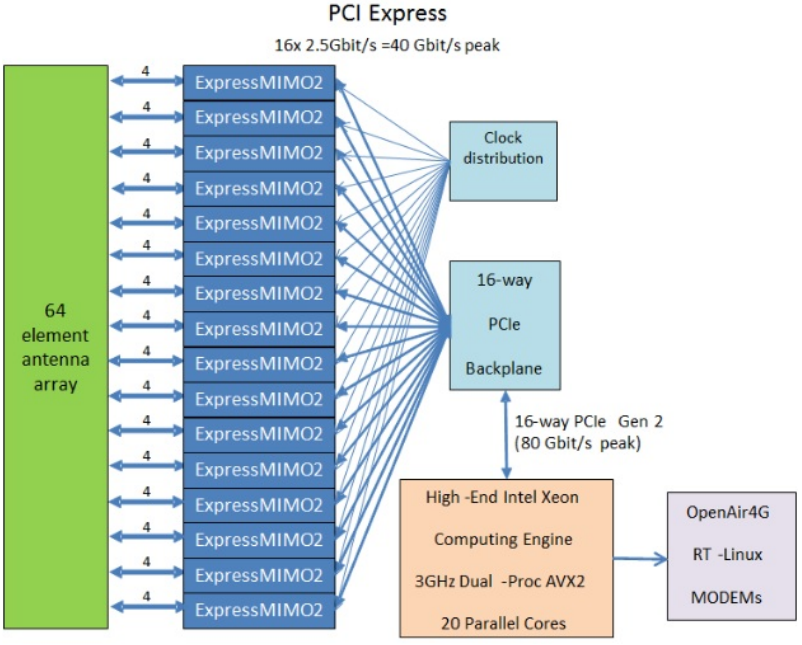


Figure 10.5: OpenAirInterface massive MIMO testbed architecture.

Let us look at the architecture in Fig. 10.5 from left to right. The 64 element antenna array is driven by 16 ExpressMIMO2 RF cards. Each card has 4 RF chains and can thus be connected to 4 independent antennas using SubMiniature version A (SMA) cables. In order to achieve both frame and frequency synchronization on all cards, we use Ettus Research's Octo-clock as an external clock distributor. The first ExpressMIMO2 card is selected as the master card who generates a 61.44MHz clock signal and a pulse per frame (every 10ms) signal as the reference. These two signals are used as external input of

Octo-clock who then outputs same signals to drive the rest 15 ExpressMIMO2 cards. The data are transmitted via PCIe interface between the ExpressMIMO2 cards and the Xeon Server. A 16-way PCIe backplane is needed to serve as the PCIe extension of the server. In order to handle the high signal processing demands, we chose to use a 20 parallel core server.

10.4.2 Hardware components

Each hardware component used in the testbed is detailed here below:

1. **Huawei antenna array**

   The planar antenna array in Fig. 10.6, provided by our partner Huawei Technologies (Paris), is composed of 20 patch antennas with 4 $\lambda/2$ spaced "H"-shaped antenna elements, optimized for the 3GPP band 38 (2.6GHz, TDD, 50MHz bandwidth). The patches are mounted on a rack with 4 antenna rows, each with adjustable height and tilt, as well as the possibility of moving patches in each row, thus offering a great flexibility for antenna element arrangement. The current arrangement in Fig. 10.6, with 4 patches in the first and last low placed on two sides and 6 patches in each of the two middle rows, forms a big "H" shape. This is mainly designed for a marketing purpose to promote "Huawei" rather than for performance reasons.

Figure 10.6: Huawei antenn array.

2. **ExpressMIMO2 card**

   ExpressMIMO2 card (Fig. 10.7) is a low cost hardware target enabling experimentation with OpenAirInterface. It is developed by OpenAir5GLab@EURECOM and can be used by OpenAirInterface soft-modem to drive up to 4 parallel RF chains with up to 20 MHz bandwidth in the range of 350-3800 MHz. It interconnects with a baseband computing engine using Gen 1 1-way PCIe (2.5 Gbit/s peak full-duplex bi-directional throughput). The board is built around a low-cost Spartan-6 FPGA (150LXT) with native PCIexpress on the FPGA fabric and coupled with 4 high-performance LTE RF ASICs, manufactured by Lime Micro Systems (LMS6002D).

The combination allows for four full-duplex or half-duplex radios to be interfaced with a desktop or laptop PC without the need for external RF.

Figure 10.7: ExpressMIMO2 card.

3. **Octo-clock**

We cascaded two Ettus Research's Octo-clock [74] in our testbed. A 61.44MHz clock signal and a pulse per frame (every 10ms) signal are generated by the master card and serves as the external input of one Octo-clock, whose one output is used to drive the other Octo-clock. These signals are then amplified to synchronize the slave cards.

Figure 10.8: Ettus Research's Octo-clock.

4. **PCIe chassis**

The testbed uses Magma's ExpressBox 16 PCIe backplane for multiplexing 16 Gen 1 1-way PCIe (40 GBit/s peak) into a single 16-way Gen2 PCIe link providing a peak data rate at 80Gbit/s [88]. It is used to host 16 ExpressMIMO2 RF cards, and is in charge of the communication between those cards with the Intel Xeon server.

Figure 10.9: Magma's ExpressBox 16 PCIe backplane.

5. **High-end Xeon Server**

   We use a high-end 20-core Xeon server (10-core dual-processor 3 GHz) with AVX2 instructions running a Linux real-time OS with low-latency kernel and OpenAirInterface LTE baseband soft-modem.

10.5 Software

This section presents the software implementation of the testbed.

10.5.1 OpenAirInterface

In the testbed, OpenAirInterface soft-modem runs on the Xeon server and drives the massive MIMO hardwares. OpenAirInterface is an open source standard-compliant implementation of a subset of Release 10 LTE for UE, eNB, MME, HSS, SGw and PGw on standard Linux-based computing equipment (Intel x86 PC/ARM architectures) [77]. OpenAirInterface comprises of two repositories: OpenAirInterface5G and OpenAirInterfaceCN, dedicated to the radio access network (RAN) and the core network (CN) implementation respectively. OpenAirInterface5G can be used for simulation/emulation, as well as real-time experimentation on off-the-shelf SDR cards, like the aforementioned ExpressMIMO2 card but also the popular USRP from National Instruments/Ettus, LimeSDR, BladeRF, and other RF platforms. It comprises of the fully compliant LTE protocol stack from the physical to the networking layer and can inter-operate with commercial LTE terminals and can be interconnected with OpenAirinterfaceCN or closed-source EPC (Enhanced Packet Core) solutions from third-parties. The objective of this platform is to provide methods for protocol validation, performance evaluation and pre-deployment system tests.

The OpenAirInterface software stack is shown in Fig. 10.10, where we use OAI as an acronym of for OpenAirInterface. The OAI soft UE and OAI soft eNB are components in OpenAirInterface5G whereas OAI soft EPC, including MME, HSS, SGw and PGw are component in OpenAirInterfaceCN. One OAI EPC can drive multiple OAI eNB. Each eNB can inter-operate with multiple OAI UEs. In Fig. 10.10, the white blocks are 3GPP layers that OpenAirInterface has implemented and the purple blocks are linux stack. On top of them, eNB, MME, SGw and PGw applications are created to ease the configuration and management. The control and data plan are differentiated by the blue and red lines linking different software components.

The software implementation of the massive MIMO testbed mainly lies in the physical layer on the eNB and UE side. eNB RRC layer is extended to allow the TM signaling between eNB and UE.

10.5.2 Integrating massive MIMO into LTE and 5G

The software extension on OpenAirInterface in order to support massive antenna array can be reprensented by the schema in Fig. 10.11. The first part is the off-line TDD channel reciprocity calibration, which is implemented in Octave; the second part consists

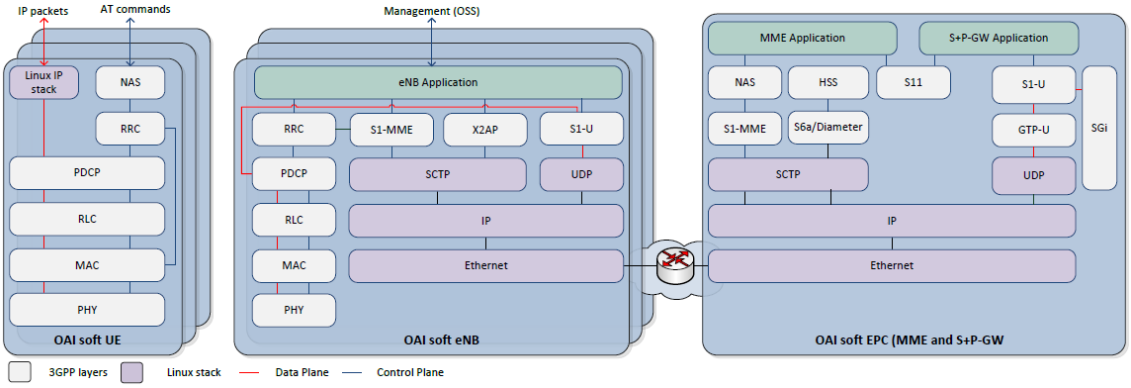


Figure 10.10: OpenAirInterface LTE software stack.

in extending the real time soft-modem to use the result from calibration and perform TM7 beamforming. In this section we detail the actions in each of these operation blocks.

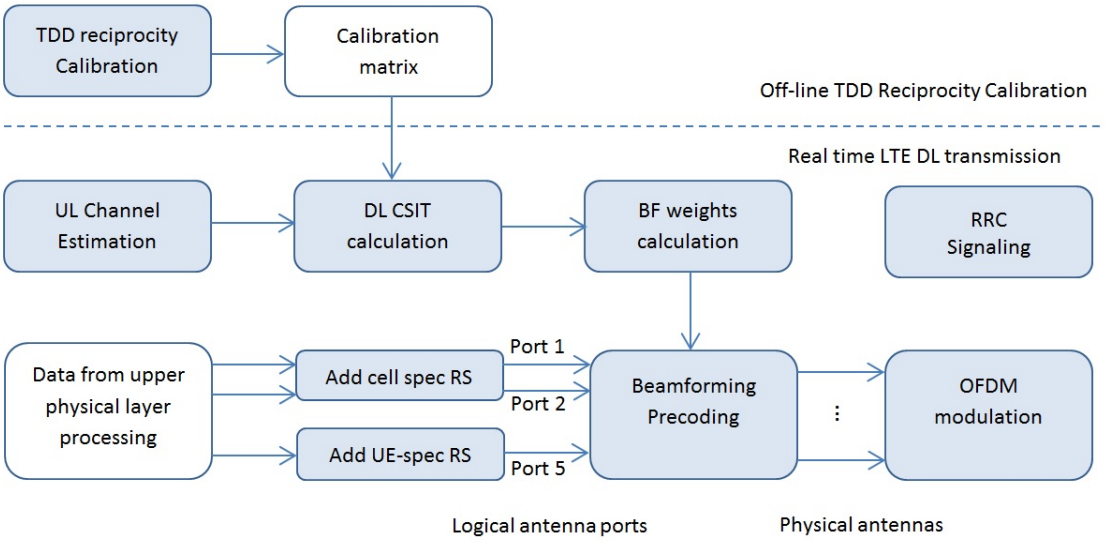


Figure 10.11: OpenAirInterface massive MIMO software implementation.

- **TDD reciprocity calibration**

  Using Octave, we have implemented various BS internal calibration methods, including group calibration as in Section 5.1, Argos calibration [25] and LS calibration [26]. Based on the bi-directional transmissions between antenna pairs or between group pairs of antennas, a calibration matrix including the calibration coefficients for each RF chain on all frequency carriers are estimated and stored in a file under Q1.15 format, which will later be read and used to obtain DL CSIT. After the calibration procedure, we switch to the OpenAirInterface real time mode. The RF Express-MIMO2 card should not be reset so that the calibration coefficients keeps unchanged during this switch.

- **UL channel estimation**

  In LTE, two types of RSs are available for UL channel estimation [34]:

- DeModulation RS (DM-RS), associated with transmissions of uplink data on the Physical Uplink Shared CHannel (PUSCH) and/or control signaling on the Physical Uplink Control CHannel (PUCCH). These RSs are primarily used for channel estimation for coherent demodulation.
- Sounding RS (SRS), not associated with uplink data and/or control transmissions, and primarily used for channel quality determination to enable frequency-selective scheduling on the uplink.

For a given UE, the DM-RS are only used in the bandwidth (i.e. Resource Blocks (RBs)) allocated to PUSCH/PUCCH, resulting in the fact that the BS can thus hardly get the whole bandwidth channel information based on it. On the contrary, it is possible to configure the SRS from higher layers to occupy the whole bandwidth for UL channel sounding. We thus implemented the SRS configuration and estimation mechanism. Each physical antenna at the BS can estimate the UL channel from the UE based on its SRS occupying the whole bandwidth.

- **DL CSIT calculation**

  In this operation, we read the calibration matrix from the file generated during the Octave reciprocity calibration, which are then applied to the instantaneously estimated UL CSI in order to assess the DL CSIT. For a 5MHz bandwidth system, 300 useful sub-carriers (i.e. resource elements (REs)) are used. In order to respect the real time operation constraint and to speed up this operation, SIMD techniques [89] are used on the sub-carriers.

  As we rely on the TDD channel reciprocity to acquire DL CSIT, we assume that the channel coherence time is higher than the frame duration (10ms). High mobility scenario is thus not in our consideration. The obtained CSIT during the UL subframes can be used to perform beamforming for the DL subframes until a new DL CSIT is calculated.

- **Beamforming weights calculation**

  Once the BS has acquired accurate enough DL CSIT, it is possible to apply different beamforming algorithms. Currently, we calculate the beamforming weights based on MRT since 1) in TM7, only one UE is considered, MRT is the optimal precoding strategy; 2) in terms of complexity, MRT consists in a simple conjugation operation on the CSIT, thus is easier to satisfy the real time constraint. Note that it is also possible to extended the current implementation to other precoding algorithms, such as ZF or MMSE, under the challenge that the introduced high complexity might make it difficult to accomplish the weights calculation during the UL period of the TDD frame.

- **Adding Cell specific RS and UE specific RS**

  In LTE DL, there exist five different types of RSs, where mainly two of them are concerned in our application: cell-specific RS and UE-specific RS. Cell specific RSs are used for channel estimation in antenna port 0-3. They are independent of the data MIMO precoding, and thus are common to all users. When beamforming schemes (TM 7-10) are used, UE specific RSs are needed to perform beamformed channel estimation at each user. UE specific RSs only exist where the OFDM frequency-time RBs are allocated for data transmission. They are precoded using the same beamforming weights as the data for each user, and are specially used for beamforming

data demodulation. Note that another type of RS - the CSI RS is introduced in Release 10, specifically for the purpose of obtaining channel state feedback for up to eight transmit antenna ports to assist the eNB in its precoding, however for our reciprocity based CSIT acquisition system, we didn't implement it.

Both Cell-specific and UE-specific RSs use quadrature phase-shift keying (QPSK) - a constant modulus modulation, in order to ensure that the Peak-to-Average Power Ratio (PAPR) of the transmitted waveform is kept low. The signal is generated based on a length-31 Gold sequence with different initialization values depending on the type of RS [34].

In our implementation, cell specific RSs are applied on antenna port 1 and 2 (or only antenna port 1), whereas UE specific RS is used on antenna port 5 to estimate the beamformed data channel in TM7. The arrangement of the DL RS in a time-frequency RB is illustrated in Fig. 10.12.

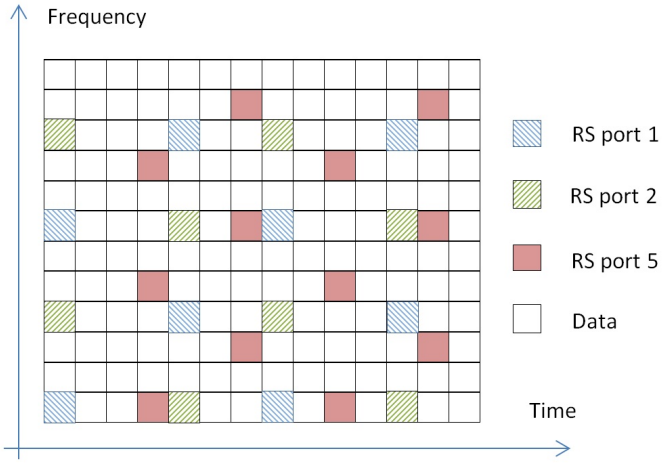


Figure 10.12: The position of cell specific and UE specific RS in a RB. The blue and green REs are the cell-specific RS for antenna port 1 and 2. The red RE is the UE-specific RE for antenna port 5 whereas the white RE is the data resource element.

- **Beamforming precoding and OFDM modulation**

  Beamforming precoding consists in multiplying the transmit signal with a different weight for each physical antenna, it precedes the OFDM modulation using the IFFT operation. Data on different logical antenna ports are mapped onto different elements in the physical antenna array. In the testbed, system and control information on the antenna port 1 (and 2) are mapped to the physical antenna array with cell-specific beamforming weights whereas data on antenna port 5 are mapped to the physical antennas using UE specific beamforming weights in order to beam the data to the target UE. This mapping method is shown in Fig. 10.13 and Fig. 10.14 for TM7 and TM8, respectively.

  In an OFDM system with N_{freq} sub-carriers and N_{ant} antennas, a total number of $N_{freq}N_{ant}$ complex value multiplication is needed for the beamforming precoding per OFDM symbol. Afterwards, the corresponding IFFT operations should be performed independently for N_{ant}. For an LTE system running in real time, it is essential to keep the whole processing time for one sub-frame (14 OFDM symbols in normal

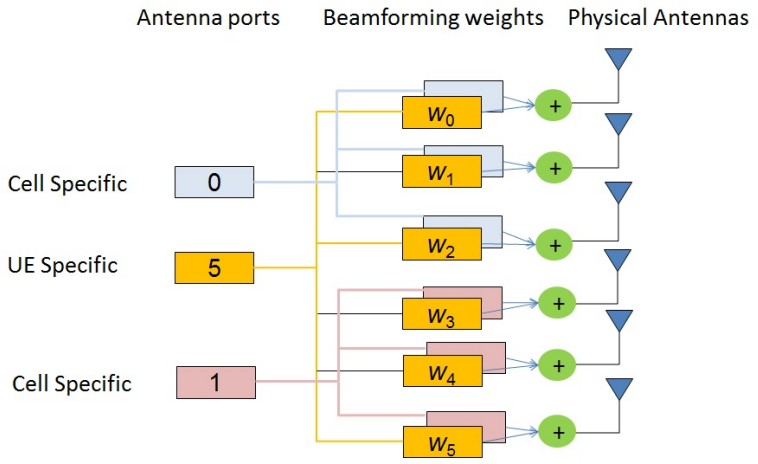


Figure 10.13: Logical antenna ports to physical antenna mapping in TM7 where antenna port 1 and 2 are mapped with cell specific beamforming weights and antenna port 5 is mapped with UE specific beamforming weights.

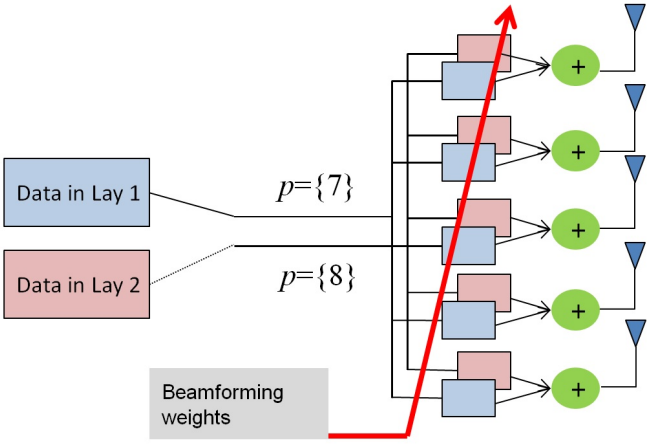


Figure 10.14: Logical antenna ports to physical antenna mapping in TM8 (for data), where the two layer data on antenna port 7 and 8 are mapped with different UE specific beamforming weights.

cyclic prefix (CP) case and 12 OFDM symbols in extended CP case) under one sub-frame duration (1ms). When the number of antennas N_{ant} grows, it becomes very difficult to meet this time constraint as the processing time of beamforming precoding and OFDM modulation increases linearly with the antenna array.

In order to make the real time processing possible, two mechanisms of parallelization are used. In the frequency domain, we use the SIMD to parallelize the beamforming precoding on adjacent sub-carriers which can in theory reduce the whole processing time by four. In the antenna space domain, we create a thread pool, where each thread takes in charge the precoding and OFDM modulation for one physical antenna, as shown in Fig. 10.15. The thread pool contains N_{ant} independent threads and has three different status. When data and control informations are ready for one time slot, we wake up all threads in the pool. The precoding and OFDM modulation for different physical antennas will run in parallel in order to reduce the processing

time. The operations in those independent threads finish one after another, and when all threads have accomplished their work, the thread pool goes into a waiting status for the next slot.

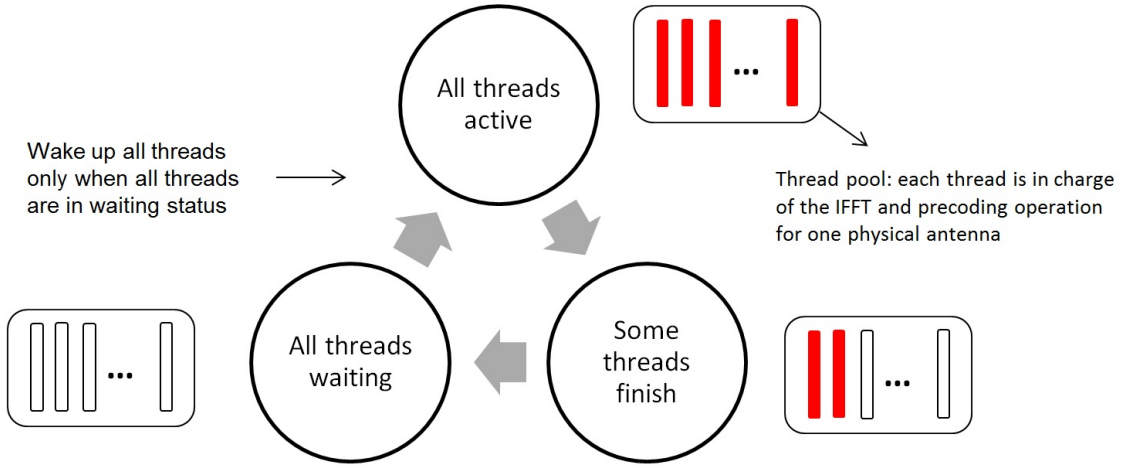


Figure 10.15: Thread pool for parallelizing the beamforming precoding and OFDM modulation on different physical antennas.

In order to compare the calculation time with and without the thread pool, we measure the execution time of the total beamforming precoding and OFDM modulation for the whole antenna array using a unitary simulator - "dlsim", dedicated to physical layer DL simulation. The simulation is performed on a x86_64 architecture machine with 4 cores. The results are summarized in Tab. 10.2 where we observe that the processing time for a single thread (i.e. thread pool is not used) increases linearly with the number of transmit antennas, as the processing is executed for antennas one by one. When thread pool is activated, the processing time for a single antenna is 168.88us, higher than the single thread case (42.45us). This is due to the overhead introduced by managing the thread pool. However, when the number of antennas increases, thanks to the parallelization, the total processing time does not increase with the number of antennas. Normally with a four core machine, at most 4 processing threads can be parallelized, however, we observe that processing for 16 Tx antennas consumes almost the same time as two antennas. This a probably due to the fact that the main time consuming part is still the thread pool management from 1 to 16 antennas. When we have 64 antennas, we have almost 4 times the processing time as the case of 16 antennas.

- **RRC (Radio Resource Control) signaling**

  In order to use transmission modes other than TM1 or TM2, the eNB needs to configure the designated transmission mode to the UE via RRC message whenever the UE establishes an RRC connection (initial connection to an LTE network, a reestablishment of an RRC connection after some kind of radio link failure or other radio connection failure or after handover to an Evolved Universal Terrestrial Radio Access Network (EUTRAN) cell). In our implementation, eNB configures the UE to use TM7 via RRCConnectionReconfiguration Message as illustrated in Fig. 10.16. Before that all the DL transmission are under TM1 or TM2 according to the number of antenna ports in Physical Broadcast Channel (PBCH). After the configuration,

Number of Tx antennas	Single thread (us)	Thread pool (us)
1	42.45	168.88
2	83.87	210.79
4	167.18	263.07
8	330.91	219.18
16	662.49	244.68
64	2645.33	839.63

Table 10.2: The execution time of beamforming precoding and OFDM modulation for different number of transmit antennas when we use single thread or a thread pool. The execution time is measured using "dlsim" simulator on a x86_64 architecture machine with 4 cores.

eNB transmits DL data using TM7, where beamforming using the massive antenna array is possible.

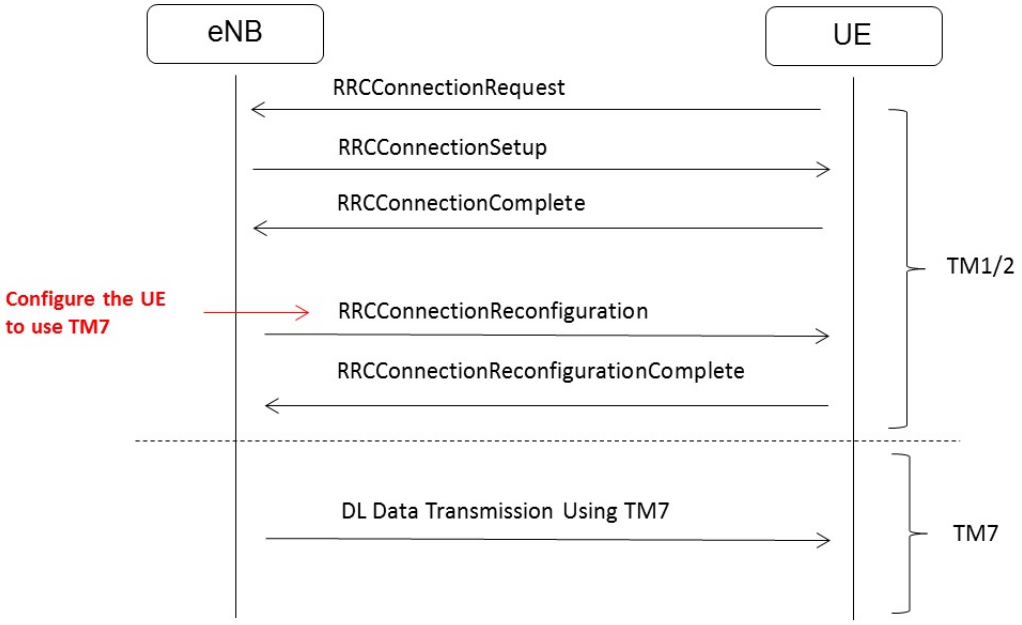


Figure 10.16: RRC message exchange between the eNB and UE during the connection setup. Transmission mode is configured in the RRCConnectionReconfiguration message. If the reconfiguration is successful, the downlink transmission switches from TM1 or TM2 to TM7.

10.6 Results and dissemination

In order to validate the TM7 physical layer implementation in OpenAirInterface, we connect the Rohde & Schwarz FSQ signal analyzer [90] to the output of one RF chain. Since the FSQ signal analyzer initially does not support TM7, we need to use a laptop having the FS-K96 OFDM vector signal analysis software [91] installed to drive and extend the capability of FSQ signal analyzer. The screen-shot of the FS-K96 vector signal analysis

software is illustrated in Fig.10.17, where we can see that both the Physical Downlink Shared Channel (PDSCH) (which carries data) and the UE specific RS are successfully decoded.

Figure 10.17: Validation of the TM7 implementation using Rohde & Schwarz FSQ Signal Analyzer

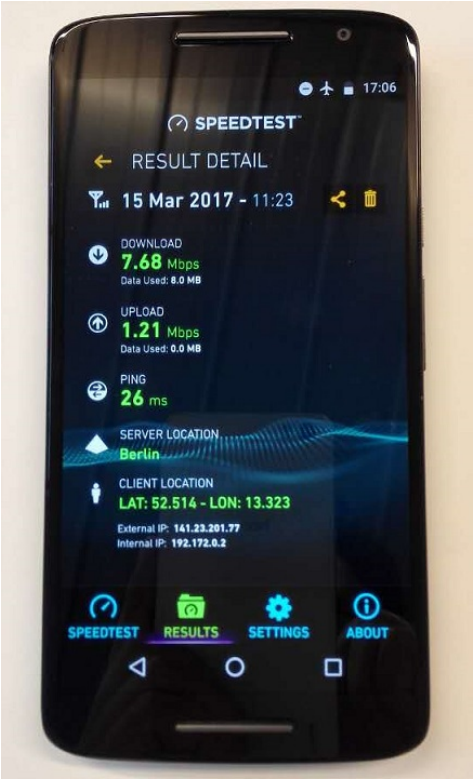


Figure 10.18: Throughput test result of TM1 using "SPEEDTEST"

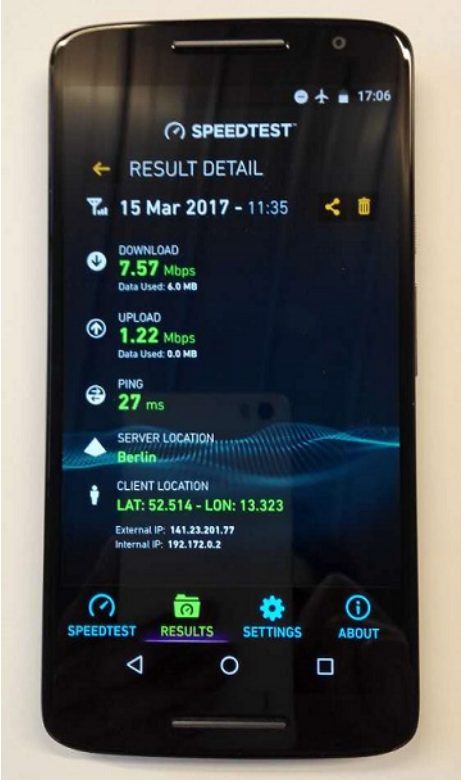


Figure 10.19: Throughput test result of TM7 using "SPEEDTEST"

The full protocol stack test is performed using "SPEEDTEST", an application for data

throughput measurement, installed on a smart-phone (Motorola X4). We compare the throughput of TM7 with TM1, which is the most basic and stable TM implementation in OpenAirInterface. The result is shown in Fig. 10.18 and Fig. 10.19. We can see that both TMs have a throughout around 7.6Mbps for the DL transmission, implying that the implementation of TM7 works well. We also notice that there are still some unsolved issues in the UL, leaving the throughput around 1.2Mbps, which is quite low.

Note that different with other massive MIMO testbeds using higher older modulation schemes (thanks to beamforming) and wider bandwidth to achieve high data rates, the data throughput of OpenAirInterface massive MIMO testbed is limited by the highest modulation and coding scheme (MCS) in LTE, the limited system bandwidth (5MHz), as well as the payload consumption on different protocol layers. As TM7 supports only one scheme transmission, it has a similar performance as TM1.

During the development of the testbed, we have successfully demonstrated our work in different conferences. In June, 2016, a reduced scale version of the testbed with 16 antennas, as in Fig. 10.20, mainly showing the the performance of TDD reciprocity calibration in a massive MIMO system was successfully shown in the European Conference on Networks and Communications (EuCNC), in Athens, Greece. Together with other demonstrators under the "Advanced Dynamic Spectrum 5G mobile networks Employing Licensed shared access (ADEL)" project [92], we have won the best booth award, after competing with other 31 European projects (Fig. 10.21).

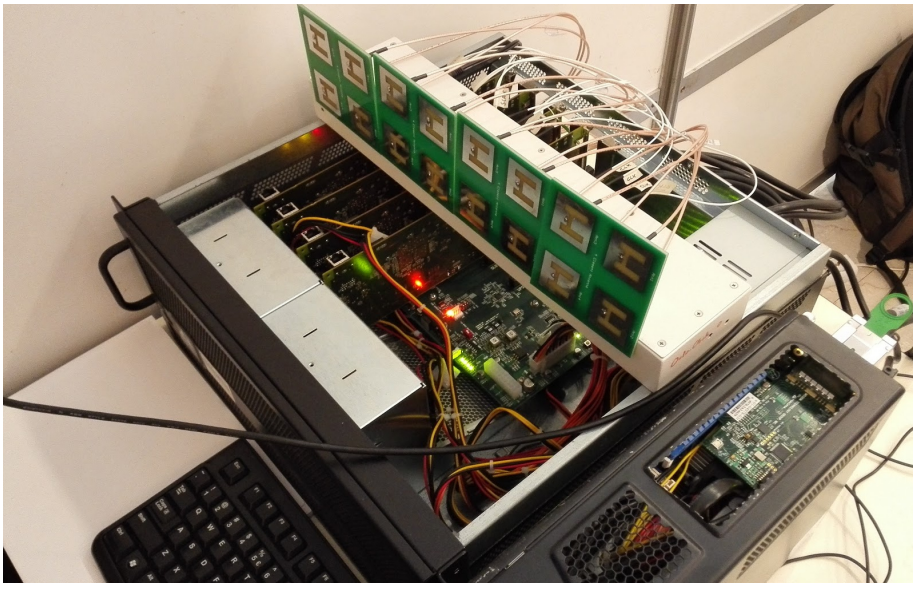


Figure 10.20: The reduced scale version of the OpenAirInterface massive MIMO testbed with 16 antennas demonstrated in EuCNC. It mainly shows the performance of TDD reciprocity calibration and the possibility to acquire near perfect CSIT in a massive MIMO system.

In March, 2017, we successfully demonstrate the interoperability of a scaled-down version (mainly for transport reasons) of the testbed with a commercial smart-phone (Motorola X4) in the 21st International ITG Workshop on Smart Antennas (WSA 2017), held in Berlin, Germany, as shown in Fig. 10.22. CSIT is acquired based on TDD channel reciprocity calibration. MRT beamforming is used to provide real time Internet service to the commercial smart-phone.

Figure 10.21: Best booth award for ADEL project, showcasting (amongst others) the first version of Eurecom's massive MIMO demonstrator. (Picture ©by Constantinos Papadias)

Figure 10.22: A scaled-down version of massive MIMO prototype with 4 antennas demonstrated in WSA Berlin showing the interoperability with commercial UE. TDD reciprocity calibration is used for CSIT acquisition. The system then performs MRT beamforming to the UE for Internet service.

10.7 Summary

This chapter presented the development work of the OpenAirInterface massive MIMO testbed. Using 16 ExpressMIMO2 cards synchronized by Octo-clock, the testbed is able to drive up to 64 antenna elements. We extended the OpenAirInterface software to support TM7 transmission. Logical antenna ports are mapped to the physical antenna array to create a narrow beam towards the served UE. UE specific RS is implemented for channel es-

timation under beamforming transmission. Based on TDD channel reciprocity calibration, the testbed is able to acquire very accurate DL CSIT, based on which, MRT beamforming weights are calculated.

We successfully demonstrated the interoperability between the testbed BS and a commercial UE using TM7, showing the feasibility to integrate massive MIMO into current LTE standard. The testbed is on continuous evolution to be able to simultaneously serve more UEs using higher order TMs in current and coming 3GPP releases.

Chapter 11

Conclusion and Future Work

This thesis addresses the problem of bringing massive MIMO from a theoretical concept to practical systems, with a special focus on TDD channel reciprocity calibration. Correctly and efficiently calibrate the Tx and Rx RF asymmetry is essential for reciprocity based CSIT acquisition in TDD massive MIMO systems.

We propose a pilot based general framework for "over-the-air" calibration. Various existing calibration methods in literature, especially BS internal calibration, suitable for massive MIMO systems, can be represented under this general framework using different antenna partition. Optimal estimation and CRB (as a performance benchmark) are derived. We also discussed coherent and non-coherent accumulation and point out that it is possible to integrate the calibration process into data service, so that the process consumes a vanishing resources.

The general framework opens up many possible innovations on reciprocity calibration. First, performing pilot exchanges using groups of antennas rather than using individual elements can speed up the calibration process. We showed that, to calibrate an antenna array with M antenna using coherent accumulation, the minimal channel uses needed is on the order of $\mathcal{O}(\sqrt{M})$ rather than $\mathcal{O}(M)$ for other existing methods in literature. Second, for a hybrid analog-digital beamforming structure system being operated in TDD mode, we propose to partition the antenna array into two groups and calibrate the system using bi-directional pilot transmission between them[1]. A reciprocity enabled hybrid beamforming system has the possibility to acquire near perfect DL CSIT, and thus significantly outperforms the case where we operate the system using traditional ways by beam training based on pre-defined beam set. Third, we also propose a calibration method that can follow the evolution of calibration coefficients. It provides a mechanism to monitor the system in a calibrated status and when there is a sporadic sudden change, it allows for a fast detection on the changed parameter.

In addition to those new calibration methods, we also carried out a measurement campaign for reciprocity calibration coefficients on real RF hardwares. The results reveal those coefficients' properties and verified, in a small scale MISO system, the widely adopted diagonal assumption on the calibration matrix in literature. We also studied the accuracy of obtained CSIT impacted by the accuracy of the calibration matrix and the instantaneously estimated UL channel.

[1]This method is valid for a sub-array architecture. For fully connected architecture, the bi-directional transmission should be performed with an assistant device or a UE

Enabled by TDD reciprocity calibration, we built up an LTE compatible massive MIMO testbed on the OpenAirInterface platform. It successfully inter-operates with commercial devices, thus demonstrates the possibility of integrating massive MIMO in current 4G systems. It also shows the feasibility of using calibration for CSIT acquisition in a real world massive MIMO testbed. The implementation identifies other challenges for system building and propose corresponding solutions.

While some interesting results have been established in this thesis, some problems are left unsolved. In Chapter 5 and 6, we compared the calibration performance of interleaved and non-interleaved antenna grouping. However, the best way to group up antennas is still an open question. Moreover, under the unified calibration framework, another dimension of innovation lies in the pilot design. We did not fully explore this topic in the current thesis, but simply use a constant amplitude Fourier pilot in most simulations.

In terms of experiment and measurement, one essential question is the intra-array channel modeling within the antenna array. The near field channel between two elements in the same antenna array strongly depends on the antenna type and antenna elements' installation. Real world measurements are needed to reveal the properties and to establish a proper model. Some interesting results are reported in [28]. We are also working hard on our side to push forward the state of the art on this research. Another question concerns the RF crosstalk and antenna mutual coupling. In Chapter 8, we verified the diagonal assumption on the calibration matrix in a small scale MIMO system. Although theoretical modeling in [19], as well as practical experience in most literature [25, 28] all point out that RF crosstalk and antenna mutual coupling can be ignored, as we've assumed in Chapter 5-7, it is still of high value to scale up the experiment in Chapter 8 to verify this assumption in a massive MIMO context.

From the point of view of testbed development and system implementation, there are multiple aspects that need to be done to evolve the current version. Today, the testbed can only create a narrow beam to a single UE with TM7. To simultaneously serve multiple UEs, development efforts are needed to enable TM8, TM9 or new TMs that might appear in future 3GPP standards. Moreover, reciprocity calibration is performed using Octave scripts when the system is initialized, it is interesting to see how this process can be integrated into the real time operation. Moreover, the co-localized massive MIMO performs signal processing in one PC, leaving it a challenging issue to handle all calculations locally. Evolving this architecture to a distributed topology, thus being aligned with the C-RAN architecture, can allow for assigning a part of the calculation tasks to the cloud whereas other tasks to distributed radio units, which can relax the real time calculation constraint in the testbed.

Bibliography

[1] Huawei, "5g : A technology vision." `http://www.huawei.com/5gwhitepaper/`, 2013.

[2] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *4th IEEE Asilomar Conf. on Sig., Sys. and Comp., (ACSSC)*, pp. 359–363, 2006.

[3] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010.

[4] G. Smith, "A direct derivation of a single-antenna reciprocity relation for the time domain," *IEEE Trans. on Antennas and Propagation*, vol. 52, pp. 1568–1577, June 2004.

[5] H. Yin, L. Cottatellucci, and D. Gesbert, "Enabling massive MIMO systems in the FDD mode thanks to D2D communications," in *48th Asilomar Conf. on Sig., Sys. and Comp.*, pp. 656–660, IEEE, 2014.

[6] A. Decurninge, M. Guillaud, and D. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Globecom Workshop on Massive MIMO: From theory to practice*, 2015.

[7] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, pp. 186–195, Feb. 2014.

[8] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Sel. Areas in Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[9] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE Journal of Sel. Topics in Sig. Process.*, vol. 8, no. 5, pp. 773–786, 2014.

[10] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Inter. Conf. on Acoustics, Speech and Sig. Process. (ICASSP)*, pp. 3249–3252, 2012.

[11] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Intern. Symp. on Info. Theory (ISIT)*, pp. 1137–1141, 2012.

[12] J. C. Guey and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode," in *Proc. IEEE 60th Veh. Technol. Conf. (VTC)*, vol. 6, pp. 4265–4269, 2004.

[13] X. Luo, "Multi-user massive MIMO performance with calibration errors," *IEEE Trans. on Wireless Commun.*, vol. 15, July 2016.

[14] W. Zhang, H. Ren, C. Pan, M. Chen, R. C. de Lamare, B. Du, and J. Dai, "Large-scale antenna systems with UL/DL hardware mismatch: achievable rates analysis and calibration," *IEEE Trans. on Commun.*, vol. 63, no. 4, pp. 1216–1229, 2015.

[15] X. Jiang, F. Kaltenberger, and L. Deneire, "How accurately should we calibrate a massive MIMO TDD system?," in *Proc. IEEE Intern. Conf. on Commun. (ICC) Workshops*, 2016.

[16] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: impact and mitigation," in *Proc. IEEE Radio and Wireless Conf. (RAWCON)*, (Boston, MA, USA), pp. 183–186, Aug. 2003.

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

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

[19] 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.*, vol. 61, pp. 3781–3793, Sep. 2013.

[20] G. Benzin, Andreas; Caire, "Internal self-calibration methods for large scale array transceiver software-defined radios," in *21th International ITG Workshop on Smart Antennas (WSA)*, (Berlin, Germany), Mar. 2017.

[21] M. Guillaud, D. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. Intern. Symp. Signal Process. and Its Applications (ISSPA)*, (Sydney, Australia), pp. 403–406, Aug. 2005.

[22] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Proc. Future Netw. and Mobile Summit*, (Florence, Italy), pp. 1–10, Jun. 2010.

[23] J. Shi, Q. Luo, and M. You, "An efficient method for enhancing TDD over the air reciprocity calibration," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, pp. 339–344, 2011.

[24] B. Kouassi, I. Ghauri, B. Zayen, and L. Deneire, "On the performance of calibration techniques for cognitive radio systems," in *Proc. IEEE Wireless Personal Multimedia Commun. (WPMC)*, pp. 1–5, Oct. 2011.

[25] C. Shepard, H. Yu, N. 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), pp. 53–64, Aug. 2012.

[26] 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 Commu.*, vol. 13, pp. 1815–1831, Apr. 2014.

[27] J. Vieira, F. Rusek, and F. Tufvesson, "Reciprocity calibration methods for massive MIMO based on antenna coupling," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, (Austin, USA), pp. 3708–3712, 2014.

[28] J. Vieira, F. Rusek, O. Edfors, S. Malkowsky, L. Liu, and F. Tufvesson, "Reciprocity calibration for massive MIMO: Proposal, modeling, and validation," *IEEE Trans. on Wireless Comm.*, vol. 16, no. 5, pp. 3042–3056, 2017.

[29] H. Papadopoulos, O. Y. Bursalioglu, and G. Caire, "Avalanche: Fast RF calibration of massive arrays," in *Proc. IEEE Global Conf. on Signal and Information Process. (GlobalSIP)*, (Washington, DC, USA), pp. 607–611, Dec. 2014.

[30] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[31] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, pp. 40–60, Jan. 2013.

[32] R. Gallager, "Circularly-symmetric gaussian random vectors," *preprint*, 2008.

[33] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[34] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS Long Term Evolution: from theory to practice*. John Wiley & Sons, 2011.

[35] H. Ji, Y. Kim, J. Lee, E. Onggosanusi, Y. Nam, B. Zhang, J.and Lee, and B. Shim, "Overview of Full-Dimension MIMO in LTE-Advanced pro," *IEEE Comm. Mag.*, vol. 55, no. 2, pp. 176–184, 2017.

[36] F. Huang, J. Wang, Y.and Geng, and D. Yang, "Antenna mismatch and calibration problem in coordinated multi-point transmission system," *IET communications*, vol. 6, no. 3, pp. 289–299, 2012.

[37] S. Han, C. Yang, G. Wang, D. Zhu, and M. Lei, "Coordinated multi-point transmission strategies for TDD systems with non-ideal channel reciprocity," *IEEE Trans. on Comm.*, vol. 61, no. 10, pp. 4256–4270, 2013.

[38] V. Jungnickel, V. Krüger, G. Istoc, T. Haustein, and C. von Helmolt, "A MIMO system with reciprocal transceivers for the time-division duplex mode," *Proc. IEEE Antennas and Propagation Society Symp.*, vol. 2, pp. 1267–1270, Jun. 2004.

[39] C. A. Balanis, *Antenna theory: analysis and design*. John Wiley & Sons, 2016.

[40] H. Wei, D. Wang, H. Zhu, J. Wang, S. Sun, and X. You, "Mutual coupling calibration for multiuser massive MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 1, pp. 606–619, 2016.

[41] H. Wei, D. Wang, and X. You, "Reciprocity of mutual coupling for TDD massive MIMO systems," in *Proc. Intern. Conf. on Wireless Commun. and Sig. Process. (WCSP)*, (Nanjing, China), pp. 1 – 5, Oct. 2015.

[42] X. Jiang, M. Čirkić, F. Kaltenberger, E. 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), pp. 4949–4953, Jun. 2015.

[43] B. Kouassi, L. Deneire, B. Zayen, R. Knopp, F. Kaltenberger, F. Negro, D. Slock, and I. Ghaur, "Design and implementation of spatial interweave LTE-TDD cognitive

radio communication on an experimental platform," *IEEE Wireless Comm.*, vol. 20, no. 2, pp. 60–67, 2013.

[44] M. Guillaud and F. Kaltenberger, "Towards practical channel reciprocity exploitation: Relative calibration in the presence of frequency offset," in *IEEE Wireless Commun. and Netw. Conf. (WCNC)*, (Shanghai, China), 2013.

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

[46] 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.

[47] J. Vieira, S. Malkowsky, Z. Nieman, K.and Miers, N. Kundargi, L. Liu, I. Wong, V. Owall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive MIMO," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshops*, (Austin, USA), pp. 287–293, 2014.

[48] X. Luo, "Robust large scale calibration for massive MIMO," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, (San Diego, CA, USA), pp. 1–6, December 2015.

[49] C. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 167–180, 1968.

[50] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. on Sig. Process.*, vol. 55, no. 6, pp. 2740–2746, 2007.

[51] R. Rogalin, O. Y. Bursalioglu, H. C. Papadopoulos, G. Caire, and A. F. Molisch, "Hardware-impairment compensation for enabling distributed large-scale MIMO," in *Proc. Information Theory and Applications (ITA) Workshop, 2013*, (San Diego, California, USA.), pp. 1–10, Feb. 2013.

[52] G. Giannakis, Y. P. Stoica, Hua, and L. Tong, *Signal Processing Advances in Communications*, ch. 7. Prentice Hall, Jan. 2000.

[53] E. de Carvalho, S. Omar, and D. Slock, "Performance and Complexity Analysis of Blind FIR Channel Identification Algorithms Based on Deterministic Maximum Likelihood in SIMO Systems," *Circuits, Sys., and Sig. Process.*, vol. 34, Aug. 2012.

[54] E. d. Carvalho and D. Slock, "Blind and Semi-Blind FIR Multichannel Estimation: (Global) Identifiability Conditions," *IEEE Trans. on Sig. Proc.*, Apr. 2004.

[55] E. d. Carvalho, J. Cioffi, and D. Slock, "Cramer-Rao bounds for blind multichannel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, (San Francisco, CA, USA), pp. 1036–1040, Nov. 2000.

[56] Z. Jiang and S. Cao, "A novel TLS-based antenna reciprocity calibration scheme in TDD MIMO systems," *IEEE Commun. Letters*, vol. PP, no. 99, 2016.

[57] C. Kim, T. Kim, and J. Seol, "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in *Proc. Globecom Workshops (GC Wkshps)*, pp. 61–65, 2013.

[58] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, 2015.

[59] J. Flordelis, F. Rusek, F. Tufvesson, E. Larsson, and O. Edfors, "Massive MIMO performance-TDD versus FDD: What do measurements say?," *arXiv preprint arXiv:1704.00623*, 2017.

[60] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Sel. Topics in Sig. Proc.*, vol. 10, no. 3, pp. 501–513, 2016.

[61] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Sel. Topics in Sig. Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[62] J. Li, F. Huang, R. Zhou, W. Chen, Z. Tian, and S. Zhou, "mmWave mobile communication under hypercellular architecture," *Journal of Comm. and Info. Netw.*, vol. 1, no. 2, pp. 62–76, 2016.

[63] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for ofdm-based large-scale MIMO systems," in *Proc. Intern. Workshop on Sig. Process. Adv. in Wireless Commun. (SPAWC)*, pp. 1–6, 2016.

[64] X. Huang, Y. J. Guo, and J. D. Bunton, "A hybrid adaptive antenna array," *IEEE Trans. on Wireless Comm.*, vol. 9, no. 5, pp. 1770–1779, 2010.

[65] Y. J. Guo, X. Huang, and V. Dyadyuk, "A hybrid adaptive antenna array for long-range mm-wave communications," *IEEE Ant. and Propag. Mag.*, vol. 54, no. 2, pp. 271–282, 2012.

[66] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Comm. Mag.*, vol. 52, no. 2, pp. 106–113, 2014.

[67] J. Nsenga, A. Bourdoux, and F. Horlin, "Mixed analog/digital beamforming for 60 GHz MIMO frequency selective channels," in *Proc. IEEE Intern. Conf. on Comm. (ICC)*, pp. 1–6, 2010.

[68] O. El Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE Intern.Conf. on Comm. (ICC),*, pp. 3724–3729, 2012.

[69] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inform. Theo. and App. Workshop (ITA)*, pp. 1–5, 2013.

[70] X. Luo and X. Wang, "How to calibrate massive MIMO?," in *Proc. IEEE Intern. Conf. on Commun. (ICC) Workshops*, 2015.

[71] G. V. Tsoulos and M. A. Beach, "Calibration and linearity issues for an adaptive antenna system," in *Proc. IEEE 47th Veh. Technol. Conf.*, vol. 3, pp. 1597–1600, May 1997.

[72] Y. Jin and F. Dai, "Impact of transceiver RFIC impairments on MIMO system performance," *IEEE Trans. Ind. Electron.*, vol. 59, pp. 538–549, Jan. 2012.

[73] I. Markovsky and S. Van H., "Overview of total least-squares methods," *Signal processing*, vol. 87, no. 10, pp. 2283–2302, 2007.

[74] "Ettus octo-clock." `https://www.ettus.com/product/details/OctoClock`. Accessed: 2017-06-18.

[75] D. Liu, W. Ma, S. Shao, Y. Shen, and Y. Tang, "Performance analysis of TDD reciprocity calibration for Massive MU-MIMO systems with ZF beamforming," *IEEE Commun. Letters*, vol. 20, no. 1, pp. 113–116, 2016.

[76] S. Blandino, F. Kaltenberger, and M. Feilen, "Wireless channel simulator testbed for airborne receivers," in *proc. IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2015.

[77] "OpenAirInterface: 5G software alliance for democratising wireless innovation." `http://www.openairinterface.org/`. Accessed: 2017-06-18.

[78] "Rice university wireless open access research platform." `http://warpproject.org/trac`. Accessed: 2017-06-21.

[79] E. Luther, "5g massive MIMO testbed: From theory to reality." `https://studylib.net/doc/18730180/5g-massive-mimo-testbed--from-theory-to-reality`, 2014.

[80] O. Edfors, "LuMaMi-A flexible testbed for massive MIMO."

[81] P. Harris, S. Zang, A. Nix, M. Beach, S. Armour, and A. Doufexi, "A distributed massive MIMO testbed to assess real-world performance and feasibility," in *Proc. IEEE 81st Veh. Tech. Conf. (VTC Spring)*, pp. 1–2, 2015.

[82] X. Wang, X. Hou, H. Jiang, A. Benjebbour, Y. Saito, Y. Kishiyama, J. Qiu, H. Shen, C. Tang, T. Tian, *et al.*, "Large scale experimental trial of 5G mobile communication systems—TDD massive MIMO with linear and non-linear precoding schemes," in *Proc. IEEE 27th Annual Intern. Symp. on Personal, Indoor, and Mobile Radio Comm. (PIMRC)*, pp. 1–5, 2016.

[83] "Samsung takes first 5G steps with advanced antenna." `https://www.pcworld.idg.com.au/article/461656/samsung_takes_first_5g_steps_advanced_antenna/`. Accessed: 2017-06-25.

[84] H. Suzuki, R. Kendall, K. Anderson, A. Grancea, D. Humphrey, J. Pathikulangara, K. Bengston, J. Matthews, and C. Russell, "Highly spectrally efficient ngara rural wireless broadband access demonstrator," in *Proc. IEEE Intern. Symp. on Commu. and Info. Tech. (ISCIT)*, pp. 914–919, 2012.

[85] "5G massive MIMO testbed." `https://www.nutaq.com/5g-massive-mimo-testbed`. Accessed: 2017-06-25.

[86] 3GPP, "Technical Specification 36.211: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 13)."

[87] 3GPP, "Technical Specification 36.211: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 14)."

[88] "Magma PCIe Chassis." `http://magma.com/products/pcie-expansion/expressbox-16-basic`. Accessed: 2017-06-18.

[89] "Single instruction multiple data." https://fr.wikipedia.org/wiki/Single_instruction_multiple_data. Accessed: 2017-06-20.

[90] "R&S FSQ signal analyzer." https://www.rohde-schwarz.com/us/manual/r-s-fsq-operating-manual-manuals-gb1_78701-29068.html. Accessed: 2017-08-10.

[91] "R&S FS-K96 OFDM vector signal analysis software." https://www.rohde-schwarz.com/us/product/fsk96-productstartpage_63493-11160.html. Accessed: 2017-08-10.

[92] "Advanced Dynamic Spectrum 5G mobile networks Employing Licensed shared access." http://www.fp7-adel.eu/. Accessed: 2017-06-18.