Channel Estimation in Large-Scale Multi-Antenna Systems for 5G and Beyond

Novel Pilot Structures and Algorithms

Karthik Upadhyya
Channel Estimation in Large-Scale Multi-Antenna Systems for 5G and Beyond

Novel Pilot Structures and Algorithms

Karthik Upadhyya

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall AS1, TUAS building of the school on 30 August 2018 at 12:00.

Aalto University
School of Electrical Engineering
Department of Signal Processing and Acoustics
Signal Processing Group
Abstract
Efficient use of the limited quantity of available spectrum to cater to the exponentially increasing demand for throughput has been the focus of communication and signal processing engineers for the past few decades. With the advent of technologies such as the Internet of things (IoT) or machine-type communications (MTC), devices and appliances around us which have predominantly been offline are being equipped with sensors that generate data and are now driving the demand for throughput. The forthcoming fifth generation (5G) standard is being developed to cater to these use cases and to also increase throughput for conventional mobile users. One of the enabling technologies of 5G is the use of antenna arrays with orders of magnitude more elements than in conventional fourth generation (4G) transceivers.
Large-scale multi-antenna systems impose constraints on channel training and transceiver architecture. In this thesis, we consider the problem of channel estimation in large-scale multi-antenna systems at conventional sub-6 GHz and millimeter-wave (mmWave) frequencies. In coherent receivers, channel state information (CSI) is obtained using training, which involves sending known pilots from the transmitter. In multi-cell networks, these pilots will have to be reused in different cells in order to limit the channel estimation overhead, resulting in a detrimental phenomenon known as pilot contamination. Pilot contamination, which causes interference and decreases throughput, is a fundamental challenge in large-scale multi-antenna systems. In the first part of this thesis, we address the issue of pilot contamination and propose using superimposed pilots for avoiding/mitigating interference. We also consider variants of superimposed pilots such as the hybrid system and staggered pilots to improve throughput.
Next, we address the problem of estimating spatial covariance matrices (SCMs) in massive MIMO systems in the presence of pilot contamination. SCMs are useful for mitigating the effects of pilot contamination, but have to be estimated from contaminated observations of the user channels, and consequently, are also contaminated. In the second part of this thesis, we propose a novel pilot structure for estimating contamination-free SCMs.
The shift to mmWave frequencies opens up large swathes of spectrum for communication, enabling the large throughputs that 5G demands. However, the channel propagation characteristics at these frequencies are markedly different from sub-6 GHz channels and communicating at mmWave frequencies imposes significant constraints on the transceiver architecture. Both factors in turn influence the design of signal processing algorithms. In the third part of the thesis, we address the problem of channel tracking in mmWave transceivers and develop novel semi-blind algorithms to track the channel with a low overhead.

Keywords
Massive MIMO, Millimeter Wave, Pilot Contamination, Superimposed Pilots, Staggered Pilots, Channel Estimation, Covariance Matrix Estimation, Channel Tracking

| ISSN (printed) | 1799-4934 | ISSN (pdf) | 1799-4942 |
| Location of publisher | Helsinki | Location of printing | Helsinki |
| Year | 2018 | Pages | 197 |
To Amma, Papa, Ammu, 
Kushi, Malli, and Inji.
Preface

The research presented in this thesis has been carried out between 2014-2018 at the Department of Signal Processing and Acoustics, Aalto University, Finland. This work would not have been possible without the patience, support, and encouragement of my supervisor, Prof. Sergiy A. Vorobyov. Prof. Vorobyov has been empathetic, compassionate, and helpful. He has also been very approachable, to the extent that I would drop by his office almost every day (or whenever I was stuck, which was more often than you’d imagine). He has made numerous opportunities available for me and I am grateful for all of them. These are qualities that rank very high on a student’s wish list and I consider myself fortunate to have him as my supervisor.

I would like to thank Prof. A. Lee Swindlehurst and Prof. Erik G. Larsson for having taken the time to be the opponents for my defense. I am also grateful to the preliminary examiners Prof. David J. Love and Prof. Osvaldo Simeone for the time and effort they have put in carefully reading through my dissertation and giving valuable comments that helped improve the manuscript.

I would like to express my gratitude to Prof. Robert W. Heath, Jr. for having hosted me at the University of Texas at Austin. Working with his fast-paced and enthusiastic research group was a rewarding experience. I am thankful to the Aalto Foundation and the Nokia Foundation for their scholarships that helped fund the research visit to UT Austin.

I wish to thank Prof. Visa Koivunen and Prof. Esa Ollila for their informative courses and critical comments in the group seminars, with special thanks to Prof. Visa Koivunen for having referred me to Nokia Bell Labs. I would also like to express my gratitude to Dr. Mikko Vehkaperä for his guidance and his extremely useful comments in the publications that we co-authored. Many thanks to my former supervisor, Prof. K.V.S. Hari, without whom, I would not have come to this lovely country.

Research does not happen without coffee and coffee is boring without friends. I am grateful to my current and former colleagues for making the workplace enjoyable. In particular, I would like to acknowledge Dr. Jari
Preface

Miettinen, Yongzhe Li, Matthew Morency, Marian Bică, Adriana Chiş, Dr. Hassan Naseri, Robin Rajamäki, Emadaldin Mozafari, Markus Yli-Niemi, Elias Raninen, Dr. Shahab Basiri, Atchut, Bhavya Omkarappa, Suhas Muniyappa, Kunal Ghosh, and Srikanth Gadicherla, and my colleagues at UT Austin - Nitin Jonathan Myers, Khurram Usman Mazher, Anum Ali, Preeti Kumari, Travis Cuvelier, and Srilakshmi Pattabiraman. Special thanks to Dr. Shailesh Singh Chauhan for his guidance and advice on a variety of issues at the beginning of my doctoral studies, Dr. Sachin Chaudhari for his assistance during my initial days in Finland, and Amit Kumar Jaiswal for putting up with me as a roommate.

I am grateful to my parents, sister, and the rest of my family members for their love and emotional support in the highs and lows of this four-year long roller-coaster ride. I owe my mother everything; she has made me what I am today. Thanks to my father for his support and my sister for always having been there for me when I needed her. I am also indebted to my late grandmother and my relatives for all that they have done for my family and me over the years.

Espoo, August 30, 2018,

Karthik Upadhya
Table of Contents

Preface 3

Contents 5

List of Publications 7

Author’s Contribution 9

List of Abbreviations 11

List of Symbols 15

1. Introduction 19
   1.1 Motivation ...................................... 19
   1.2 Research Objectives ............................ 20
   1.3 Contributions .................................. 21
   1.4 Author’s Independent Contribution ............. 22
   1.5 Thesis Structure ............................... 22

2. Massive MIMO 23
   2.1 Point-to-Point MIMO ........................... 23
   2.2 Multi-User MIMO ............................... 25
   2.3 Massive MU-MIMO ............................. 27
      2.3.1 Favorable Propagation Conditions ........ 27
      2.3.2 Single-cell Massive MIMO With Perfect CSI at BS 28
      2.3.3 Channel Estimation ......................... 29
   2.4 Challenges .................................... 31
      2.4.1 Pilot Contamination ......................... 31
      2.4.2 Hardware Constraints ....................... 33

3. Channel Estimation in TDD Massive MIMO 35
   3.1 Overview ...................................... 35
### Contents

3.2 State-of-the-Art in Channel Estimation for TDD Massive MIMO .................................. 36

3.2.1 Pilot Decontamination at the BS .......... 36

3.2.2 Pilot Contamination Avoidance ............ 41

3.3 Contribution .................................. 44

3.4 Superimposed Pilots ........................... 44

3.4.1 Hybrid System ............................. 47

3.4.2 Staggered Pilots ............................ 49

3.4.3 Iterative Data-Aided Channel Estimation ... 51

4. Covariance Matrix Estimation for Massive MIMO 55

4.1 Background and Motivation ...................... 55

4.2 Challenges in Estimating the Spatial Covariance Matrices 57

4.2.1 Estimating Spatial Covariance Matrices When $N < M$ ........................ 58

4.2.2 Covariance Matrix Estimation in Massive MIMO in the Presence of Pilot Contamination . .... 59

4.3 Contributions .................................. 61

4.4 Estimation using Randomly Phase-Shifted Pilots in Publication VI ............................. 62

5. Channel Estimation and Tracking in Millimeter-Wave MIMO 65

5.1 Motivation ..................................... 65

5.2 mmWave Channel Characteristics ............... 66

5.2.1 Diffraction .................................. 66

5.2.2 Scattering and Penetration Losses .......... 66

5.2.3 Distance-Dependent Path-loss .............. 67

5.2.4 Channel Representation ..................... 68

5.3 mmWave Architectures ........................ 69

5.4 Channel Estimation in mmWave MIMO ......... 72

5.5 Hybrid Beamforming ............................ 77

5.6 Channel Tracking in mmWave MIMO ............ 78

5.7 Contribution .................................. 80

5.7.1 Method Proposed in Publication VII ........ 80

5.7.2 Method Proposed in Publication VIII ....... 81

6. Conclusion and Future Work 83

Bibliography 85

Errata 95

Publications 97
List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.


VII K. Upadhya, R. W. Heath Jr, S. A. Vorobyov. Tracking abruptly chang-

Author’s Contribution

Publication I: “Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO”

The author proposed the idea, performed the analysis and simulations with input from the co-authors.

Publication II: “Downlink performance of superimposed pilots in massive MIMO systems in the presence of pilot contamination”

The author proposed the idea, performed the analysis and simulations with input from the co-authors.

Publication III: “Time-multiplexed / superimposed pilot selection for massive MIMO pilot decontamination”

The author proposed the idea, performed the analysis and simulations with input from the co-authors.

Publication IV: “Superimposed pilots are superior for mitigating pilot contamination in massive MIMO”

The author proposed the idea, performed the analysis and simulations with input from the co-authors.
Author’s Contribution

**Publication V: “Downlink performance of superimposed pilots in massive MIMO systems”**

The author proposed the idea, performed the analysis and simulations with input from the co-authors.

**Publication VI: “Covariance matrix estimation for massive MIMO”**

The author proposed the idea, performed the analysis and simulations with input from the co-author.

**Publication VII: “Tracking abruptly changing channels in mmWave systems using overlaid data and training”**

The author proposed the idea, performed the analysis and simulations with input from the co-authors.

**Publication VIII: “Low-overhead receiver-side channel tracking for mmWave MIMO”**

The author proposed the idea, performed the analysis and simulations with input from the co-authors.
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4G</td>
<td>Fourth generation</td>
</tr>
<tr>
<td>5G</td>
<td>Fifth generation</td>
</tr>
<tr>
<td>ADC</td>
<td>Analog-to-digital converter</td>
</tr>
<tr>
<td>AGC</td>
<td>Automatic gain controller</td>
</tr>
<tr>
<td>AM</td>
<td>Alternating maximization</td>
</tr>
<tr>
<td>AoA</td>
<td>Angle of arrival</td>
</tr>
<tr>
<td>AoD</td>
<td>Angle of departure</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive white Gaussian noise</td>
</tr>
<tr>
<td>BER</td>
<td>Bit error rate</td>
</tr>
<tr>
<td>BS</td>
<td>Base station</td>
</tr>
<tr>
<td>CRLB</td>
<td>Cramér-Rao lower-bound</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel-state information</td>
</tr>
<tr>
<td>CSIT</td>
<td>Channel state information at transmitter</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier transform</td>
</tr>
<tr>
<td>DL</td>
<td>Downlink</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-maximization</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency-division duplexing</td>
</tr>
<tr>
<td>FEC</td>
<td>Forward error correction</td>
</tr>
<tr>
<td>i.i.d</td>
<td>Independent and identically distributed</td>
</tr>
</tbody>
</table>
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IoT</td>
<td>Internet of things</td>
</tr>
<tr>
<td>LDPC</td>
<td>Low-density parity check</td>
</tr>
<tr>
<td>LMMSE</td>
<td>Linear minimum mean-squared error</td>
</tr>
<tr>
<td>LOS</td>
<td>Line-of-sight</td>
</tr>
<tr>
<td>LP</td>
<td>Linear programming</td>
</tr>
<tr>
<td>LS</td>
<td>Least-squares</td>
</tr>
<tr>
<td>LTE</td>
<td>Long-term evolution</td>
</tr>
<tr>
<td>M-ZF</td>
<td>Multi-cell zero-forcing</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum a posteriori</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-input multiple-output</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum-likelihood</td>
</tr>
<tr>
<td>MM</td>
<td>Majorization-minimization</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum mean-squared error</td>
</tr>
<tr>
<td>mMTC</td>
<td>Massive MTC</td>
</tr>
<tr>
<td>MMV</td>
<td>Multiple measurement-vector</td>
</tr>
<tr>
<td>mmWave</td>
<td>Millimeter-wave</td>
</tr>
<tr>
<td>MR</td>
<td>Maximum-ratio</td>
</tr>
<tr>
<td>MRC</td>
<td>Maximum-ratio combiner</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean-squared error</td>
</tr>
<tr>
<td>MTC</td>
<td>Machine-type communication</td>
</tr>
<tr>
<td>MU</td>
<td>Multi-user</td>
</tr>
<tr>
<td>MUSIC</td>
<td>Multiple signal classification</td>
</tr>
<tr>
<td>NLOS</td>
<td>Non-line-of-sight</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal frequency-division multiplexing</td>
</tr>
<tr>
<td>OMP</td>
<td>Orthogonal matching pursuit</td>
</tr>
<tr>
<td>PARAFAC</td>
<td>Parallel factor analysis</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature amplitude modulation</td>
</tr>
<tr>
<td>RF</td>
<td>Radio frequency</td>
</tr>
<tr>
<td>RMS</td>
<td>Root mean-square</td>
</tr>
<tr>
<td>RP</td>
<td>Regular pilot</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>RZF</td>
<td>Regularized zero-forcing</td>
</tr>
<tr>
<td>SBL</td>
<td>Sparse Bayesian learning</td>
</tr>
<tr>
<td>SCM</td>
<td>Spatial covariance matrix</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal-to-interference-plus-noise ratio</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SP</td>
<td>Superimposed pilot</td>
</tr>
<tr>
<td>TDD</td>
<td>Time-division duplexing</td>
</tr>
<tr>
<td>UL</td>
<td>Uplink</td>
</tr>
<tr>
<td>ULA</td>
<td>Uniform linear array</td>
</tr>
<tr>
<td>V2V</td>
<td>Vehicle-to-Vehicle</td>
</tr>
<tr>
<td>V2X</td>
<td>Vehicle-to-Everything</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless local area network</td>
</tr>
<tr>
<td>ZF</td>
<td>Zero-forcing</td>
</tr>
</tbody>
</table>
### List of Symbols

#### Latin Letters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{A}_R$</td>
<td>Matrix with steering vectors corresponding to AoAs as columns</td>
<td></td>
</tr>
<tr>
<td>$\mathbf{A}_T$</td>
<td>Matrix with steering vectors corresponding to AoDs as columns</td>
<td></td>
</tr>
<tr>
<td>$A_R$</td>
<td>Effective aperture of receive antenna</td>
<td></td>
</tr>
<tr>
<td>$A_T$</td>
<td>Effective aperture of transmit antenna</td>
<td></td>
</tr>
<tr>
<td>$B_c$</td>
<td>Coherence bandwidth.</td>
<td></td>
</tr>
<tr>
<td>$C_d$</td>
<td>Number of symbols in the DL time slot.</td>
<td></td>
</tr>
<tr>
<td>$C_{dl}$</td>
<td>DL channel capacity.</td>
<td></td>
</tr>
<tr>
<td>$C_{m \times n}$</td>
<td>Set of $m \times n$ complex valued matrices.</td>
<td></td>
</tr>
<tr>
<td>$C_u$</td>
<td>Number of symbols in the UL time slot.</td>
<td></td>
</tr>
<tr>
<td>$C_{ul}$</td>
<td>UL channel capacity.</td>
<td></td>
</tr>
<tr>
<td>$d$</td>
<td>Distance between transmitter and receiver.</td>
<td></td>
</tr>
<tr>
<td>$\mathbf{d}$</td>
<td>Vector of symbols transmitted in the DL.</td>
<td>diag operator.</td>
</tr>
<tr>
<td>$\operatorname{diag}{\cdot}$</td>
<td>Expectation operator.</td>
<td></td>
</tr>
<tr>
<td>$\mathbf{F}$</td>
<td>Precoding matrix</td>
<td></td>
</tr>
<tr>
<td>$\mathbf{F}_{\text{BB}}$</td>
<td>Baseband Precoding matrix</td>
<td></td>
</tr>
<tr>
<td>$\mathbf{F}_{\text{RF}}$</td>
<td>RF Precoding matrix</td>
<td></td>
</tr>
<tr>
<td>$g_{j\ell k}$</td>
<td>Component of the channel vector corresponding to small-scale fading between user $(\ell, k)$ and BS $j$.</td>
<td></td>
</tr>
<tr>
<td>$G_R$</td>
<td>Gain of receive antenna</td>
<td></td>
</tr>
<tr>
<td>$G_T$</td>
<td>Gain of transmit antenna</td>
<td></td>
</tr>
<tr>
<td>$H$</td>
<td>Channel matrix between transmitter and receiver.</td>
<td></td>
</tr>
<tr>
<td>$h_{j\ell k}$</td>
<td>Channel vector between user $(\ell, k)$ and BS $j$.</td>
<td></td>
</tr>
<tr>
<td>$I(\mathcal{U}<em>{\text{RP}}, \mathcal{U}</em>{\text{SP}})$</td>
<td>Total weighted interference in the hybrid system with user partitions $\mathcal{U}<em>{\text{RP}}$ and $\mathcal{U}</em>{\text{SP}}$.</td>
<td></td>
</tr>
<tr>
<td>$I_{\text{RP} - \text{dl}}$</td>
<td>Interference power caused by user $(\ell, k)$ in the DL if assigned to $\mathcal{U}_{\text{RP}}$.</td>
<td></td>
</tr>
<tr>
<td>$I_{\text{RP} - \text{ul}}$</td>
<td>Interference power caused by user $(\ell, k)$ in the UL if as-</td>
<td></td>
</tr>
</tbody>
</table>
List of Symbols

\( I_{\text{SP}-\text{dl}} \)  
Interference power caused by user \((\ell, k)\) in the DL if assigned to \(U_{\text{SP}}\).

\( I_{\text{SP}-\text{ul}} \)  
Interference power caused by user \((\ell, k)\) in the UL if assigned to \(U_{\text{SP}}\).

\( K \)  
Number of users per cell.

\( L \)  
Number of cells in the network.

\( M \)  
Number of BS antennas.

\( N_s \)  
Number of data streams.

\( n_{\text{dl}} \)  
AWGN vector in the DL.

\( n_{\text{ul}} \)  
AWGN vector in the UL.

\( P \)  
Number of channel paths.

\( P_{\ell k} \)  
Superimposed pilot transmitted by user \(k\) of cell \(\ell\).

\( P_{\text{dl}} \)  
Transmit power in the DL.

\( P_T \)  
Transmit power.

\( P_{\text{ul}} \)  
Transmit power in the UL.

\( Q_{jm} \)  
Spatial covariance matrix of channel estimate of user \((j, m)\).

\( r \)  
Pilot reuse factor.

\( \mathbb{R}_+ \)  
Set of positive real-valued scalars.

\( R_{\ell j k} \)  
Spatial covariance matrix of user \((\ell, k)\) at BS \(j\).

\( r^{\text{RP}} \)  
Pilot reuse factor for users transmitting RP.

\( r^{\text{SP}} \)  
Pilot reuse factor for users transmitting SP.

\( s \)  
Vector of symbols transmitted in the UL.

\( s_{\ell k} \)  
Vector of symbols transmitted in the UL by user \(k\) of cell \(\ell\).

\( T_c \)  
Coherence time.

\( \mathcal{U} \)  
Set of users the network.

\( \mathcal{U}_{\text{RP}} \)  
Set of users in the hybrid system transmitting RP.

\( \mathcal{U}_{\text{SP}} \)  
Set of users in the hybrid system transmitting SP.

\( W \)  
Combining matrix

\( W_{\text{BB}} \)  
Baseband Combining matrix

\( W_{\text{RF}} \)  
RF Combining matrix

\( x_{\ell k} \)  
Vector containing UL data transmitted by user \(k\) of cell \(\ell\).

\( y_{\text{dl}} \)  
Received vector in downlink.

\( y_{\text{ul}} \)  
Received vector in uplink.

**Greek Letters**

\( \alpha \)  
Complex path-gain

\( \beta_{j \ell k} \)  
Large-scale path-loss coefficient between user \((\ell, k)\) and BS \(j\).

\( \gamma \)  
Shrinkage coefficient.

\( \delta \)  
Antenna spacing

\( \delta_{\text{Kr}}(\cdot) \)  
Kronecker-delta function

\( \theta \)  
Angle of arrival.

\( \lambda \)  
Wavelength of the radio signal.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\mu^2$ is the fraction of UL power allocated to pilot.</td>
</tr>
<tr>
<td>$\xi_{dl}$</td>
<td>Weight for the interference power in the DL.</td>
</tr>
<tr>
<td>$\xi_{ul}$</td>
<td>Weight for the interference power in the UL.</td>
</tr>
<tr>
<td>$\Pi [n]$</td>
<td>Pilot allocation matrix in coherence block $n$.</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$\rho^2$ is the fraction of UL power allocated to data.</td>
</tr>
<tr>
<td>$\sigma_k$</td>
<td>$k$th singular value of the channel.</td>
</tr>
<tr>
<td>$\tau$</td>
<td>RP pilot length.</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>Scaled unitary matrix with columns used as RP.</td>
</tr>
<tr>
<td>$\phi_{lk}$</td>
<td>RP transmitted in the UL by user $(\ell, k)$.</td>
</tr>
<tr>
<td>$\psi$</td>
<td>Angle of departure.</td>
</tr>
<tr>
<td>$\Omega_p [n]$</td>
<td>Set of users assigned pilot sequence $p$ in coherence block $n$.</td>
</tr>
</tbody>
</table>

**Notation**

- $|\cdot|$ : Cardinality of a set.
- $1_{\{S\}}$ : Indicator function on set $S$.
- $(\cdot)^*$ : Complex conjugate of a matrix.
- $(\cdot)^H$ : Conjugate-transpose of a matrix.
- $(\cdot)^T$ : Matrix transpose.
- $(\cdot)^\dagger$ : Moore-Penrose pseudo-inverse.
- $(\ell, k)$ : Tuple that denotes user $k$ in cell $\ell$.
- $\mathcal{CN}(\mu, R)$ : Complex Gaussian distribution with mean $\mu$ and covariance matrix $R$.
- $\emptyset$ : Null set.
1. Introduction

1.1 Motivation

With limited quantity of available spectrum, improving the spectral efficiency to cater to the exponentially increasing demand for data rates has engaged communication and signal processing engineers for the past few decades. With the advent of technologies such as the internet of things (IoT) or machine-type communication (MTC), devices and appliances around us which have predominantly been offline are now being equipped with sensors that generate data. Making use of this data, in many cases, requires it to be sent over a communication link to a central location, with the nature of the sensed data specifying the communication requirements. For instance, support for high-bandwidth high-reliability communication with low latency is required in vehicle-to-vehicle (V2V) networks which have been envisaged to enable future autonomous vehicles to share sensor data [1, 2]. On the other extreme, we have massive MTC (mMTC) where the focus is on a massive number of battery-operated devices transmitting intermittently at low-data rates (of the order of 10 Kb/s) with uplink (UL) dominant traffic [3]. Therefore, a strong and versatile communication backbone is required to address these diverse needs. Moreover, with the advent of these devices, the demand for data rates is now machine-driven rather than human-driven and therefore, the design of future communication networks should take this into account.

The next-generation fifth generation (5G) communication standard is being developed to address these diverse use-cases. One of the enabling technologies of 5G is the use of antenna arrays with orders of magnitude more elements than in conventional fourth generation (4G) long-term evolution (LTE) transceivers (which have upto 8 antenna elements). The large-scale antenna arrays offer significantly higher array and diversity gains which result in higher spectral and UL energy efficiencies at low computational complexities. Consequently, transceivers equipped with
these large antenna arrays are expected to be a standard feature of 5G [4] and beyond-5G cellular networks. Another important shift in 5G is the push to use the relatively unused spectrum in millimeter-wave (mmWave) frequencies. mmWave, with carrier frequencies in the range $30 - 300$ GHz, offers large contiguous blocks of bandwidth of upwards of $1$ GHz. For instance, the IEEE 802.11ad wireless local area network (WLAN) standard uses $1.88$ GHz bandwidth at $60$ GHz carrier frequency [5] while, in contrast, 4G LTE at sub-$6$ GHz frequency supports a bandwidth of only $20$ MHz.

The gain in spectral efficiency promised by transceivers with large antenna arrays is contingent on the availability of accurate channel-state information (CSI) for beamforming. In practice, the CSI has to estimated at the receiver from observations made when known pilot sequences are sent from the transmitter. Estimating the channel in a multi-cell multi-user environment with large antenna arrays and with a low overhead mandates using time-division duplexing (TDD) and UL pilot reuse [6, 7, 8]. Pilot reuse results in coherent interference in the UL and downlink (DL) and is termed as *pilot contamination* [9, 6]. Pilot contamination also reduces the coherent beamforming gain offered by the large antenna array. The coherent interference and decreased beamforming gains negatively impact the UL and DL throughputs [10], thereby diminishing the benefits of large-scale antenna arrays.

Large-scale antenna arrays also impose unique constraints on transceiver architecture, especially at mmWave frequencies. The mmWave channel also exhibits very different propagation characteristics in comparison with the sub-$6$ GHz channel. These two critical differences mandate novel signal processing algorithms for mmWave communication links.

### 1.2 Research Objectives

In this thesis, we consider sub-$6$ GHz massive multi-user (MU)-multiple-input multiple-output (MIMO) and mmWave MIMO transceivers. The first objective of this thesis is to develop methods for avoiding pilot contamination by suitably modifying the nature of pilot transmission at the user terminal.

The spatial covariance matrices (SCMs) of individual users at the base station (BS) are useful for decontaminating the channel estimates [11, 12, 13]. These SCMs have to be estimated at the BS from observations of the individual user channels. However, due to pilot contamination, observations of individual user channels are contaminated with the channel vectors of users in neighboring cells which share the same pilot, and using these contaminated observations directly will result in the estimates of the SCMs also becoming contaminated. Therefore, the problem of estimating the individual SCMs in the presence of pilot contamination is challenging.
The second objective of this thesis is to develop pilot structures and algorithms to obtain contamination-free estimates of the individual user SCMs at the BS in the presence of pilot contamination.

mmWave transceivers use large antenna arrays and generate narrow beams to compensate for the increased path-loss at mmWave frequencies. These narrow beams render the mmWave communication link sensitive to user mobility since the communication link is susceptible to changes in the angles of departure and arrival of the channel paths at the transmitter and receiver. Furthermore, mmWave communication links are also sensitive to blockages because of the higher penetration losses at mmWave frequencies. Consequently, the time-varying mmWave channel has to be tracked at both ends of the communication link in order to maintain sufficient signal-to-noise ratio (SNR). Developing algorithms for mmWave channel tracking is further complicated by the hardware constraints imposed on the transceiver architecture. Therefore, the third objective of this thesis is to develop algorithms to track changes in the mmWave channel with a low overhead while satisfying the hardware constraints on the transceiver architecture.

1.3 Contributions

The main contributions of this thesis are described as follows

1. We have proposed using superimposed pilots (SPs) to avoid pilot contamination for estimating the channel in massive MU-MIMO. We have also proposed variations of SP such as staggered pilots and the hybrid system to obtain higher throughputs by reducing the inter and intra-cell interference in SP. This contribution is summarized in Chapter 3.

2. We have proposed a novel pilot structure and algorithm for estimating asymptotically contamination-free individual user SCMs in the presence of pilot contamination. The proposed method has the benefit of not requiring UL synchronization between users in the different cells. This contribution is summarized in Chapter 4.

3. We have proposed two semi-blind methods for channel tracking in point-to-point mmWave MIMO transceivers. In the first method, the transmitter transmits pilots in the null-space of the channel matrix and data in its signal space. The receiver then estimates the data and the time-variations in the channel jointly. In the second method, we consider the scenario in which the angles of arrival (AoAs) of the paths change at the receiver while the corresponding angles of departure (AoDs) at the transmitter are fixed. We propose an algorithm to track the receiver-side
channel for this scenario with a low overhead. Both these methods satisfy the hardware constraints imposed by mmWave MIMO. This contribution is summarized in Chapter 5.

1.4 Author’s Independent Contribution

The main results of this thesis have been published in two journal articles and five conference papers. An additional journal article is currently in the last stage of review (after revision). The author of this thesis is responsible for the theoretical studies, algorithm development, and numerical results in all the publications included in this dissertation (Publications I-VIII). The co-authors helped with planning the research and in writing and revising the publications.

1.5 Thesis Structure

This thesis is divided into an introduction part which summarizes the contributions in Publications P.I - P.VIII, and a collection of the eight original publications P.I - P.VIII. The attached publications include the original theory, methods, and results that are presented in this dissertation. The introductory part of this thesis is structured as follows. In Chapter 2, we briefly recap the concepts of point-to-point MIMO and MU-MIMO, introduce massive MIMO, and discuss its potential benefits. We then describe some of the challenges in massive MIMO that are addressed in this thesis. Methods for mitigating the effect of pilot contamination are discussed in Chapter 3. SPs for massive MIMO along with its variants are introduced in this chapter. In Chapter 4, methods for estimating the individual user SCMs in massive MIMO are discussed. We also introduce a novel pilot structure for estimating the SCM in the presence of pilot contamination. Chapter 5 contains an overview of mmWave architectures and channel estimation and precoding/combining algorithms. Methods for channel tracking are also discussed in the chapter.
2. Massive MIMO

MIMO technology has been a topic of interest for the past two decades and MU-MIMO has made its way into standards such as 4G LTE and IEEE 802.11 (WiFi). Massive MIMO is a variant of MU-MIMO with the potential to offer significantly higher spectral and energy efficiencies at low computational complexities, making it one of the enabling technologies for 5G communication systems [4, 14, 15, 16].

Before we look at massive MIMO in further detail, we will briefly review point-to-point and MU-MIMO technologies, discuss its limitations, and describe how massive MIMO overcomes these limitations and what it has to offer.

2.1 Point-to-Point MIMO

Point-to-point MIMO is an elementary version of a MIMO system in which a BS with \( M \) antennas communicates with a user terminal with \( K \) antennas, as shown in Fig. 2.1. Let \( H \in \mathbb{C}^{M \times K} \) be the channel in the UL. Then, under the assumption of channel reciprocity, the received observations at the BS and the user terminal in the UL and DL, respectively, are

\[
y_{ul} = \sqrt{\frac{P_{ul}}{K}} H s + n_{ul} \quad (2.1)
\]

\[
y_{dl} = \sqrt{\frac{P_{dl}}{M}} H^T d + n_{dl} \quad (2.2)
\]

where \( P_{ul} \) and \( P_{dl} \) are the powers with which symbols \( s \in \mathbb{C}^K \) and \( d \in \mathbb{C}^M \) are transmitted in the UL and DL, respectively, and \((\cdot)^T\) denotes the matrix transpose. The vectors \( s \) and \( d \) are assumed to be such that \( \mathbb{E}\{\|s\|^2\} = K \) and \( \mathbb{E}\{\|d\|^2\} = M \), where \( \mathbb{E}\{\cdot\} \) is the expectation operator and \( \|\cdot\| \) is the \( \ell^2 \) norm of a vector. The additive noise vectors \( n_{ul} \) and \( n_{dl} \) are assumed to be zero-mean and independent and identically distributed (i.i.d) complex Gaussian random variables with unit variance. If the CSI is available only at the receiver, which is the BS in the UL and the user terminal in the DL,
the channel capacity in the UL and DL can then be obtained as [17]

\[
C_{\text{ul}} = \log_2 \left| I_K + \frac{P_{\text{ul}}}{K} H H^H \right| = \min\{M,K\} \sum_{k=1} \log_2 \left( 1 + \frac{P_{\text{ul}}}{K} \sigma_k^2 \right) \tag{2.3}
\]

\[
C_{\text{dl}} = \log_2 \left| I_M + \frac{P_{\text{dl}}}{M} H H^H \right| = \min\{M,K\} \sum_{k=1} \log_2 \left( 1 + \frac{P_{\text{dl}}}{M} \sigma_k^2 \right) \tag{2.4}
\]

where \(\sigma_k\) is the \(k\)th singular value of \(H\) and \((\cdot)^H\) denotes conjugate transpose.

The channel capacities \(C_{\text{ul}}\) and \(C_{\text{dl}}\) depend on the values of \(M\) and \(K\) and the richness of the scattering environment. With rich scattering, the channel coefficient at each antenna is uncorrelated, and consequently, the communication link benefits from both an array and a spatial multiplexing gain. The former increases with both \(M\) and \(K\) and improves the link SNR, while the latter allows for a maximum of \(N_s \triangleq \min\{M,K\}\) streams to be transmitted in parallel. Increasing \(N_s\) when the channel can support
these streams (such as with rich scattering) results in a capacity increase through spatial multiplexing. In the other extreme, with limited scattering and a rank-one channel, such as in line-of-sight (LOS) conditions, spatial multiplexing is not possible and the communication link only benefits from the array gain. Since the SNR, which is a function of the array gain, is inside the logarithm, the amount of increase in the throughput diminishes with increasing \( M \).

In channel conditions that support multiple streams, exploiting this feature requires increasing \( M \) and \( K \), with the latter increasing the cost and complexity of the user terminal since each antenna needs a dedicated radio-frequency (RF) chain. The number of antennas \( K \) is also limited by the size of the device since the antennas have to be separated by half the wavelength in order to prevent undesirable grating lobes in the generated beams while ensuring that the channel coefficients between different antennas are uncorrelated.

MU-MIMO overcomes some of the aforementioned limitations of point-to-point MIMO, and is discussed in detail in Section 2.2.

### 2.2 Multi-User MIMO

In MU-MIMO, the \( K \) antennas in a point-to-point MIMO user terminal is separated across \( K \) single-antenna terminals\(^1\) (as shown in Fig. 2.2). The BS then communicates with the \( K \) terminals over the same time-frequency resource. Denoting \( h_k \in \mathbb{C}^M \) as the channel between user \( k \) and the BS, the received symbols in the UL can be written as

\[
y_{ul} = \sqrt{P_{ul}} \sum_{k=1}^{K} h_k s_k + n_{ul} = \sqrt{P_{ul}} H s + n_{ul}
\]

where \( H \triangleq [h_1, \ldots, h_K] \) and \( s \triangleq [s_1, \ldots, s_K]^T \). The UL symbols \( s_k \) are normalized such that \( \mathbb{E} \{ |s_k|^2 \} = 1 \). The received signal by user \( k \) in the DL can be written as

\[
y_{dl,k} = \sqrt{P_{dl}} h_k^T d + n_{dl,k} .
\]

Stacking the elements \( \{ y_{dl,k} \}_{k=1}^{K} \) in a column vector, we get

\[
y_{dl} = \sqrt{P_{dl}} H^T d + n_{dl}
\]

where \( n_{dl} \triangleq [n_{dl,1}, \ldots, n_{dl,K}]^T \). Assuming that the CSI is available at the BS in the UL and both the BS and the user terminals in the DL, the sum

\(^1\)The user terminals could also have multiple antennas which would then allow multiple simultaneous data streams between a multi-antenna user terminal and the BS under favorable propagation conditions. However, to describe the salient features of MU-MIMO over point-to-point MIMO, single-antenna terminals are sufficient, which is what we will consider in this section.
capacity in the UL and DL is then given as [18]

\[
C_{ul} = \log_2 \left| I_K + P_{ul} H^H H \right| \quad (2.8)
\]

\[
C_{dl} = \max_{p \geq 0} \log_2 \left| I_M + P_{dl} H PH^H \right| \quad (2.9)
\]

where \( P = \text{diag} \{ p \} \) is a diagonal matrix with the elements of \( p \) on its diagonals. Non-linear methods such as successive interference cancellation (SIC) and dirty paper coding approach these theoretical limits [19, 20].

In comparison with point-to-point MIMO, MU-MIMO is less sensitive to channel propagation conditions. This is because the channel vectors between the user terminals and the BS are more likely to be independent than in point-to-point MIMO owing to the greater separation between user terminals. Moreover, propagation conditions such as LOS, which are
severely detrimental to point-to-point MIMO, are less stressing for MU-MIMO. From the point of view of hardware complexity, MU-MIMO requires only single antenna terminals thereby addressing an important hardware limitation of point-to-point MIMO terminals. However, achieving the UL and DL spectral efficiencies in (2.8) and (2.9) requires complicated nonlinear signal processing at the BS and the user terminals. This limitation is addressed by massive MIMO which is explained in detail in the subsequent section.

2.3 Massive MU-MIMO

Massive MIMO, introduced in the seminal papers [6, 9], is a variant of MU-MIMO with $M$ being much larger than $K$. In [6], it was shown that under favorable propagation conditions, the achievable throughput increases logarithmically in $M$, and that linear precoding and combining at the BS are asymptotically optimal. In addition, under these propagation conditions, the number of users supported, and consequently, the spatial-multiplexing gain also increases with $M$. All these benefits are under the assumption that perfect CSI is available at the BS. In Section 2.3.1, we describe in detail the nature of the propagation conditions that are favorable for large antenna systems. CSI acquisition and the effect of imperfect CSI at the BS are described in Sections 2.3.3 and 2.4.1.

2.3.1 Favorable Propagation Conditions

Let $h_k \in \mathbb{C}^M$ be the channel vector of an arbitrary user $k$ at the reference BS. We say that the channel exhibits favorable propagation when the channel vectors of any two users are asymptotically orthogonal, i.e.,

$$\frac{1}{M} h_k^H h_{k'} \to 0, \quad \forall \ k \neq k'. \tag{2.10}$$

As an example, consider the rich-scattering environment in which the channel coefficients at each antenna are i.i.d complex Gaussian random variables, i.e., $h_k \sim \mathcal{CN}(0, \beta_k I_M)$, where $\beta_k \in \mathbb{R}_+$ is the large-scale path-loss coefficient. Assuming that the channel vectors of any pair of users are independent, we have from the law of large numbers that

$$\frac{1}{M} h_k^H h_{k'} \xrightarrow{a.s.} \frac{1}{M} \mathbb{E}\{h_k^H h_{k'}\} = 0, \quad \forall \ k \neq k'. \tag{2.11}$$

$$\frac{1}{M} \|h_k\|^2 \xrightarrow{a.s.} \frac{1}{M} \mathbb{E}\{\|h_k\|^2\} = \beta_k, \quad \forall \ k \tag{2.12}$$

where $\xrightarrow{a.s.}$ signifies almost-sure convergence. From (2.11), we see that the Rayleigh-fading channel satisfies asymptotic orthogonality. The large array at the BS with independent channel coefficients provides a diversity
order of $M$, and as $M \to \infty$, the increasing diversity order renders the communication link insensitive to small-scale fading. This phenomenon is called channel hardening \cite{6, 10} and results in the inner product of the channel vector $h$ and a beamforming vector $v$, which is obtained from $h$, becoming equal to its average value asymptotically, i.e., $v^H h \xrightarrow{a.s.} \mathbb{E}\{v^H h\}$.

In addition, consider the narrow-band LOS channel with AoA $\theta$ at the BS. If the BS is equipped with a uniform linear array (ULA) with elements spaced half the wavelength apart from each other, the steering vector corresponding to AoA $\theta$ can be defined as $a(\theta) \triangleq [1, e^{j\pi \sin(\theta)}, e^{j2\pi \sin(\theta)}, \ldots, e^{j(M-1)\pi \sin(\theta)}]^T$. Then the channel vector of user $k$ with AoA $\theta_k$ can be written as \cite{19}

$$h_k = \sqrt{\beta_k} e^{-j\frac{2\pi d_k}{\lambda}} a(\theta_k)$$ (2.13)

where $d_k$ is the distance between user $k$ and the reference BS and $\lambda$ is the wavelength of the carrier. For large antenna arrays, we have the property \cite{11}

$$\frac{1}{M} a^H(\theta_k) a(\theta_{k'}) \to \delta_{K\ell}(\theta_k, \theta_{k'})$$ (2.14)

where $\delta_{K\ell}(-)$ is the Kronecker delta function. From (2.14), it is straightforward to show that the LOS channel in (2.13) satisfies the asymptotic orthogonality condition in (2.10) provided each user has a distinct AoA at the BS. The asymptotic orthogonality is a consequence of the large array gain at the BS allowing the BS to form narrow beams towards each user with the beam-width progressively decreasing with increasing $M$. Note the difference between the LOS and the independent Rayleigh fading scenarios where the former channel is deterministic and the latter is random.

### 2.3.2 Single-cell Massive MIMO With Perfect CSI at BS

Consider a single-cell MIMO system with $K$ single-antenna users and $M$ antennas at the BS. We assume that the channel vector $h_k$ can be written as

$$h_k = \sqrt{\beta_k} g_k$$ (2.15)

where $g_k \in \mathbb{C}^M$ accounts for the small-scale fading or the array response. Setting $D \triangleq \text{diag}\{\beta_1, \ldots, \beta_K\}$ and $G \triangleq [g_1, \ldots, g_K]$, the channel matrix $H$ can be written as $H = GD^{1/2}$. As a consequence of asymptotic orthogonality, we have \cite{7}

$$H^H H = D^{1/2} G^H G D^{1/2} \approx MD$$ (2.16)
Substituting (2.16) into (2.8) and (2.9), the asymptotic capacity can be obtained as [7]

\[
C_{ul} = \log_2 |I_K + P_{ul} H^H H| \\
\approx \log_2 |I_K + P_{ul} MD| \\
= \sum_{k=1}^{K} \log_2 (1 + P_{ul} M \beta_k) \\
(2.17)
\]

\[
C_{dl} = \max_{p^T 1 = 1, p \geq 0} \log_2 |I_M + P_{dl} HPH^H| \\
\approx \max_{p^T 1 = 1, p \geq 0} \log_2 |I_K + P_{dl} MD| . \\
(2.18)
\]

With a maximum-ratio (MR) combiner used for detection in (2.5), as a consequence of asymptotic orthogonality, we obtain the following at the output of the combiner [7, 16]

\[
\hat{s} = H^H y_{ul} = \sqrt{P_{ul}} H^H H s + H^H n_{ul} \approx \sqrt{P_{ul}} M D s + H^H n_{ul} . \\
(2.19)
\]

Similarly, with an MR precoder, the transmitted signal is \( d = M^{-1/2} H^* D^{-1/2} P^{1/2} x \), where \( x \in \mathbb{C}^K \) is the source information vector such that \( E \{ xx^H \} = I \). Normalizing \( d \) by \( \sqrt{M} \) ensures that the average transmit power is \( E \{ ||d||^2 \} = 1 \). Then, the received vector of symbols at the \( K \) users in the DL is given as

\[
y_{dl} = \sqrt{P_{dl}} M H^T H^* D^{-1/2} P^{1/2} x + n_{dl} \approx \sqrt{P_{dl}} M D^{1/2} P^{1/2} x + n_{dl} . \\
(2.20)
\]

From (2.19) and (2.20), it can be seen that for large \( M \), the spectral efficiency with MR precoder and combiner is the same as that of (2.17) and (2.18), implying that the simple linear MR precoder and combiner with computational complexity proportional to \( M \) are optimal with perfect CSI at the BS [7]. This is a shift in paradigm with respect to MU-MIMO in which computationally complex non-linear processing is required to achieve capacity.

Another consequence of (2.17) is that, with perfect CSI, the UL transmit power required to achieve a particular spectral efficiency is inversely proportional to \( M \) [21].

### 2.3.3 Channel Estimation

While the results in Section 2.3.2 are obtained by assuming the availability of perfect CSI at the BS, in practice, the channel has to be estimated from received observations. In conventional MIMO systems, CSI is obtained at the receiver by transmitting an orthogonal pilot from each transmitter antenna. The amount of time-frequency resource used for pilot transmission
is then proportional to the number of transmitter antennas and is independent of the number of receive antennas. Conventional MU-MIMO is implemented in either TDD or frequency-division duplexing (FDD) modes. In the former, the UL and DL transmissions occupy the entire available bandwidth but in separate time-slots, whereas in the latter, the UL and DL transmissions are in different frequency bands but happen simultaneously.

With FDD, since UL and DL transmissions are at different frequency bands, the CSI in both these bands are different. Consequently, the CSI corresponding to both the UL and DL channels need to be available at the BS for precoding and combining. Estimating the UL channel requires the $K$ users to transmit orthogonal pilots with an overhead proportional to $K$. However, for estimating the DL channel, the BS has to transmit $M$ orthogonal pilots (one pilot per antenna element) and the channel estimates obtained at the user terminal have to be fed back to the BS through the UL channel. The DL channel estimation requires an overhead proportional to $M$, which becomes prohibitive when $M$ is large. Furthermore, the channel is coherent over a certain time and frequency, which is referred to as a coherence block. Channels across multiple coherence blocks are assumed independent (in a block-fading model), and therefore, the channel coefficients in each of these coherence blocks have to be separately estimated. For a coherence time of 1 ms (corresponding to a maximum user velocity of 135 km/h at 2 GHz carrier frequency) and coherence bandwidth 200 kHz (corresponding to a maximum delay spread of 5μs), the channel is coherent for 200 kHz × 1 ms = 200 symbols. Since both pilots and data have to be transmitted in a given coherence block which is of the order of a few hundred symbols, the pilot overhead severely constrains the size of the antenna array at the BS. Despite several alternative methods having been proposed in literature for reducing the channel estimation overhead and for designing the precoders and combiners [22, 23, 24, 25], FDD is not a preferred option for implementing massive MIMO [8].

On the other hand, with TDD, if the transmit and receive RF chains at the BS are properly calibrated, the UL and DL channels can be assumed to be reciprocal. Consequently, the channel estimated in the UL can be used for designing the precoder for transmission in the DL. Since channel estimation in the UL can be accomplished with users transmitting orthogonal pilots, the overhead for channel estimation in the UL is proportional only to $K$ and is independent of $M$. Therefore, for a given $K$, $M$ can be arbitrarily large. As in the case with perfect CSI, if each user is assigned a unique pilot, the achievable rate increases logarithmically in $M$ when the estimated CSI is used in conjunction with linear precoding and combining at the BS [21]. However, the UL transmit power required to achieve a particular spectral efficiency decreases as $\sqrt{M}$ instead of as $M$ when perfect CSI is available.

TDD along with channel reciprocity at the BS allows for the pilot se-
quence length to become independent of $M$ and depend only on $K$. However, in practice, it is impractical to assign unique orthogonal pilots when the coherence time is small and when $K$ is large. The latter is typical in multi-cell environments. This necessitates pilot reuse across cells resulting in a phenomenon called pilot contamination. In addition, using large antenna arrays enforces certain hardware constraints on the transceiver architecture, especially at mmWave frequencies. The issue of channel estimation in multi-cell massive MIMO systems in addition to the implication of large-scale antenna arrays on the hardware architecture are described in the next section.

2.4 Challenges

The benefits of massive MIMO, described in the previous sections, are dependent on propagation conditions, channel estimate quality, BS hardware architectures, etc. In this section, we will describe two challenges in massive MIMO which are relevant to this thesis, namely, pilot contamination and the hardware constraints in mmWave transceivers.

2.4.1 Pilot Contamination

As described in Section 2.3.3, the channel is estimated in TDD massive MIMO through UL pilots transmitted by the users. Consider a multi-cell network with $L$ cells and $K$ single-antenna users per cell. We denote user $k$ in cell $\ell$ using the tuple $(\ell, k)$. Let BS $j$ be the reference BS. Then, in order to obtain the channel estimate of $(j, k)$ without interference, each of the $LK$ users in the network has to be assigned a unique orthogonal pilot. However, assigning $LK$ orthogonal pilots necessitates reserving $LK$ symbols in the UL time-slot for pilot transmission. With the channel having to be estimated in every coherence block, the overhead due to pilot transmission becomes prohibitive when $LK$ is large, which is the case in practice. Therefore, in order to reduce the overhead, pilots need to be reused across cells, which leads to interference in the UL and DL. As an illustration, let $\Phi \in \mathbb{C}^{\tau \times \tau}$ be a scaled unitary matrix such that $\Phi^H\Phi = \tau I_\tau$, where $\tau \geq K$ is the length of the UL pilot. We assume for simplicity that user $k$ in each of the $L$ cells uses the $k$th column of $\Phi$, i.e., $\phi_k$ as its pilot sequence. Denoting the channel vector between user $(\ell, k)$ and BS $j$ as $h_{j\ell k} \in \mathbb{C}^M$, if all the pilot transmissions are synchronized, the received observations at BS $j$ during pilot transmission, denoted as $Y_j^{(p)} \in \mathbb{C}^{M \times \tau}$, can be written as

$$Y_j^{(p)} = \sqrt{P_{ul}} \sum_{\ell=1}^L \sum_{k=1}^K h_{j\ell k} \phi_k^T + N^{(p)} \quad (2.21)$$
where $N^{(p)} \in \mathbb{C}^{M \times \tau}$ is the matrix of additive white Gaussian noise (AWGN) at the BS during pilot transmission, with each element being i.i.d and distributed as $\mathcal{CN}(0,1)$. Then, the least-squares (LS) estimate of the channel vector of user $m$ in cell $j$ can be obtained as

$$\hat{h}_{jm} = \frac{1}{\tau \sqrt{P_{ul}}} y_j^{(p)} \phi_m^* = \underbrace{h_{jjm}}_{\text{Desired Channel}} + \sum_{\ell \neq j} \underbrace{h_{j\ell m}}_{\text{Interfering Channels}} + \frac{1}{\tau \sqrt{P_{ul}}} N^{(p)} \phi_m^*.$$  \hfill (2.22)

As a consequence of reusing pilots, it can be seen from (2.22) that the estimate of the channel is contaminated by the channel vectors of users that reuse the same pilot as that of the reference user. Since the error in the channel estimate is correlated with channel vectors of other users in the system, the BS forms beams in the directions of these interfering users in addition to the desired users both in the UL and DL, thereby reducing the beamforming gain for the desired user and causing coherent interference to other users that share the pilots. This phenomenon is referred to as pilot contamination [6, 7, 10]. As a result of pilot contamination, the component corresponding to coherent interference remains as $M \to \infty$ even though the non-coherent interference vanishes. Consequently, when the LS channel estimate is used in an MR or zero-forcing (ZF) precoder/combiner, pilot contamination results in a ceiling on the asymptotic throughput. The UL and DL signal-to-interference-plus-noise ratios (SINRs) of user $(j, m)$ as $M \to \infty$ are given as

$$\text{SINR}_{ul}^{jm} = \frac{\beta_{jjm}^2}{\sum_{\ell \neq j} \beta_{j\ell m}^2}$$  \hfill (2.23)

$$\text{SINR}_{dl} = \frac{\nu_{jm} \beta_{jjm}^2}{\sum_{\ell \neq j} \nu_{\ell m} \beta_{j\ell m}^2}$$  \hfill (2.24)

where the parameters $\nu_{\ell k}, \forall (\ell, k)$ normalize the transmit power at the BS to unity. This is in stark contrast to the result obtained in Section 2.3.2 where the spectral efficiency scales logarithmically in $M$ without bound when perfect CSI is available at the BS. A detailed survey of the literature on mitigating pilot contamination can be found in Chapter 3.

In this thesis, we will consider using SPs for channel estimation. With SP, the pilots are transmitted alongside data at a reduced power. This results in a larger set of pilots becoming available, thereby allowing for a reduced reuse of pilots. Our contribution has also been summarized in Chapter 3.

Recently, it was shown in a seminal work that the ceiling on the achievable throughput due to pilot contamination can be eliminated under certain conditions on the SCM [12, 13]. This result is contingent on the availability...
of estimates of the SCMs of the individual users at the BS. Since covariance matrix estimation is performed in the presence of pilot contamination, estimating contamination-free covariance matrices is challenging and non-trivial.

In this thesis, we propose a novel pilot structure that provides asymptotically contamination-free SCM estimates. We describe the problem of SCM estimation as well as our contribution in detail in Chapter 4.

2.4.2 Hardware Constraints

Massive MIMO is characterized by a large number of antenna elements. An antenna element separation of $\lambda/2$ is sufficient to ensure that the channel coefficients between antenna elements are uncorrelated in an environment with rich scattering, thereby providing diversity and channel hardening, while preventing grating lobes in LOS propagation conditions. However, at sub-6 GHz frequencies, which have been the mainstay of cellular communication for the past four decades, the half-wavelength spacing between antenna elements ranges from 2 cm to 0.2 m. With constraints on the physical dimensions of the antenna arrays, the large inter-element spacing at sub-6 GHz frequencies limits the number of elements in a massive MIMO array.

On the other hand, with mmWave frequencies in the range 30 – 100 GHz, the half-wavelength antenna element spacing is in the range 1.5 – 5 mm allowing for orders of magnitude more antenna elements in the same physical area. Consequently, mmWave MIMO is a more promising candidate for large-scale antenna arrays than conventional sub-6 GHz transceivers. However, the small inter-element spacing and large transmission bandwidths impose constraints on the transceiver architecture which in turn necessitates novel signal processing algorithms. These constraints and their impact on signal processing algorithms are discussed in detail in Chapter 5. In this thesis we develop two algorithms for channel tracking at mmWave frequencies under the mmWave hardware constraints.
Massive MIMO
3. Channel Estimation in TDD Massive MIMO

3.1 Overview

CSI is essential for designing the precoder and combiner at the BS and therefore plays a crucial role in realizing the promised gains of massive MIMO. In massive MIMO, the CSI is estimated at the BS using UL pilots. Under the assumption of channel reciprocity, the estimated CSI is utilized by the BS for both precoding and combining.

Wireless channels are selective in both time and frequency, implying that they are only valid over a finite time-interval and frequency range. A commonly used approach in communication literature to model such doubly-selective channels is block-fading. In this model, the channel vector is assumed to be constant for $T_c$ seconds over a bandwidth of $B_c$ Hz, allowing for a coherence block with $C = B_c T_c$ channel uses, and channel vectors in two different coherence blocks are assumed to be independent. Under this model, it is clear that pilots need to be transmitted every $C$ channel uses to re-estimate the channel. If each user is to be assigned a unique pilot, the channel-estimation overhead increases linearly in the number of users reducing the number of available channel uses for transmitting data.

As demonstrated in Section 2.4.1, reusing pilot sequences results in the channel estimates of a reference user being contaminated by the channel vectors of the users in the neighboring cells [6, 7, 26]. The contaminated channel, when used to design the precoder and combiner results in interference in the DL and UL. It is therefore necessary to mitigate pilot contamination to reduce the amount of interference and in turn increase the spectral efficiency.
3.2 State-of-the-Art in Channel Estimation for TDD Massive MIMO

Several methods have been proposed recently to mitigate the effects of pilot contamination. In a broad sense, these methods can be classified into two types (i) methods that utilize the differences between user transmissions or channel properties of the users to decontaminate pilot contamination at the BS; (ii) methods that avoid or mitigate pilot contamination by modifying user transmissions or pilot structure.

3.2.1 Pilot Decontamination at the BS

As mentioned earlier, pilot contamination is a consequence of not being able to assign each user a dedicated pilot. Methods that decontaminate the pilots at the BS utilize properties such as limited scattering, linear independence between the user covariance matrices, and asymptotic orthogonality between user channels to differentiate between users and separate their channel vectors at the BS.

**Blind Subspace Methods**

As described in Section 2.3.1, under favorable propagation conditions, the channel vectors of any pair of users at the BS are asymptotically orthogonal. Let \( Y \in \mathbb{C}^{M \times C_u} \) be the received observations at the \( M \) BS antennas in the \( C_u \) symbols in the UL time-slot. Let \( X \in \mathbb{C}^{LK \times C_u} \) be the matrix of UL data transmitted by the \( K \) users in each of the \( L \) cells and \( H \in \mathbb{C}^{M \times LK} \) be the channel matrix between these users and the reference BS. Then, \( Y \) can be written as

\[
Y = \sqrt{P_{ul}}HX + N
\]

where, \( N \in \mathbb{C}^{M \times C_u} \) denotes the additive noise at the BS. In turn, the covariance matrix of the received data in a particular time-slot can be written as

\[
R_Y \triangleq \mathbb{E}_{X,N} \{ YY^H \} = P_{ul}HH^H + I = P_{ul}GDG^H + I
\]

where \( \mathbb{E}_{X,N} \{ \cdot \} \) denotes that the expectation is only over the random variables \( X \) and \( N \). As a consequence of asymptotic orthogonality, when \( M \) is large, the eigenvectors of the covariance matrix \( R_Y \) are approximately the channel vectors \( h_k \) upto a scalar multiple [27]. The eigenvectors corresponding to the \( K \) largest eigenvalues are then computed from the sample-covariance matrix \( \hat{R}_Y \triangleq YY^H / C_u \) and an estimate of the scalar multiple is obtained by comparing the eigenvectors to the observations received during pilot transmission. In [28], the channel estimate is decontaminated by projecting the LS estimate of the channel onto a subspace spanned by the eigenvectors of \( \hat{R}_Y \) corresponding to the \( K \) largest eigenvalues.
The efficacy of the blind-subspace methods in [27] and [28] in reducing interference is dependent on the separation between the signal and interference subspaces, which in turn is dependent on the difference between the eigenvalues of $R_Y$. From (3.2), the eigenvalues corresponding to the signal subspace are approximately of the form $MP_{ul}^k + 1$. Therefore, the difference between the eigenvalues is proportional to the difference between the large-scale path-loss coefficients of the users in the reference and interfering cells, and the signal and interference subspaces overlap when the users are close to the cell edges when $M$ is finite. The conditions for separability between the signal and interference subspaces have been quantified in [29].

Blind subspace methods suffer a performance degradation when two users that share the same pilot have similar values of large-scale path-loss coefficient. In [30], the authors propose a maximum a posteriori (MAP)-based channel estimation method which is shown through simulations to significantly outperform the blind subspace-based method while being robust to channel conditions in which the large-scale path-loss coefficients of the reference and interfering users are similar.

Other Semi-Blind Methods

Another form of semi-blind channel estimation is to treat the detected UL data as pilots, effectively increasing the pilot sequence length. In [31], a semi-blind method has been proposed for channel estimation in single-cell massive MIMO systems, wherein each user is assumed to transmit a unique pilot followed by UL data. The channel and data are jointly estimated to maximize the likelihood function using the expectation-maximization (EM) algorithm.

It is shown that as $M \to \infty$ the deterministic Cramér-Rao lower-bound (CRLB) corresponds to the case when all the symbols that are transmitted in the UL are perfectly known. In addition, the stochastic CRLB when $M \to \infty$ corresponds to the case when the users transmit orthogonal pilot sequences for the whole length of the UL time-slot. Simulation results show that the proposed EM method achieves the CRLB at moderately high-SNR and that the channel estimation mean-squared error (MSE) decreases for large $M$ to the value corresponding to when all UL symbols are used for transmitting pilots. These results are useful since they show promise for methods that utilize the transmitted UL data to improve the quality of the channel estimate.

However, this paper considers only the single-cell scenario and therefore, the error component in the MSE corresponds to only the AWGN at the BS. It will therefore be interesting to extend the results to the multi-cell scenario and obtain bounds on the performance of data-aided algorithms.

In [32], the authors propose an iterative channel estimation method in which the estimated payload data are treated as pilots to obtain the linear
minimum mean-squared error (LMMSE) estimate of the channel. Work [33] proposes a space-alternating generalized EM for jointly estimating the channel and data. In [34], the payload symbols are assumed to be drawn from a finite constellation, and some of the estimated symbols after a hard-decision operation are treated as pilots. The set of payload symbols that are treated as pilots are chosen based on a reliability function that is defined in the paper. In these methods, since the users transmit symbols that are independent and zero-mean, the mean-squared distance between the payload data-streams of the reference and interfering users increases with the number of payload-data symbols in the UL time-slot, which in turn improves the efficacy of these methods.

The dependencies (or correlation) within the sequence of payload data transmitted by a user can also be used to separate the user channels at the BS, provided that the nature of this dependency is different for different users. The transmitted data symbols are dependent on each other when they are the output of an error control code, which in modern communication systems is either low-density parity check (LDPC) or turbo codes. An \((n, k)\) code with \(k\) message bits and \(n\) code bits \((n > k)\) contains \(2^k\) valid codewords from a possible set of \(2^n\) codewords. For LDPC codes, the set of valid codewords depends on the permutation used at the interleaver. In [35], the authors propose using different permutations of the interleavers for the reference and interfering users, thereby changing the set of valid codewords between them. Using the linear programming (LP) relaxation of the LDPC decoder, the set of valid codewords are added as constraints to a minimum-variance-based combiner design problem, so that the output of the combiner is always a valid codeword of the reference user. As a result, the combiner actively rejects the interference from the users sharing the same pilots since their set of valid codewords is different from that of the reference user.

**Pilot Decontamination Using Second-order Statistics of the Channel**

The received pilots and data signals at the BS are typically correlated in the spatial and/or temporal dimensions, and different users exhibit different amounts of correlation across these dimensions. The differences between the user covariance matrices can potentially be used to separate the user channels and decontaminate the channel estimates in the respective domains. Moreover, the second-order statistics are valid for a longer duration and have to be estimated less frequently in comparison with the channel vectors. For instance, the SCM of the channel is valid for one to two orders of magnitude longer than the coherence time [36, 37, 38]. Therefore, mitigating pilot contamination using covariance estimates requires a lower overhead than utilizing longer-length pilot sequences.

Channel environments with limited scattering are characterized by a spatially correlated channel. Considering a ULA with steering vector \(a(\theta)\)
corresponding to the AoA $\theta$, the channel vector of user $k$ comprised of $P$ paths can be written as

$$h_k = \frac{1}{\sqrt{P}} \sum_{p=0}^{P-1} \alpha_{kp} a(\theta_{kp})$$

where $\alpha_{kp}$ is the channel coefficient of the $p$th path and is assumed to be distributed as $\mathcal{CN}(0, \beta_k)$.

Let $p_{\theta_k}$ be the probability density function of the AoAs of the paths $\{\theta_{kp}\}_{p=1}^P$, then, assuming that $\{\alpha_{kp}\}_{p=0}^{P-1}$ and $\{\theta_{kp}\}_{p=0}^{P-1}$ are independent, the covariance matrix of the reference user $R_k$ can be written as

$$R_k = E\{h_k h_k^H\} = \beta_k \int_0^\pi a(\theta) a(\theta)^H p_{\theta_k}(\theta) d\theta .$$

Let $\hat{h}_k$ be the LS channel estimate of user $k$, $\tilde{h}_k \triangleq h_k - \hat{h}_k$ be the estimation error, and $R_k^i \triangleq E\{\tilde{h}_k \tilde{h}_k^H\}$ be the SCM of the estimation errors in the LS channel estimate of user $k$. $R_k^i$ is essentially the sum of the SCMs of the users that interfere with user $k$ and the additive noise. Then, the LMMSE channel estimate can be obtained as

$$\hat{h}_{k,\text{LMMSE}} = R_k (R_k + R_k^i)^{-1} \hat{h}_k .$$

If the AoAs of the reference and interfering users are distributed such that the reference and interfering users have non-overlapping angular supports, i.e., the supports of the probability density functions $p_{\theta}(\cdot)$ corresponding to the reference and interfering users do not overlap, it is shown in [11] that the signal spaces of $R_k$ and $R_k^i$ are asymptotically orthogonal to each other. As a result, the MSE of the LMMSE channel estimate (3.5) approaches that for the interference-free case asymptotically in $M$.

Furthermore, the quantity of residual interference is dependent on the amount of overlap of the angular supports and magnitude of the large-scale path-loss coefficients of the interfering users. Since the set of users that interfere with the reference cell is dependent on the pilot allocation, an algorithm is proposed in [11] that assigns pilots so as to minimize the MSE of the channel estimate.

The spatial covariance information can also be used to improve the blind pilot decontamination method in [28]. It was mentioned in Section 3.2.1 that blind methods are less-effective in the finite $M$ regime due to subspace-leakage when the reference and interfering users have similar large-scale path-loss coefficients. On the other hand, the efficacy of the method in [11], detailed in the previous paragraph, is limited by the overlap of the angular supports between the user and interference channels and is less sensitive to the large-scale path-loss coefficients.

Noticing that the methods in [11] and [28] are complementary to each other the authors in [39] propose a covariance-aided blind-pilot decontamination method. In this method, the covariance information is utilized
to design a spatial filter for removing the interference from outside the angular support of the reference user channel in the received observations. The filtered output is then used to compute the sample covariance matrix in [28], and the channel estimate is projected onto the subspace spanned by the eigenvectors corresponding to the $K$ largest eigenvalues. As expected, the proposed method outperforms the methods in [11] and [28].

In the wideband scenario, in addition to the angular domain, variations in the support of the channel impulse response of different users can be used to separate them in the temporal domain. In [40], the authors estimate the wideband channel covariance matrix under the assumption that the channel is sparse in the angular and temporal domains. Since the covariance matrices are estimated in the presence of pilot contamination, the covariance matrix of the reference user is contaminated with the covariance matrices of users that transmit the same pilot. The authors propose a supervised/unsupervised clustering method to obtain the spatio-temporal channel clusters corresponding to the reference user and decontaminate the covariance matrix estimate. The decontaminated covariance matrix estimates are then utilized in an LMMSE channel estimator to eliminate pilot contamination from interfering users that have a different temporal or angular support when compared with that of the reference user.

The aforementioned works utilize the asymptotic orthogonality between the signal subspaces of the covariance matrices of the reference and interfering users to decontaminate the channel estimates. However, when this orthogonality condition is not satisfied, such as in the case of overlapping angular supports in [11], the presence of pilot contamination imposes a ceiling on the asymptotic UL and DL throughput.

In the seminal work [13], the authors observed that the LMMSE channel estimates of the reference and interfering users are asymptotically linearly independent when their covariance matrices are asymptotically linearly independent. The linearly independent channel estimates, when used in a minimum mean-squared error (MMSE) or multi-cell zero-forcing (M-ZF) precoder/combiner, results in the throughput increasing logarithmically in $M$ without bound. The following two assumptions on the SCMs are shown in [13] to result in linearly independent LMMSE channel estimates.

- For any user $k$ in cell $\ell$, the SCM at BS $j$ is such that $\lim \inf_M \frac{1}{M} \text{tr}(R_{jk}) > 0$ and $\lim \sup_M \|R_{jk}\|_2 < \infty$ as $M \to \infty$.

- For any user $k$ in cell $j$ with $\lambda_{jk} \triangleq [\lambda_{j1k}, \ldots, \lambda_{jLk}]^T \in \mathbb{R}^L$ and $\ell' = 1, \ldots, L$

  $$\lim \inf_M \inf_{\lambda_{jk} \lambda_{j'k} = 1} \frac{1}{M} \left\| \sum_{\ell=1}^L \lambda_{j\ell k} R_{j\ell k} \right\|_F^2 > 0. \quad (3.6)$$

Another result derived in [13] is that the unbounded increase in the UL
and DL rates can be obtained even when the element-wise LMMSE channel estimate is used with MMSE precoding/combining, provided the diagonals of the covariance matrices are asymptotically linearly independent. This is an important result from a practical standpoint since computing the element-wise LMMSE estimate is simple and requires estimating only the diagonals of the covariance matrices.

### 3.2.2 Pilot Contamination Avoidance

An alternative approach is to design the pilots transmitted by the users such that the impact of pilot contamination at the BS is minimized. This class of methods focuses on utilizing longer pilots, allocating pilots to users intelligently based on their location or the amount of interference they cause, and designing the pilot sequences based on the channel properties to avoid pilot contamination.

**Protocol-Based Methods**

A simple approach to avoid pilot contamination is to allocate more UL symbols for pilot transmission, thereby allowing for a larger set of orthogonal pilots to be shared across a larger number of cells.

Let \( \tau = rK \) symbols be used for pilot transmission where \( r \) is the pilot reuse factor defined as the number of cells over which the \( \tau \) pilots are shared. When \( r > 1 \), only a subset of the \( \tau \) pilots are used in each cell. Let \( \gamma_{\tau} \) be the SINR for a given \( \tau \), then a lower bound on the UL capacity \( R \) is given as

\[
R = \left(1 - \frac{rK}{C_u}\right) \log_2 (1 + \gamma_{\tau}) .
\] (3.7)

For a particular \( K \), utilizing a larger \( r \) lowers pilot contamination, thereby leading to a higher SINR \( \gamma_{\tau} \). However, a larger \( r \) results in a smaller pre-log factor because of the larger pilot transmission overhead, implying that there exists a trade-off between the pre-log factor and \( \gamma_{\tau} \).

In [41, 42], the authors derive expressions for the ergodic sum UL and DL achievable rates of the users in a cell when the pilots are reused across \( r \) cells. Based on these expressions, the number of users that maximize a lower bound on the ergodic capacity for a particular value of \( r \) is computed in [42]. It is shown that when \( M \to \infty \) the number of users that need to be scheduled to maximize the UL and DL sum throughputs is proportional to the number of symbols in each coherence block \( C \), and that the spectral efficiency is maximum when half of the symbols in each coherence block is allocated for pilot transmission, i.e., \( \tau = C/2 \). Through simulations, the authors analyze the optimal values of \( K \) and \( r \) for a finite \( M \), and conclude that \( r = 3 \) is often a decent choice to maximize the spectral efficiency.

Fractional pilot reuse [43] can be viewed as a generalization of the concept of integer pilot reuse described in [42], where users (rather than cells, as
in the case of integer pilot reuse) are divided into disjoint sets and are allocated pilots based on their large-scale path-loss coefficients.

Users that are at the cell-edge cause significantly higher interference in the UL and DL than users close to their BSs because of the proximity of the former to the interfering BSs. Therefore, a reasonable approach is to allocate the pilots such that users that cause/perceive the most interference are allocated pilots that are reused with a larger pilot reuse factor than users that cause less interference.

A larger number of pilot sequences can also be made available to the BS by overlaying the pilot and data transmissions. Our contributions in Publication I - Publication V are in this context and are explained in greater detail in Section 3.3.

Another approach to avoid pilot contamination is to stagger UL pilot transmissions. In this regard, there are two main approaches: (i) users in a cell transmit pilots when the neighboring cells transmit DL data [44, 45] and (ii) the users in a cell transmit pilots when the users in the neighboring cells transmit UL data [46, 47]. In both cases, the set of cells in the network is divided into $\Gamma$ subsets and the users in a given subset $A_\gamma$ are assumed to transmit their pilots simultaneously.

In the former approach, the reference BS perceives interference from UL pilot transmissions from users in its subset and powerful DL signals from the BSs in the other $\Gamma - 1$ subsets. In [45], it is shown that the interference from DL data transmitted by the BSs in the $\Gamma - 1$ subsets, when users in $A_\gamma$ transmit their pilots, vanishes asymptotically in $M$, and the only remaining interference is the pilot contamination from the interfering users within $A_\gamma$. As a result, it is shown that staggering the UL and DL pilot transmissions results in substantial gains in the asymptotic achievable throughput.

It has to be noted that in [45], only asymptotic expressions for the achievable rate have been derived, and the claims made are based on these expressions. In [46], the authors obtained approximate expressions for the achievable rates for a finite $M$ for both cases (i) and (ii). For the latter approach, in which users of the $\Gamma - 1$ interfering subsets transmit UL payload data when the users in the reference subset transmit pilots, it is shown that the achievable throughput can be improved, with respect to the case when all the pilots are aligned, by varying the transmit powers of the pilots and data. Based on these expressions, [46] concludes that method (i) outperforms (ii) when the number of users is small.

User-Location-Aided Methods
If the SCM can be accurately parameterized with the geographical location of the user, a method similar to [11] can be used to perform pilot allocation.

In [48], the authors assume an elevated BS with a ring of scatterers around the user terminal. The SCM is then parameterized by the mean
angle of the scatterers, given by the physical angle of the user with respect to the BS, and the angular spread which is dependent on the size of the ring and the distance of the user terminal from the BS. In [49], the authors consider a Ricean channel in which the LOS component is parameterized by the location of the user terminal. In both the aforementioned methods, assuming that all the user locations are known, a metric is defined for the amount of interference from an interfering user based on its location and steering angle. The pilots are then allocated to minimize this metric.

The efficacy of location-aided methods is contingent on the availability of precise location information, which may not always be feasible, for example, in the absence of indoor localization. Channel characteristics also vary depending on the nature of the propagation environment around the user. For instance, a user indoors experiences a different angle-delay profile when compared to a pedestrian outdoors. Experimental validation is needed to ensure an accurate mapping between user location and channel knowledge at the BS. Developing methods that are robust to these discrepancies is also an interesting research direction.

**Pilot Design**
UL pilots can be designed to take advantage of the channel statistics as well as modulation characteristics of the transmitted signal, thereby providing additional means to separate users.

In [50], the channel is assumed to be correlated across different UL time-slots and the temporal covariance matrices of the reference and interfering users are utilized for designing the pilot sequences such that the Doppler spectra of the desired and interfering users at the BS do not overlap. The designed pilots are orthogonal in the Doppler domain and thereby avoid pilot contamination.

For wideband orthogonal frequency-division multiplexing (OFDM) transmission, [51] proposed adjustable phase-shift pilots which allow for the user transmissions to be separated in the delay domain.

In [52] the authors extend the method in [51, 50] for the wide-band case. To design the pilots, the authors use a combination of the limited angular scattering of the channel, the small root mean-square (RMS) delay spread of the channel (in comparison with the duration of the OFDM symbol), in conjunction with the temporal covariance matrices to separate the user transmissions in the angle, delay, and Doppler domains, respectively.

**Precoding Methods**
Methods based on precoding have been proposed in [53, 54]. The authors in [54] proposed a distributed single-cell method for precoding for mitigating pilot contamination. The authors in [53] proposed a precoding method that uses coordination between cells to eliminate the effects of pilot contamination, theoretically yielding infinite SINRs.
3.3 Contribution

In Publications I-V, we consider channel estimation and pilot contamination avoidance by using longer length pilot sequences which are obtained by overlaying pilots and data. Publications IV and V are extensions of Publications I-III. Therefore, we will consider only Publications IV and V (henceforth, referred to as [55] and [56], respectively) for the rest of this chapter.

The main contributions of these publications are as follows

- Methods for channel estimation using longer length pilot sequences obtained by overlaying pilots and data.

- Deriving expressions for the achievable rates in the UL and DL with channel estimates obtained using these pilots.

- Formulating and developing the concept and design of a hybrid system which contains users transmitting both SPs and regular pilots (RPs).

- Deriving the Bayesian CRLB of the channel estimate obtained from SP.

- Showing that staggered pilots are a particular case of SPs and can trade off intra-cell interference for channel estimation overhead

3.4 Superimposed Pilots

In conventional communication systems, training-based or semi-blind approaches are typically used for estimating the channel at the receiver. These approaches involve transmitting a known pilot sequence either on dedicated symbols (henceforth referred to as RP) or alongside the data at a reduced power (henceforth referred to as SP).

Methods for channel estimation with SPs have been extensively studied for MIMO systems [57, 58, 59, 60]. However, for conventional MIMO, the data transmitted alongside pilots in SPs results in a poorer quality channel estimate as compared with RP, which in turn reduces the SNR. Earlier works on SP have focused on accommodating this loss in SNR in exchange for a reduced pilot overhead. For instance, when the coherence time of the channel is small, such as in cases of high user-mobility, transmitting dedicated pilots for channel estimation is infeasible/expensive, and SPs are an attractive and viable alternative.

On the other hand, in multi-cell multi-user MIMO, SPs allow for a larger number of orthogonal UL pilots to be transmitted; their number being
limited by the number of symbols in the UL time-slot. Specifically, with SP, the vector transmitted by user \((\ell, k)\) is of the form

\[
s_{\ell k} = \rho x_{\ell k} + \mu p_{\ell k}
\]

(3.8)

where \(x_{\ell k} \in \mathbb{C}^{C_u}\) is the vector of data symbols and \(p_{\ell k} \in \mathbb{C}^{C_u}\) is the pilot. The parameters \(\rho^2\) and \(\mu^2\) are the fractions of transmit power assigned to pilots and data, respectively, such that \(\rho^2 + \mu^2 = 1\). The pilots \(p_{\ell k}\) are taken from the columns of the scaled unitary matrix \(P \in \mathbb{C}^{C_u \times C_u}\). In contrast, with RP, \(s_{\ell k}\) is of the form

\[
s_{\ell k} = [\phi_{\ell k}^T, x_{\ell k}^T]^T,
\]

and the pilots \(\phi_{\ell k}\) are taken from the columns of the scaled unitary matrix \(\Phi \in \mathbb{C}^{\tau \times \tau}\) with \(K \leq \tau \leq C_u\).

With SP, the larger set of pilots facilitates reduced pilot-reuse and a lower inter-cell interference from pilot contamination. However, transmitting pilots alongside data causes intra-cell interference, since the data sequences are not orthogonal across users. This allows the designer to trade-off one type of interference for the other.

SP for massive MIMO has also been considered in [61, 62, 63]. In [61], the authors considered SP for massive MIMO and derived approximate expressions for the UL and DL achievable rates when \(C_u \geq LK\), and compared the performance of SP with RP with \(r_{RP} = 1\). In [62], exact expressions for the UL achievable rates are derived for the general case of \(C_u \leq LK\), and the UL spectral and energy efficiency of SP is compared with that of RP with optimized \(r_{RP}\). The authors of [62] conclude that SP and RP offer similar spectral and energy efficiencies when the pilot reuse factor of RP \(r_{RP}\) is optimized.

In [55] and [56], we compare SP and its variants with RP for channel estimation in massive MIMO. In [55], we obtain approximate expressions for the UL achievable rate when the LS channel estimate obtained from SP is used in an MR combiner, which is then compared with the achievable rate of RP with \(r_{RP} = 1\). These expressions are obtained under the assumption that \(C_u \geq KL\) and inter-cell interference from cells which reuse SPs are negligible. Exact expressions for both UL and DL achievable rates with SP and MR precoder and combiner for the general case of \(C_u \leq KL\) are derived in [56].

Note that the exact expressions for UL throughput derived in [56] is different from that obtained in [62]. Two expressions for the UL achievable rate with SP are obtained in [62]. In obtaining the first expression, the pilot transmitted alongside the data is treated as interference, and consequently, the resulting expression underestimates the achievable rate. Whereas, in obtaining the second expression, the pilots are assumed to be removed perfectly, and consequently, the achievable rate is overestimated. In [56], we side-step the aforementioned issue by multiplying the received observations with a unitary matrix that relegates all the interference from the transmitted pilot to a single UL symbol. This symbol is then discarded since we are only interested in a lower-bound on the UL channel capacity.
The remaining \( C_u - 1 \) symbols are free from the interference caused by the pilot of the reference user, and consequently, standard methods in massive MIMO literature are used to compute an expression for the achievable rate.

The asymptotic MSE and achievable rates of SP are dependent on the fractions of transmit power allocated to data and pilots. The optimal values of \( \rho^2 \) and \( \mu^2 \) are obtained by maximizing a lower bound on the UL rate as [56]

\[
\rho_{\text{opt}}^2 = \left(1 + \sqrt{M\kappa}\right)^{-1} \quad (3.9)
\]

\[
\mu_{\text{opt}}^2 = \left(1 + \frac{1}{\kappa\sqrt{M}}\right)^{-1} \quad (3.10)
\]

where \( \kappa \) is a constant dependent on the large-scale path-loss coefficients and is defined in [56]. Note that in (3.9), \( \rho_{\text{opt}}^2 \) is inversely proportional to \( \sqrt{M} \), as a result of which, the component of the MSE corresponding to the overlaid data decreases at a rate of \( \sqrt{M} \).

In [56], the Bayesian CRLB is derived for SPs under the condition that \( C_u \geq LK \). We impose this condition specifically to evaluate the effect of the data transmission on the estimation error. The MSE is then shown to asymptotically achieve a close approximation of the CRLB. In addition, the MSE for the more general case of \( C_u \leq LK \) asymptotically achieves the MSE of the channel estimate obtained from RP with reuse factor \( r_{\text{RP}} = r_{\text{SP}} \).

Since a lower \( \rho^2 \) implies a better quality channel estimate, the component of the inter and intra-cell interference power from data transmission also reduces at a rate of \( \sqrt{M} \). While this reduction in interference does not significantly improve the UL throughput, there is a notable improvement in the DL achievable rate, and the asymptotic DL throughput of SP with reuse factor \( r_{\text{SP}} \) is the same as that of RP with the same reuse factor.

For finite \( M \), with the aforementioned values of \( \rho^2 \) and \( \mu^2 \), SPs in general, offer a higher UL and DL achievable rate in comparison with RP when \( r_{\text{RP}} = 1 \) [55, 56]. However, SP and RP are shown to be comparable when the pilot reuse factor \( r_{\text{RP}} \) is optimized for a particular coherence length [62, 63].

The performance of SP is limited by self-interference resulting from transmitting pilots alongside data. Since the reference users are closest to the reference cell, the intra-cell interference forms the largest component of interference in both the UL and DL. In order to minimize the impact of this intra-cell interference, we propose the following approaches

- A hybrid system containing both SP and RP
- Staggered pilots in which users in the reference cell utilize orthogonal pilots, thereby eliminating intra-cell interference.
Iterative data-aided channel estimation to estimate and remove the data that is overlaid with pilots.

3.4.1 Hybrid System

One of the main features of SP is that it does not require a dedicated set of symbols for pilot transmission. This property can be used to overlay users transmitting SP over a set of users transmitting RP without causing any interference to the latter, and increase the overall throughput of the network. We consider two approaches to utilize users transmitting SP in a hybrid system.

In the first approach, we add users transmitting SPs to an optimized network (in terms of UL and DL spectral efficiency) with users transmitting only RPs. The combined throughput of both users transmitting RPs and SPs is shown to be higher than that of the network in which users transmit only RPs, as is shown by the following theorem [55].

Theorem 1. In a system that employs time-multiplexed pilots and is designed to maximize the UL and DL sum-rate (such as the scheme described in [42]), let $K$ be the optimal number of users per cell, $L$ be the total number of cells in the system, $\tau > 0$ be the optimal number of symbols used for pilot training, $r^{RP}$ be the optimal pilot-reuse factor, and $C_u - \tau$ and $C_d$ be the number of data symbols in the UL and DL slots, respectively. Then, with $M \to \infty$, there exists a hybrid system, that uses both RP and SP, which is capable of supporting $C_u - \tau$ additional users and offers a higher sum-rate in the UL and DL than the optimal system that only employs time-multiplexed pilots.

The proof of this theorem is straightforward given the structure of the hybrid system in Fig. 3.1. The users that employ SPs cease from transmitting data or pilots when users that employ RPs transmit their pilots. Since the channel vectors of users are asymptotically orthogonal, even with MR precoding and combining, the users transmitting SPs do not interfere with the users transmitting RPs, and as a result, the UL and DL throughput of the users transmitting RPs is unchanged. The users transmitting SPs also offer a non-zero UL and DL throughput, and as a result, the hybrid system offers a higher asymptotic throughput than a system that only employs RPs.

SPs and RPs exhibit complementary behavior with respect to the achievable spectral efficiencies when $r^{RP} = 1$, in the sense that SP offers higher UL and DL throughput with respect to RP when the users are at the edge of a cell and lower UL and DL throughput when the users are close to the BS. In the second approach, given a network with users transmitting only RPs, we utilize the aforementioned observation and propose a simple
framework and algorithm to partition the set of users $\mathcal{U}$ into two disjoint sets $\mathcal{U}_{RP}$ and $\mathcal{U}_{SP}$ which contain users that transmit RPs and SPs, respectively [55, 64, 56]. The partitioning is accomplished by minimizing the total interference in the UL and the DL.

Let $\xi_{ul}$ and $\xi_{dl}$ be non-negative weights for the UL and DL interference powers such that $\xi_{ul} + \xi_{dl} = 1$. Let $I_{RP}^{ul}$ and $I_{RP}^{dl}$ be user $(\ell, k)$’s contribution to the interference power in the UL and DL when assigned to $\mathcal{U}_{RP}$, and similarly, let $I_{SP}^{ul}$ and $I_{SP}^{dl}$ be the corresponding interference powers when user $(\ell, k)$ is assigned to $\mathcal{U}_{SP}$. Then, the weighted interference power in the UL and DL, which is given as

$$T_{RP}^{\ell k} \triangleq \xi_{ul} I_{RP}^{ul} + \xi_{dl} I_{RP}^{dl},$$

is used to construct the objective function. For a given partition $\mathcal{U}_{RP}$ and $\mathcal{U}_{SP}$ of the total set of users $\mathcal{U}$, the sum of the weighted interference of all the users in the UL and DL can be written as

$$I(\mathcal{U}_{RP}, \mathcal{U}_{SP}) = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \left( T_{RP}^{\ell k} 1\{((\ell,k)\in\mathcal{U}_{RP})\} + T_{SP}^{\ell k} 1\{((\ell,k)\in\mathcal{U}_{SP})\} \right).$$

(3.13)

We utilize (3.13) as the objective function to obtain the optimal partitioning $\mathcal{U}_{RP}$ and $\mathcal{U}_{SP}$ through solving the following optimization problem

$$\left(\mathcal{U}_{RP}, \mathcal{U}_{SP}\right) = \arg \min_{\mathcal{U}_{RP}\subseteq \mathcal{U}} \min_{\mathcal{U}_{SP}\subseteq \mathcal{U}} I(\mathcal{U}_{RP}, \mathcal{U}_{SP})$$

subject to $\mathcal{U}_{RP} \cup \mathcal{U}_{SP} = \mathcal{U}$

$$\mathcal{U}_{RP} \cap \mathcal{U}_{SP} = \varnothing$$

(3.14)

where $\varnothing$ is the null-set. The constraints in (3.14) are non-convex and solving the optimization problem requires a complexity that is exponential
in the number of users. In [55, 64], we propose a simple greedy algorithm which utilizes asymptotic expressions of $I(\mathcal{U}_R, \mathcal{U}_S)$ to partition the users into these two sets.

From the simulation results in [56], it is clear that the proposed greedy algorithm exploits the complimentary behavior of the users in the hybrid system and offers a sum UL and DL throughput higher than that of either RP or SP alone.

### 3.4.2 Staggered Pilots

The interference from a user transmitting SP increases with its proximity to the reference BS. As a result, the intra-cell interference resulting from transmitting pilots alongside data is typically the strongest source of interference in the UL and DL. This component of intra-cell interference can be eliminated by assigning orthogonal pilots to users in the reference cell. The resulting pilot structure is called staggered pilots and has been investigated earlier from a different perspective in [46, 47].

Staggered pilots are an intermediate choice between SP and RP when $r_{sp} = r_{rp}$ (c.f. Figs. 3.2-3.4). In Fig. 3.2, we see that no two cells that share the $\tau = r_{rp}K$ pilots interfere with each other since they transmit orthogonal pilots. Only cells that reuse the $r_{rp}$ pilots interfere with each other. On the other extreme, with SP, all users in all cells interfere with each other since the users overlay data with the pilots. Staggered pilots is in between both these extremes since users within a cell do not interfere with each other, whereas two users in different cells do irrespective of whether they share the pilot or not.

A consequence of eliminating the intra-cell interference is that, for the same pilot transmission overhead and with optimized values of $\rho^2$ in (3.9), the DL throughput of staggered pilots is very close to that of RP with reuse factor $r_{sp}$ [56].

In Figs. 3.5 and 3.6, the DL achievable rate of RP with various pilot reuse factors, SP, and staggered pilots is plotted against $M$ when the LS channel estimate is used in an MR and ZF precoder at the BS. The simulation is performed with $L = 91$ hexagonal cells with inter-BS separation of 1km. Each cell has $K = 5$ users. The UL time-slot has $C_u = 35$ symbols. The users are assumed to be uniformly distributed in the cells. More details of the simulation setup can be found in [56]. SP and staggered pilots use a pilot reuse factor of $r_{sp} = 7$.

It can be seen in Fig. 3.5 that the DL throughput of staggered pilots is very close to that of RP with $r_{rp} = 7$ despite requiring only 14.29% of the UL overhead. Similarly, with the ZF precoder in Fig. 3.6, staggered pilots achieve around 90% of the throughput obtained by RP with $r_{rp} = 7$. Given that services such as high-definition video streaming require considerably larger throughput in the DL than in the UL, using staggered pilots is an
Figure 3.2. Pilot structure of RP. When the users in cell $j$ transmit pilots, the users in the remaining $r^{RP} - 1$ cells that share the set of $\tau$ orthogonal pilots effectively transmit zeros. Only users in the cells that reuse the same set of pilots as those of cell $j$ transmit their pilots.

Figure 3.3. Pilot structure of superimposed. The pilot and data transmissions cover the entire UL slot.
3.4.3 Iterative Data-Aided Channel Estimation

As mentioned earlier, the dominating component of interference from the users that share the set of $C_u$ SPs is from the UL data that is transmitted alongside the pilot. This component of interference in the channel estimate can be reduced by jointly estimating the channel and data.

In the absence of any structure within the channel or payload data, the likelihood function is the only available metric for semi-blind channel estimation. Similar to the method proposed in [31], an EM or alternating maximization (AM) can be used to iteratively maximize the log-likelihood and obtain estimates for the channel and data.

However, when the UL symbols are continuous random variables, such as in the case of a Gaussian constellation, such iterative data-aided methods that maximize the likelihood of the channel and/or UL data are not suitable. Since, in the absence of constraints on the UL data, an iterative algorithm can choose $\hat{x}$ to maximize the likelihood by penalizing the MSE of $\hat{h}$. While finite constellations could also exhibit such undesirable behavior, the possibility of that is limited especially with lower-order constellations such as 16-quadrature amplitude modulation (QAM). However, with finite constellations, the optimization problem involving data estimation (even

Figure 3.4. Pilot structure of staggered pilots. In contrast with RP, the pilot transmissions cover the entire UL slot, as with SP. However, in contrast with SP, the users in a cell do not transmit pilots and data simultaneously.
Figure 3.5. DL Sum Rate vs $M$ with MR precoder. SP and Staggered pilots offer an asymptotic DL throughput equivalent to that of RP with $r_{RP} = 7$, even though the UL overhead is only as much as that of RP with $r_{RP} = 1$.

Figure 3.6. DL Sum Rate vs $M$. The solid lines are obtained using the ZF precoder and the dashed lines are obtained with the MF precoder.

after AM) is non-convex, with the complexity increasing exponentially in the number of users in the reference cell.

In [55], a heuristic iterative data-aided algorithm has been proposed under the assumption that the data is drawn from a finite-constellation. With each iteration, the algorithm is guaranteed to maximize the approximate UL SINR. However, since the expressions for the UL SINR are only
approximate, the algorithm is not guaranteed to maximize the actual UL SINR. From simulations, we see that the proposed method significantly reduces the intra-cell interference with lower-order constellations, and that the proposed algorithm performs better than its non-iterative counterpart [55]. Similar improvements in the BER and SINR performance of RPs are also possible, as demonstrated in [31] for the single-cell case.

Note that the method proposed in [55] employs MR combining for data detection. An alternative would be to employ computationally complex non-linear methods such as coded SIC at the BS for data detection, and use this data in the subsequent iteration for estimating the channel. However, one of the salient features of massive MIMO is that simple linear precoding and combining are capacity achieving (with RP and in the absence of pilot contamination), thereby obviating the need for SIC-type methods. Nonetheless, there are preliminary works that consider SIC for massive MIMO with RP [65], and utilizing SIC in iterative data-aided methods could be a potential research direction.

In conclusion, this improvement in both BER and (approximate) SINR performance of the proposed heuristic algorithm makes iterative algorithms practically relevant for mitigating interference when either RP or SP is employed. Developing methods that offer a convergence guarantee while requiring a low computational complexity as well as methods that exploit structure within in transmitted data (such as in [35]), are also interesting problems for future research.
4. Covariance Matrix Estimation for Massive MIMO

4.1 Background and Motivation

As mentioned in Chapter 3, pilot contamination is a consequence of the finite coherence time and coherence bandwidth of the channel since the number of orthogonal pilot sequences that can be transmitted is limited by the number of channel uses in the coherence block. For example, consider a coherence bandwidth of $B_c = 200 kHz$ and a coherence time of $T_c = 1 ms$, which allows for a user communicating at 2 GHz carrier frequency with a $5 \mu s$ channel delay spread at the BS to travel at a velocity of 135 km/h [38]. Hence, for such a user, the channel can be assumed to be static for $B_c T_c = 200$ channel uses which have to be shared between payload data and UL pilots. As a result, longer pilot sequences use up the portion of the coherence interval that can be allocated for data transmission which in turn results in a reduced spectral efficiency.

On the other hand, the SCM varies approximately one to two orders of magnitude slower than the channel vectors [36, 37]. For example, in [38], the channel statistics are assumed to be constant for the system bandwidth of $B_s = 10 MHz$ and frame-length of $T_s = 0.5 s$, which results in the SCM being constant over $\tau_s = (B_s T_s) / (B_c T_c) = 25000$ coherence blocks. Consequently, SCMs need to be estimated less frequently in comparison with the actual channel vectors, and therefore require a lower overhead in comparison with estimating the channel vectors in each coherence block. The coherence block and frame structure for TDD massive MIMO with additional pilots for estimating the SCMs is shown in Fig. 4.1.

Once estimated, these SCMs can be used to decontaminate the channel estimates [11, 12, 13, 40]. In channels with limited scattering and non-overlapping user angular supports, the signal spaces of different user SCMs are asymptotically orthogonal [11], which results in an asymptotically pilot contamination-free LMMSE channel estimate. However, when the angular support of the user channels overlap, the interference resulting
from pilot contamination can be reduced by suitably allocating the pilots to minimize this overlap. In addition, when the SCMs satisfy certain conditions (detailed in Section 3.2.1), the throughput can be shown to scale logarithmically in the number of antenna array elements when the LMMSE channel estimate is used in conjunction with the LMMSE precoder and/or combiner [13].

Since SCMs are valid for a considerably longer duration than the channel coherence time, and since they can be used to decontaminate the channel estimate, utilizing the SCM for pilot decontamination is a preferred alternative to using longer-length pilot sequences. However, estimating these covariance matrices for large antenna arrays in the presence of pilot contamination is challenging for two reasons, namely:

- The number of samples needed to estimate the SCMs increases with the number of elements in the antenna array.

- Estimating the SCMs of individual users in the presence of pilot contamination is not straightforward.

These two issues, and possible workarounds, are discussed in detail in the subsequent section.
Covariance Matrix Estimation for Massive MIMO

\( h_k \sim \mathcal{CN}(0, R_k), \ w_1 \sim \mathcal{CN}(0, I) \)

\( \hat{h}_1 = h_1 + h_2 + h_3 + w_1 \)

\[ \mathbb{E}\{ \hat{h}_1 \hat{h}_1^H \} = \sum_{k=1}^{3} R_k + I \]

![Figure 4.2](image)

**Figure 4.2.** Illustration of the challenge in acquiring the individual user SCM in the presence of pilot contamination. Here, User 1 is the reference user and Users 2 and 3 are interfering users.

### 4.2 Challenges in Estimating the Spatial Covariance Matrices

We denote the SCM of user \((\ell, k)\) at BS \(j\) as \(R_{j\ell k}\). Computing the LMMSE channel estimate requires estimates of \(R_{j\ell k}\) corresponding to the reference users as well as the SCM of the LS channel estimate. Specifically, let \(\hat{h}_{j\ell k}\) be the LS estimate of \(h_{j\ell k}\) at BS \(j\). If \(Q_{jm} \triangleq \mathbb{E}\{\hat{h}_{j\ell m}^* \hat{h}_{j\ell m}^H\}\) and \(R_{j\ell k} \triangleq \mathbb{E}\{h_{j\ell k}^* h_{j\ell k}^H\}\) are the covariance matrices of the channel estimate and the channel vectors, respectively, the LMMSE estimate of the channel of user \((\ell, k)\) is given as

\[
\hat{h}_{j\ell k}^{\text{LMMSE}} = R_{j\ell k} Q_{jm}^{-1} \hat{h}_{j\ell k}.
\]

A large antenna array at the BS is a characteristic feature of massive MIMO. With \(M\) antennas at the BS, at least \(M\) uncorrelated samples of the channel are required to obtain a full-rank SCM. This is important especially in the case of \(Q_{jm}\) since it has to be invertible, which is particularly problematic when \(M\) is large, as in the case of massive MIMO, since the number of samples \(N\) has to scale with \(M\).

A possible workaround is to estimate only the diagonals of the SCM. In [13], it is shown, under certain conditions, that using only the diagonals of the SCM to obtain an element-wise LMMSE channel estimate also results in an unbounded logarithmic increase in the UL/DL spectral efficiency with respect to \(M\). This is a useful result since estimating only the diagonal elements of the SCM requires significantly fewer observations whose number is independent of the dimensions of the antenna array. Therefore,
utilizing the estimates of the diagonals of the SCM is a viable substitute to estimating the full SCM, albeit at the cost of a higher channel estimation MSE and lower spectral efficiency when compared to the latter.

The second issue with estimating the SCMs is that the theoretical results derived in [11, 12, 13] are contingent on the availability of the individual covariance matrices \( R_{j\ell k} \), \( \forall (\ell, k) \). Estimating these SCMs is not straightforward since the user channel vectors \( h_{j\ell k} \), \( \forall (\ell, k) \) are observed in the presence of pilot contamination. Thus, using these contaminated channel estimates for estimating the SCM will result in the estimated SCM being contaminated with the SCMs of other users in the system that employ the same orthogonal pilots. This issue is illustrated in Fig. 4.2 where the objective is to estimate the SCM of User 1. Users 2 and 3 are interfering users belonging to the neighboring cells and transmit the same pilot as User 1. The BS receives the contaminated observation \( \hat{h}_1 \) which when used to estimate \( R_1 \) results in a contaminated SCM estimate.

Note also that, while the issue of estimating a large covariance matrix with limited observations can be addressed by estimating a diagonal surrogate as mentioned earlier, the problem of estimating these diagonal elements in the presence of pilot contamination still remains.

A few recent works have addressed both the aforementioned issues of estimating the large-dimensional SCMs with a small sample-size in the presence of pilot contamination. These works, their shortcomings, and our contribution are discussed in the subsequent sections of this chapter.

### 4.2.1 Estimating Spatial Covariance Matrices When \( N < M \)

When the number of available samples \( N \) to estimate the SCM is less than \( M \), the estimated SCM will have to be regularized to ensure that it has full rank [38].

One approach to regularize the estimated SCM is to shrink it towards a target matrix [66]. Based on the approach described in [66], the authors in [67] propose shrinking the non-diagonal entries of the estimated SCM towards zero and derive the optimal shrinkage coefficient that minimizes the MSE of the estimate. The regularized SCM is guaranteed to be full-rank and positive definite.

Imposing structure on the SCM can also reduce the number of samples required to estimate it. For instance, under limited scattering in the angle/delay domain, the narrow-band/wide-band SCM can be assumed to be low rank, which in turn can be used to reduce the number of samples to estimate it [40].

Under the limited scattering assumption, the channel vector of each user is a linear combination of the steering vectors which are parameterized by the AoAs of the received paths. Utilizing this property, a method such as in [68] can be used to improve the SCM estimate through an iterative
In each iteration, estimates of the AoAs are obtained from the sample SCM using a parametric estimation method such as root-multiple signal classification (MUSIC). The estimated AoAs are then used to refine the sample SCM by eliminating undesirable cross-terms, which are significant when the sample-size used to estimate the SCM is small.

In [48], the authors parameterize the SCM with the location of the user. The received paths at the BS are assumed to be originating from a ring of scatterers around the user. Given the user location, the mean angle of the ring of scatterers is given by the physical angle between the user and the BS, and the angular spread is obtained from the radius of the ring and the distance of the user terminal from the BS. This information can then be used to compute the SCM at the BS. However, the use of such a method requires precise knowledge of the user location, which may not always be available. Moreover, the channel propagation conditions also change with the nature of the environment around the user, and therefore, extensive experimental validation may be required before such methods can be used in practice.

### 4.2.2 Covariance Matrix Estimation in Massive MIMO in the Presence of Pilot Contamination

In [38], the authors propose two methods that employ unique and dedicated pilots for estimating the individual SCMs. In both methods, the estimate $\hat{Q}_{jm}$ is obtained by computing the sample SCM of the LS channel estimate using observations from $N_Q$ coherence blocks and then regularizing it by shrinking its non-diagonal elements. The sample SCM of the LS channel estimate can be obtained as

$$\hat{Q}_{jm}^{(sample)} = \frac{1}{N_Q} \sum_{n=1}^{N_Q} \hat{h}_{jjm}^{(n)} (\hat{h}_{jjm}^{(n)})^H$$

(4.2)

where the superscript $n$ indicates that the LS estimates of the channel have been obtained in the $n$th coherence block.

The sample SCM is then regularized using the method described in [67] as

$$\hat{Q}_{jm} = \gamma \hat{Q}_{jm}^{(sample)} + (1 - \gamma) \hat{Q}_{jm}^{(diag)}$$

(4.3)

where $\gamma \in [0, 1]$ is the shrinkage coefficient and $\hat{Q}_{jm}^{(diag)}$ is a diagonal matrix containing only the diagonal entries of $\hat{Q}_{jm}^{(sample)}$. Note that $\hat{Q}_{jm}$ is a full-rank and positive-definite matrix.

For estimating $R_{j\ell k}$, in the first method, each user transmits a unique pilot for $N_R$ coherence blocks to obtain an LS estimate of its individual channel vector at the reference BS. Then, $R_{j\ell k}$ is straightforwardly obtained by computing the sample SCM of the interference-free channel vectors.
In the second method, all the users that interfere with user \((\ell, k)\) transmit pilots simultaneously in \(N_R\) coherence blocks, while the user \((\ell, k)\) maintains radio silence. The sample SCM obtained from these observations, denoted \(\hat{Q}_{j,-\ell k}\), contains the sum of the SCMs of all the interfering users. \(\hat{R}_{j\ell k}\) is then obtained by subtracting \(\hat{Q}_{j,-\ell k}\) from \(\hat{Q}_{jm}\) as
\[
\hat{R}_{j\ell k}^{(\text{diff})} = \hat{Q}_{jm}^{(\text{sample})} - \hat{Q}_{j,-\ell k}^{(\text{sample})}.
\]
(4.4)

In both the aforementioned methods, \(\hat{R}_{j\ell k}^{(\text{sample})}\) and \(\hat{R}_{j\ell k}^{(\text{diff})}\) converges almost surely to \(R_{j\ell k}\) as \(N_R \to \infty\). When \(N_R \leq M\), the estimated \(\hat{R}_{j\ell k}\) is regularized as in (4.3). However, note that this type of regularization for \(\hat{R}_{j\ell k}^{(\text{diff})}\) need not result in a full-rank or positive-definite matrix. Through simulations, it has been demonstrated in [38] that the second method outperforms the first one. With the second method, the channel estimate using the sample SCMs approaches the LMMSE estimate when \(N_R\) is close to \(M\).

In [69], the pilots are allocated to users in the reference and interfering cells systematically across consecutive coherence blocks. Let the \(LK\) users be indexed with a single index \(k\). Let \(\Pi [n] \in \{0, 1\}^{LK \times \tau}\) with \(\tau < LK\) be the matrix representing the pilot allocation in coherence block \(n\), with \(\Pi [n]_{kp} = 1\) implying that user \(k\) is assigned pilot \(p\), and \(\Omega_p [n]\) be the set of users assigned pilot \(p\) in coherence block \(n\). Then, if the pilot transmissions in all the cells are synchronized, the SCM of the channel estimate of users transmitting pilot \(p\) is given as
\[
Q_p [n] = P_{ul} \sum_{k \in \Omega_p [n]} R_k + I.
\]
(4.5)

Stacking the vectorized versions of \(Q_p [n]\) and \(R_k\) into the matrices \(B_R\) and \(B_{Q[n]}\) defined as
\[
B_R = [\text{vec}(R_1), \ldots, \text{vec}(R_{LK})]
\]
(4.6)
\[
B_{Q[n]} = [\text{vec}(Q_1 [n]), \ldots, \text{vec}(Q_\tau [n])]
\]
(4.7)
we have the relation
\[
B_{Q[n]} = B_R \Pi [n] + \text{vec}(I)1^T.
\]
(4.8)

Note that \(\Pi [n]\) is a tall matrix, and therefore, \(B_R\) cannot be recovered uniquely in (4.8). However, using a set of \(N\) pilot allocations over \(N\) coherence blocks, i.e., \(\Pi = [\Pi_1, \ldots, \Pi_N]\), we can write
\[
B_Q = [B_{Q[1]}, \ldots, B_{Q[N]}] = B_R \Pi + \text{vec}(I)1^T.
\]
(4.9)

If \(\Pi\) is chosen such that it has full row-rank, \(B_R\) can be recovered as
\[
B_R = (B_Q - \text{vec}(I)1^T) \Pi^T
\]
(4.10)
where $(\cdot)^\dagger$ denotes the Moore-Penrose pseudo-inverse.

With this pilot allocation strategy, the maximum-likelihood (ML) estimate of the diagonals of the SCM is obtained in [69] and is shown to converge almost surely to their actual values without requiring dedicated training symbols for estimating the individual SCMs. This method is extended in [70] wherein a framework as well as a greedy algorithm are proposed for optimizing the pilot allocations for estimating the SCM while maximizing a network utility function.

In [71], the SCM estimation problem with time-varying pilot allocation across consecutive coherence blocks is cast as an estimation problem with missing data. The missing data problem is solved using EM and the estimated SCMs are shown to be asymptotically free of pilot contamination.

In [40], an estimate of the wideband channel covariance matrix of a user is recovered from the contaminated covariance matrix using a clustering algorithm. Under the assumption of limited scattering, the channel can be modeled to be composed of a few clusters in the delay and angle domain. A supervised/unsupervised learning algorithm is then proposed to identify the clusters corresponding to the reference user. These clusters are then utilized to extract the individual user covariance matrix from the contaminated covariance matrix.

Note that, unlike for [38] and [69], no theoretical guarantees are available for [40] and [71].

In the aforementioned methods, the authors implicitly assume that the BSs coordinate the pilot transmissions across all cells. For instance, employing unique pilots in [38] across different cells necessitates that all users in the network have perfect timing synchronization and utilize the same symbols in the UL time-slot for pilots. Similarly, the system model and the methods in [69, 70] assume that all the users are perfectly synchronized. While this is a common assumption in massive MIMO literature, assuming that all users are synchronized is infeasible in practice. Furthermore, requiring all the users to use the same set of UL symbols requires coordination between BSs and large cyclic prefixes to account for the time delay from distant users.

4.3 Contributions

In Publication VI (henceforth referred to as [72]), we address the problem of estimating the SCM of an arbitrary user in the presence of pilot contamination and in the absence of BS coordination. The main contribution of this publication is a novel pilot structure for estimating the SCM.
4.4 Estimation using Randomly Phase-Shifted Pilots in Publication VI

In the proposed method, each user in the reference cell is assumed to transmit two pilots in each coherence block. The second pilot is multiplied by a randomly generated phase-shift. The phase-shifts are realizations of the random variables \( \{\Theta_t\}_{t=1}^T \) which are distributed such that \( E\{e^{j\Theta_t}\} = 0, \forall t \) and are mutually independent for different values of \( t \). Here, \( T \) is the number of disjoint subsets into which the set of all the cells in the network is partitioned. The cells inside each of the \( T \geq 1 \) subsets are assumed to be perfectly synchronized.

The LS channel estimates obtained from both the pilots for an arbitrary user \((c, u)\) can be written as

\[
\begin{align*}
\hat{h}_{jcu}^{(1)} &= h_{jcu} + \alpha_{cu}^{(1)} + e_{cu}^{(1)} + w_j^{(1)} \\
\hat{h}_{jcu}^{(2)} &= h_{jcu} + e^{-j\Theta} \alpha_{cu}^{(2)} + e^{-j\Theta} \epsilon_{cu}^{(2)} + e^{-j\Theta} w_j^{(2)}
\end{align*}
\]

where, for \( p \in \{1, 2\} \), \( \alpha_{cu}^{(p)} \) corresponds to the interference from the subset of cells that are synchronized with cell \( c \), \( \epsilon_{cu}^{(p)} \) corresponds to asynchronous UL or DL transmissions from the cells outside the subset, and \( w_j^{(p)} \) is the additive noise. These terms have been defined in [72] and have been omitted here for the sake of brevity.

Since the random variable \( \Theta_t \) is independent of the channel vectors and
Covariance Matrix Estimation for Massive MIMO

<table>
<thead>
<tr>
<th>Method</th>
<th>Supports asynchronous networks</th>
<th>Pilot overhead for estimating ( R_{jℓk} ), ( ∀(ℓ,k) )</th>
<th>Offers theoretical guarantees</th>
</tr>
</thead>
<tbody>
<tr>
<td>[38]</td>
<td>-</td>
<td>( N_RLK )</td>
<td>+</td>
</tr>
<tr>
<td>[69]</td>
<td>-</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>[40]</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Proposed method</td>
<td>+</td>
<td>( NK )</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 4.1. Important attributes of the proposed and state-of-the-art methods for covariance estimation in massive MIMO.

the UL data, we have

\[
R_{j\hat{h}_{jcu}^{(1)}\hat{h}_{jcu}^{(2)}} = \mathbb{E}\left\{ \hat{h}_{jcu}^{(1)} \left( \hat{h}_{jcu}^{(2)} \right)^H \right\} = R_{jcu}. \tag{4.13}
\]

In the absence of the random phase-shift, the channel estimation errors in \( \hat{h}_{jcu}^{(1)} \) and \( \hat{h}_{jcu}^{(2)} \) are correlated, since \( \alpha_{cu}^{(1)} \) and \( \alpha_{cu}^{(2)} \) are correlated. The random phase-shift to the second pilot decorrelates these terms and the resulting sample cross-correlation matrix converges in probability to the true covariance matrix asymptotically in \( N \). However, for a finite number of observations, the sample cross-correlation matrix is not Hermitian symmetric, and therefore needs to be regularized to ensure both Hermitian symmetry and full rank.

A simple illustration of the idea behind the proposed method is given in Fig. 4.3. Each user transmits two pilots. The first pilot is as in Fig. 4.2 and results in a contaminated observation. For the second pilot, user \( k \) applies a random phase-shift \( e^{j\Theta_k} \). The realizations of the random variable corresponding to the reference user \( \Theta_1 \) are assumed to be known at the BS and therefore, the phase-shift on \( h_1 \) can be compensated. It is then straightforward to see that the estimation errors in \( \hat{h}_{11} \) and \( \hat{h}_{12} \) are uncorrelated, and their cross-correlation gives the SCM of the reference user.

In order to obtain the SCM of an arbitrary user, a BS has to be synchronized with the user in question and have knowledge of the realizations \( \theta_{tn} \) of the random variable \( \Theta_t \) corresponding to that user. This requirement is less demanding when compared with that of perfect timing synchronization and/or simultaneous transmission of UL pilots required by existing methods [38, 69, 40]. The attributes of the proposed and state-of-the-art methods for covariance estimation in massive MIMO are summarized in Table 4.1.

In Figs. 4.4 and 4.5, the MSE and UL achievable rate of the proposed method is compared with the method in [38] for different values of \( N \). Figs. 4.4 and 4.5 are plotted for \( L = 7 \) cells with \( M = 100 \) and \( K = 5 \) users per cell. The channel statistics are assumed to be constant for \( \tau_s = 25000 \)
Covariance Matrix Estimation for Massive MIMO

Figure 4.4. Normalized MSE of channel estimate of an average user vs. $N$. Copyright 2018 IEEE.

Figure 4.5. Achievable rate in the UL vs. $N$. Copyright 2018 IEEE.

coeherence blocks. The users are assumed to be distributed at a distance of 120 m from the BS with the BSs separated by 300 m. The received paths are assumed to be uniformly distributed over an angular spread of $20^\circ$ with mean AoA given by the angle between the reference between the BS and the users. It can be seen that the proposed method significantly outperforms the method in [38].

Note that in [72], the system model assumes perfect symbol-level timing synchronization for users in all cells. However, it is straightforward to show that the proposed method works even in the general case when no symbol-level timing or frequency synchronization is assumed across the cells.
5. Channel Estimation and Tracking in Millimeter-Wave MIMO

5.1 Motivation

Carrier frequencies less than 6 GHz have been the mainstay of conventional terrestrial communication links owing to their good propagation properties. The last decade has seen the birth of several new technologies such as IoT, MTC, unmanned aerial vehicles, and self-driving vehicles. These emerging technologies involve numerous devices and vehicles communicating with each other as well as the infrastructure around them, and the limited available bandwidth at sub-6 GHz frequencies is insufficient to cater to the needs of these technologies.

To address this bottleneck, an alternative would be to communicate in the relatively unused mmWave band which have carrier frequencies in the range of 30 - 300 GHz [1, 2, 73]. Owing to their higher carrier frequencies, these communication links offer large contiguous blocks of spectrum of upwards of 1 GHz, which is orders of magnitude larger than what is offered by their sub-6 GHz counterparts [74, 5]. These large chunks of spectrum have the potential to offer unprecedented individual and network-level throughputs and enable the aforementioned applications.

However, mmWave signals face a hostile propagation environment characterized by diffuse scattering, higher penetration losses, and lower diffraction. These propagation effects result in mmWave communication links being predominantly LOS with a few non-line-of-sight (NLOS) clusters. In the next section, we will explain the characteristics of the mmWave propagation environment as well as its impact on the design of both the transceiver architecture and the algorithms for precoding/combining and channel estimation.
5.2 mmWave Channel Characteristics

Since mmWave signals have very small wavelengths in comparison with the objects in the environment that they interact with, they experience a distinct set of propagation characteristics when compared with sub-6 GHz signals. Understanding these propagation effects is essential for designing signal processing algorithms for mmWave transceivers. These propagation effects are detailed in the remainder of this section.

5.2.1 Diffraction

Diffraction is an important propagation characteristic in sub-6 GHz systems since it allows for signal coverage around corners and obstacles [75]. The Fresnel zone between the transmitter and receiver, which is the set of diffracted angles that interfere constructively at the receiver, specifies the diffraction angle. The width of the Fresnel zone is a function of the signal wavelength and becomes narrower with decreasing wavelength, thereby resulting in reduced diffraction at mmWave frequencies. Furthermore, the size of the object required to occlude the Fresnel zone is smaller when it is narrow, implying that small objects can significantly attenuate the signal.

5.2.2 Scattering and Penetration Losses

A radio wave propagating in a medium undergoes reflection when it interacts with an object with a different set of electrical properties than the medium. When the surface of the object is rough, the signal also undergoes diffuse scattering. The fractions of signal power corresponding to specular and diffuse components of the reflected signal is determined by the roughness of the reflecting surface, with the latter component dominating when the surface is rough [76].

Since the effective roughness of a surface increases with decreasing wavelength, mmWave signals experience higher diffuse scattering when compared with sub-6 GHz signals [77]. This scattering results in small-scale fading and rapid variations in the received signal power over travel distances of a few wavelengths [78]. Reflection and scattering are important mechanisms for obtaining coverage in mmWave networks in the absence of the LOS component. Many objects such as clothing, building walls, and trees are excellent reflectors of mmWave signals [79, 80].

mmWave signals experience significant penetration losses from stationary objects in the environment in the range of 2.7 dB to 35.3 dB depending on the type of building material [81]. Brick walls attenuate the signal by 25.3 dB and the human body blockage can attenuate the signal by 30-40 dB [82]. Objects close to the antenna array such as finger or dirt can also alter the beam pattern and significantly attenuate the transmitted/received
signal. As a result, the LOS path, as well as several of the NLOS paths resulting from reflection and scattering are attenuated by objects in the environment. This phenomenon, in conjunction with a lower diffraction, reduces the richness of the scattering environment with only the unblocked LOS path and dominant NLOS clusters providing coverage. Consequently, the mmWave channel is sparse in the angular domain.

5.2.3 Distance-Dependent Path-loss

In addition to penetration losses, mmWave signals experience a higher path-loss for fixed transmitter and receiver gains due to the smaller wavelengths employed. Compensating for these losses by keeping the antenna effective aperture constant (or increasing the gain) imposes constraints on the transceiver hardware.

To demonstrate this, let $P_T$, $G_T$, and $A_T$ be the transmitted power, gain and aperture of the transmit antenna, respectively. Similarly, let $P_R$, $G_R$, and $A_R$ be the corresponding parameters for the receiver. Then, assuming that the transmitter and receiver beams are oriented towards each other, we have from the Friis transmission formula that

$$P_R = \frac{\lambda^2}{(4\pi d)^2} G_R G_T P_T \quad (5.1)$$

where $\lambda$ is the wavelength of the transmitted signal and $d$ is the separation between the transmitter and the receiver. The transmitter (or receiver) gain is related to the effective aperture of the antenna as

$$G_T = \frac{4\pi}{\lambda^2} A_T \quad (5.2)$$

For a fixed $G_T$ and $G_R$, it can be observed from (5.1) that the received power is proportional to the squared of the wavelength. Consequently, communicating at higher frequencies with fixed-gain antennas/antenna arrays increases the path-loss and reduces the received power. In contrast, it can be seen from (5.1) and (5.2) that holding the effective apertures $A_T$ and $A_R$ constant will more than compensate for the path loss [84].

Note that the received power in (5.1) is contingent on the beams of the transmitter and receiver being oriented towards each other. In practice, this is accomplished using steerable antenna arrays in which the antenna elements have to be spaced by at most $\lambda/2$ to prevent undesirable grating lobes. With carrier frequencies between $30 - 300$ GHz, the spacing between the antennas are in the range of $0.5 - 5$ mm. This narrow spacing impacts the choice of RF elements since they have to fit within the limited space available.

Furthermore, the power consumption of an ADC increases linearly with the sampling frequency for a given architecture [85, 86], and running full-resolution analog-to-digital converters (ADCs) on the baseband signal
results in a power consumption in the range 15 - 795 mW at bandwidths of 36 MHz - 1.8 GHz per ADC [87]. The large power consumption of ADCs and other components of the RF chain in conjunction with the narrow spacing between antenna elements renders it infeasible to utilize an RF chain for each antenna element.

5.2.4 Channel Representation

Since mmWave channels have a dominant LOS path with a few NLOS clusters, the channel can be represented as a sum of steering vectors corresponding to these paths. Let $\theta_p, \psi_p, \alpha_p$ be the AoA, AoD and complex channel coefficient of path $p$, respectively. For a ULA with $M$ antenna elements with spacing $\delta$ between the elements, the steering vector corresponding to angle $\theta$ is given as

$$a(\theta) = \left[ 1, e^{-j \frac{2\pi}{\lambda} \sin(\theta)}, e^{-j \frac{2\pi}{\lambda} 2\sin(\theta)}, \ldots, e^{-j \frac{2\pi}{\lambda} (M-1) \sin(\theta)} \right]^T.$$  (5.3)

Let $a^T(\psi)$ and $a_R(\theta)$ be the steering vectors corresponding to the transmitter and receiver. The narrow-band channel $H \in \mathbb{C}^{M_R \times M_T}$ at the receiver can be written as

$$H = \sum_{p=0}^{P-1} \alpha_p a_R(\theta_p) a^T(\psi_p) = \bar{A}_R \bar{D} \bar{A}_H^T$$  (5.4)

where $\bar{A}_R \triangleq [a_R(\theta_0), \ldots, a_R(\theta_{P-1})]$, $\bar{A}_T \triangleq [a_T(\psi_0), \ldots, a_T(\psi_{P-1})]$, and $\bar{D} \triangleq \text{diag} \{\alpha_0, \ldots, \alpha_{P-1}\}$.

Equation (5.4) can easily be extended for the wide-band channel and for the case of two-dimensional arrays by accounting for the time-delay of the received paths and redefining the steering vectors in terms of the azimuth and elevation. However, for the rest of this chapter, we will restrict ourselves to the narrow-band case with the ULA since our contributions in Publication VII and Publication VIII are for this scenario. Representations for wide-band channels with multi-dimensional antenna arrays can be found in [87, 88].

Another representation of (5.4) can be obtained in which the matrix $\bar{D}$ can be replaced by a sparse matrix. This representation allows for the channel estimation problem to be treated as a sparse-recovery problem as will be discussed in Section 5.4. Let $\{\theta_q\}_{q=1}^{G_R}$ and $\{\psi_q\}_{q=1}^{G_T}$ be the set of $G_R$ and $G_T$ quantized AoAs and AoDs, respectively. Then, (5.4) can be written as

$$H = \bar{A}_R \bar{D} \bar{A}_T^H \approx A_R D A_T^H$$  (5.5)

where the columns of $A_R \in \mathbb{C}^{M_R \times G_R}$ and $A_T \in \mathbb{C}^{M_T \times G_T}$ are steering vectors corresponding to the quantized AoAs and AoDs, and $D$ is a sparse
Figure 5.1. Architecture of fully-digital transceivers typically used in sub-6 GHz transceivers.

Figure 5.2. RF elements in a sub-6 GHz transceiver. For the rest of the chapter, the term 'RF chain' will be used to refer to the elements inside the dashed box.

matrix. The approximation in (5.5) is because the error resulting from off-grid AoAs and AoDs are neglected in this representation.

With $A_R$ and $A_T$ replaced by the discrete Fourier transform (DFT) matrices $V_R$ and $V_T$, we obtain a beamspace representation of the channel $H_b$ as

$$H = V_R H_b V_T^H.$$  \hspace{1cm} (5.6)

The matrices $V_R$ and $V_T$ well approximate the left and right singular subspaces of $H$ and therefore, the matrix $H_b$ is sparse.

5.3 mmWave Architectures

Conventional sub-6 GHz MIMO transceivers utilize fully digital architectures wherein each antenna is connected to a dedicated RF chain (c.f.
Fig. 5.3. mmWave MIMO architecture based on hybrid precoding/combining.

Fig. 5.1). The block diagram of the RF chain is shown in Fig. 5.2. Since fully-digital architectures are not feasible for mmWave transceivers with large antenna arrays (as explained in Section 5.2.3), various alternative architectures have been proposed in recent literature to replace the fully-digital transceiver and satisfy the power and space constraints imposed by mmWave communication. These architectures can be broadly divided into two categories, namely,

- A hybrid architecture in which the precoding/combining is performed in both RF and baseband. The RF precoding/combining is implemented using an RF lens or a network of analog phase-shifters and/or switches.

- A low resolution architecture wherein each antenna element has a dedicated RF chain with a low-resolution ADC.

With hybrid beamforming, the number of RF chains and ADCs used is much smaller than the number of antennas, and each antenna is connected to one or more RF chains through a network of analog phase-shifters and/or switches (c.f. Fig. 5.3). The precoding and combining operations are then split across the RF and base-band with the former accomplished using the analog elements.

However, this split is not straightforward since the design of the analog precoder/combiner is a non-convex problem, owing to the element-wise constraints imposed by the analog elements. With phase-shifter networks, these constraints arise from the fact that the phase-shifters can only change the phase of the signal by quantized phase-shifts, constraining the elements of the analog precoding/combining matrix to possess unit-amplitude with quantized phase-shifts.
Switch networks, on the other hand, perform antenna selection and connect the $M_{\text{RF}}$ RF chains to $M_{\text{RF}}$ of the $M$ antenna elements, constraining the analog precoding/combining matrix to contain only binary values. Analog precoders/combiners that are formed with a mix of both phase-shifters and switches will have a combination of both the constraints on the elements of the corresponding matrix.

A variant of the hybrid architecture is a lens-based architecture in which the analog precoding/combining is performed using an RF lens instead of phase-shifters/switches. The $M$ antenna elements are fed to an RF lens front-end which performs the spatial Fourier transform, thereby enabling the $N_s$-dimensional baseband precoder/combiner to access the beamspace of the channel. With a properly designed lens front-end, the different feed antennas to the RF lens excite spatially orthogonal beams. The $N_s$ data streams are then mapped onto these orthogonal beams through a mmWave beam-selector to excite the corresponding antenna elements.

An alternative to the hybrid architecture, in which $M_{\text{RF}} \ll M$ ADCs are used for $M$ antenna elements, is to use a low-resolution ADC at each antenna element. For a given ADC architecture and sampling frequency, the power consumed by an ADC increases exponentially with the number of bits [85, 86]. The lower power requirement allows low-resolution ADCs to be employed for each antenna element (c.f. Fig. 5.4). We will not be discussing low-resolution architectures in further detail. Interested readers can refer [87] and the references therein for more information.

An important consequence of these new architectures is that existing algorithms for designing the precoding and combining matrices in sub-6 GHz MIMO have to be suitably modified to satisfy the aforementioned hardware constraints. Furthermore, as mentioned earlier, mmWave signals experience higher losses due to penetration and scattering, and the resulting channel is predominantly sparse in the angular domain. This channel sparsity can be used to reduce the computational complexity required to
design the precoder and combiner.

The hardware constraints imposed by mmWave communication also affects channel estimation. For instance, with the hybrid architecture, the beams are steered by setting the values of the phase-shifters and/or switches. For a particular choice of phase-shifter values, the transmitter/receiver has access to only a low-dimensional subspace of the $M_R \times M_T$ channel corresponding to the directions of the transmit and receive beams. Therefore, several observations are required to populate the entire channel matrix. This is in contrast to the fully-digital case in which only one observation is required to obtain all the values in a column of the channel matrix. Again, utilizing the sparsity of the channel, the number of observations required to estimate the full channel can be significantly reduced through compressive sensing and sparse recovery.

In the subsequent sections, we will discuss existing methods for channel estimation, channel tracking, and precoder/combiner design for mmWave transceivers with hybrid architectures.

### 5.4 Channel Estimation in mmWave MIMO

Channel estimates are necessary for designing the precoders and combiners. With analog and hybrid architectures, the channel can be accessed only at the output of the RF precoder/combiner, and consequently, the complete $M_R \times M_T$ channel matrix is unobservable at the baseband. As a result, methods for channel estimation for fully-digital transceivers, typically used at sub-6 GHz frequencies, cannot be directly used in the mmWave context.

In addition, as mentioned earlier in Section 5.2, the propagation characteristics of mmWave signals result in a channel that is sparse in the angular domain. The sparsity potentially allows for the transceiver to populate the large $M_R \times M_T$ dimensional matrix $H$ with significantly fewer observations. The low-rank nature of the channel also reduces the overhead required for the receiver to feedback the CSI since only the singular vectors corresponding to the few non-zero singular values have to be fed back.

Channel estimation in mmWave transceivers is typically accomplished using one of three approaches, namely:

- Beam-training
- Compressive sensing and sparse recovery
- Low-rank matrix completion.

With beam-training, the transmitter and receiver generate spatially
orthogonal beams that span non-overlapping angular regions. The receive signal strength is evaluated for each beam pair and the beam-pair with the highest received signal strength is selected for data transmission. A maximum of $M$ spatially orthogonal beams are possible with $M$ antenna elements, and therefore, the transmitter and receiver together can generate a maximum of $M_T \times M_R$ beam pairs. In its naivest form, beam-training involves exhaustively searching through each of the $M_T \times M_R$ beams to find the one that maximizes the received signal power. When $M_T$ and $M_R$ are of the order of hundreds of elements, the estimation overhead becomes prohibitively large, thereby restricting the number of symbols available for data transmission in a coherence block.

A more efficient approach would be to use a set of hierarchical beams from a multi-resolution codebook [89, 90, 91, 92]. With this codebook, the transmitter and receiver initialize beam-training with a set of wide beams that divide the angular domain into a number of sectors. The receiver computes the power of the received signal for each pair of transmit and receive beams/sectors. The pair of sectors at the transmitter and receiver that generate the highest receive signal power are then chosen for the next round, and the receiver feeds back the selected transmit sector to the transmitter. The chosen sectors are then divided into narrower sub-sectors and the iteration proceeds till a target angular resolution or signal power is achieved.

While hierarchical beam-training has been well known for use with analog beamforming [5], an extension to the hybrid architecture is possible in which the beam pattern for a sector is approximated by using all the available RF chains instead of only a single RF chain [91]. The use of multiple RF chains available in a hybrid architecture results in a beam pattern that is closer to the desired beam pattern compared to the case when only a single RF chain is employed. Quite expectedly, the error between the desired beam pattern and the actual one realized with $M_{RF}$ RF chains reduces with increasing $M_{RF}$.

Several variants of the beam-training protocol have been proposed in recent literature [90, 93, 94, 95, 96, 92]. In [90, 93], the beam-training protocol is modified to obviate the need for receiver to feed back the strongest sector to the transmitter. The training overhead in a hierarchical beam-search-type method is reduced by overlapping different beam-training sectors in [96], at the cost of a lower SNR. A method for designing beam patterns for uniform planar arrays by minimizing the MSE between the desired and actual beams has been proposed in [97].

In [94], it is noticed that using only a subset of the antenna elements results in better approximating the broad beams used in the initial stages of the hierarchical beam search algorithm. The authors propose an architecture consisting of a combination of analog phase-shifters and switches, with the latter being used to select the subset of antennas used to generate...
the broad beams.

Beam training can lead to significant variations in the receive power when various beam pairs are tested, resulting in frequent automatic gain controller (AGC) resets. This problem is addressed in [95] by coding the beams with orthogonal codes and spreading out the transmitted energy.

Using only the received power could result in erroneous beam selections at low SNR in the initial stages of the beam-training algorithm. This can be addressed by using hypothesis testing which involves comparing the received signal to a threshold that is obtained by imposing constraints on the false-alarm rate [92].

Beam training is conceptually simple and requires a low computational complexity to implement; the latter feature possibly explaining its adoption in the IEEE 802.11ad Wi-Fi standard. However, in a multi-user scenario, the training overhead scales linearly with the number of users and the per-user training overhead increases with the number of spatially multiplexed streams.

In addition, to estimate a channel with $P$ paths, the overhead is proportional to $P^2$ [91]. If the number of users or streams are large enough such that the channel estimation overhead becomes a bottleneck, more sophisticated methods such as compressive sensing and sparse-recovery have to be used for estimating the channel in the context of multi-user and multi-stream communications.

Compressive sensing and sparse recovery methods utilize the sparse nature of the mmWave channel to reduce the number of observations needed to estimate it [98]. For channel estimation, the transmitter is assumed to use a sequence of $N_T$ precoding vectors $\{f_m\}_{m=1}^{N_T}$ during training. For each of these precoding vectors, the receiver makes $N_R$ measurements using combining vectors $\{w_n\}_{n=1}^{N_R}$. Then, setting $F \triangleq [f_1, \ldots, f_{N_T}]$ and $W \triangleq [w_1, \ldots, w_{N_R}]$, the received observations during training is given as

$$Y = \sqrt{P} W^H F + Q$$

where $Q$ is the noise at the output of the combiner. Vectorizing (5.7) and using the channel representation in (5.5), we obtain

$$y \triangleq \text{vec}(Y) = \sqrt{P} \left( F^T A_T^* \circ W^H A_R \right) \text{diag} \{ D \} + q \approx \sqrt{P} \left( F^T A_T^* \otimes W^H A_R \right) d + q$$

(5.8)

where $q \triangleq \text{vec}(Q)$, $d \triangleq \text{vec}(D)$, and $\circ$ denotes the Khatri-Rao product.

With the measurement matrix defined as $\Psi \triangleq \sqrt{P} \left( F^T A_T^* \otimes W^H A_R \right)$, an estimate of the sparse vector $\hat{d}$ can be recovered from $y$ as

$$\hat{d} = \arg \min_d \| d \|_0$$

subject to $\| y - \Psi d \| < \epsilon$  

(5.9)
where $\epsilon$ is a design parameter. The non-convex optimization problem in (5.9) can be solved by relaxing the $\| \cdot \|_0$ operator with the $\ell_1$ norm or using a sparse recovery algorithm such as orthogonal matching pursuit (OMP) or sparse Bayesian learning (SBL) [99]. The channel estimate $\hat{H}$ is obtained from $\hat{d}$ as

$$\hat{H} = A_R \hat{D} A_T^H. \quad (5.10)$$

The left and right singular vectors of $\hat{H}$ are used to design the precoder and combiner during data transmission after accounting for the hardware constraints [91, 100].

The choice of $F$ and $W$ in (5.7) defines the angular directions in which the channel is probed. A commonly used method is to generate $N_T$ and $N_R$ pseudo-random precoding and combining vectors. With analog or hybrid architectures, the pseudo-random precoding vectors are obtained by setting the analog phase-shifters and/or switches to random values in the constraint set for each measurement. These beamforming vectors generate diffused beams in random directions.

If all the quantized analog phase-shifts $e^{j\theta}$ in the range $\theta \in [0, 2\pi]$ are equally likely, the average received power is the same in all directions. However, when random beamforming is performed with switch-based analog precoders, more power is transmitted towards the broadside of the array. Despite this difference, the channel estimates obtained through phase-shifter and switch-based architectures are shown through simulations to result in similar MSEs [101].

One of the benefits of using pseudo-random precoding and combining vectors is that the overhead for channel estimation in a multi-user environment is independent of the number of users that are associated with a BS, provided the channel is trained in the DL [102]. With this setup, proposed in [102], the BS uses a set of $N_T$ pre-defined pseudo-random vectors $\{f_m\}_{m=1}^{N_T}$ during training. The users then make $N_R$ measurements using randomly generated combining vectors $\{w_n\}_{n=1}^{N_R}$ and reconstruct their channels using any off-the-shelf sparse-recovery algorithm. Each user then feeds back the values and indices of the non-zero elements of $\hat{d}$ to the BS to reconstruct the channel and design the precoder for data transmission.

This ability to estimate the channel of all users simultaneously addresses an important shortcoming of beam-training-based methods, which is that the overhead for channel estimation increases proportionally with the number of users. However, one of the drawbacks with this setup is that the mobile user-terminal, running on a limited energy budget, bears the computational complexity of estimating the channel. The sparse recovery formulation in (5.9) also leads to a biased channel estimate since the AoAs and AoDs are quantized to obtain the measurement matrix.

As with beam-training, several variants of the compressive sensing-based
Channel Estimation and Tracking in Millimeter-Wave MIMO

approaches have been proposed recently [103, 104, 105, 98, 106]. In [103], the authors propose compressed beam-selection in which the channel and the precoder-combiner pair are jointly recovered from the observations when the analog architecture is used at both ends of the communication link.

As mentioned earlier, channel estimation in the DL requires the mobile terminal to solve (5.9). This issue is addressed in [104] wherein the users transmit their training beams using a layered frame structure in the UL and the channel is estimated through a parallel factor analysis (PARAFAC) decomposition at the BS. With this method, it is shown that the training overhead required in the UL is less than in the case when the users employed orthogonal pilots.

References [98] and [105] address the estimation errors resulting from AoA and AoD quantization in (5.5) by modifying the measurement matrix and using a Newton-type method to obtain the off-grid components. A framework to train users in the DL that have diverse SNR and mobility conditions is proposed in [106].

Wide-band channel estimation can also be cast as a sparse-recovery problem [105, 88, 107], and the angular support of the wide-band channel is constant across all sub-carriers [88]. Consequently, the channel can be recovered using sparse-recovery methods that leverage multiple measurement-vectors (MMVs).

An alternative approach to estimate the channel is to use the property that \( H \) is low-rank. Then, \( \hat{H} \) can be recovered from \( Y \) in (5.7) by solving the following optimization problem

\[
\hat{H} = \arg \min_H \ \text{rank}(H) \quad \text{subject to} \quad \| Y - \sqrt{P} W H F \| < \epsilon .
\]  

(5.11)

The non-convex objective in (5.11) can be relaxed by replacing it with the nuclear norm [108] or the atomic norm [109]. The solution to (5.11) does not require quantizing the AoAs and AoDs, and consequently, offers a lower MSE than with grid-based approaches [109].

An alternative to estimating the full CSI is to estimate only the dominant subspaces of the low-rank mmWave channel since the knowledge of the dominant singular vectors of the channel are sufficient for precoder/combiner design (singular values are required for power allocation) [110]. Methods for estimating the dominant signal subspace through Krylov subspace methods and through estimating the low-rank covariance matrix have been proposed in [111] and [112, 113, 114], respectively.
5.5 Hybrid Beamforming

The matrices $F$ and $W$ for channel estimation are chosen such that the directions corresponding to all the singular vectors of the channel are probed, since the objective is to recover the singular vectors corresponding to the strongest singular values. These singular vectors are then used to design $F$ and $W$ for maximizing the spectral efficiency of the communication link.

The beam-training method described in Section 5.4 for channel estimation can also be viewed as a method to design the precoders and combiners from channel observations [87]. In the absence of errors and with a spatially sparse channel, $r$ distinct spatially orthogonal beam-pairs identified during beam-training are approximates of the left and right singular vectors of the channel matrix corresponding to the $r$ largest singular values.

However, with the CSI available at the transmitter and receiver, more sophisticated techniques such as multi-user MIMO and interference cancellation can be applied. In addition, designing $F$ and $W$ to maximize the spectral efficiency results in a higher throughput than with beam-training [100].

With hybrid beamforming, the precoding and combining matrices are of the form $F = F_{RF} F_{BB}$ and $W = W_{RF} W_{BB}$. When $H$ is available at the receiver, the matrices $F_{RF}$, $F_{BB}$, $W_{RF}$, and $W_{BB}$ can be designed to maximize the mutual information which, when Gaussian symbols are transmitted over the mmWave channel, is given by [17]

$$ I(F_{RF}, F_{BB}, W_{RF}, W_{BB}) = \log_2 \left| I_{N_s} + \frac{P_T}{N_s} R_n^{-1} W_{BB}^H W_{RF}^H H F_{RF} F_{BB} F_{BB}^H F_{RF}^H H^H W_{RF} W_{BB} \right| $$

(5.12)

where $R_n = W^H W$ is the noise covariance matrix after combining. Jointly optimizing $F$ and $W$ in (5.12) is often found to be intractable and the non-convex constraints on $F_{RF}$ and $W_{RF}$ make finding the global optimum unlikely [100]. In [100], the authors decouple the optimization over $F$ and $W$ by optimizing over the variables separately.

Assuming that the receiver uses a fully-digital architecture, an approximation of the mutual information can be obtained as [100]

$$ I(F_{RF}, F_{BB}) = \log_2 \left| I_{N_s} + \frac{P_T}{N_s} R_n^{-1} H F_{RF} F_{BB} F_{BB}^H F_{RF}^H H^H \right| $$

$$ \approx \log_2 \left| I_{N_s} + \frac{P_T}{N_s} \Sigma_s^2 \right| - (N_s - \| V_s H F_{RF} F_{BB} \|_F^2) $$

(5.13)

where $\Sigma_s$ is a diagonal matrix containing the $N_s$ largest singular values and $V_s$ contains the corresponding right singular vectors of $H$. Setting $F_{opt} \triangleq V_s$, (5.13) can be approximately maximized by solving the following
optimization problem \cite{100}

\[
\left( F_{\text{RF}}^{\text{opt}}, F_{\text{BB}}^{\text{opt}} \right) = \arg \min_{F_{\text{RF}}, F_{\text{BB}}} \| F_{\text{opt}} - F_{\text{RF}} F_{\text{BB}} \|_F
\]

subject to \( F_{\text{RF}} \in F_{\text{RF}} \)

\[
\| F_{\text{RF}} F_{\text{BB}} \|_F^2 = N_s
\]  

(5.14)

where \( F_{\text{RF}} \) is the set of feasible RF precoders. For analog phase-shifters, \( F_{\text{RF}} \) is the set of \( M_T \times M_{\text{RF}} \) matrices with elements that are unit-amplitude and have quantized phase-shifts. Similarly, for switches, \( F_{\text{RF}} \) is such that each column of \( F_{\text{RF}} \) is a binary vector with a single one and zeros elsewhere \cite{101}.

With a constraint set \( W_{\text{RF}} \) on \( W_{\text{RF}} \), \( W \) can be designed by minimizing the MSE of the data \( s \) at the output of the combiner.

\[
\left( W_{\text{RF}}^{\text{opt}}, W_{\text{BB}}^{\text{opt}} \right) = \arg \min_{W_{\text{RF}}, W_{\text{BB}}} E \left\{ \| s - W_{\text{BB}}^H W_{\text{RF}}^H y \|_2^2 \right\}
\]

subject to \( W_{\text{RF}} \in W_{\text{RF}} \).

(5.15)

Equation (5.15) can be shown to be equivalent to minimizing the weighted Frobenius norm of the difference between the MMSE combiner \( W_{\text{MMSE}} \) and \( W_{\text{RF}} W_{\text{BB}} \) \cite{100}. The latter problem is given as

\[
\left( W_{\text{RF}}^{\text{opt}}, W_{\text{BB}}^{\text{opt}} \right) = \arg \min_{W_{\text{RF}}, W_{\text{BB}}} \left\| \left( y y^H \right)^{1/2} (W_{\text{MMSE}} - W_{\text{RF}} W_{\text{BB}}) \right\|_F
\]

subject to \( W_{\text{RF}} \in W_{\text{RF}} \).

(5.16)

With analog phase-shifters capable of continuous phase-shifts, \( F_{\text{opt}} \) and \( W_{\text{MMSE}} \) can be exactly realized with the hybrid architecture if \( M_{\text{RF}} \geq 2 N_s \) \cite{115}. However, more sophisticated techniques are required when using quantized phase-shifters or switches in the RF precoder.

Using the sparsity of the mmWave channel, \cite{100} proposed using OMP to jointly compute the analog and digital precoders/combiners in (5.14) and (5.16). For the analog phase-shifter-based architecture, and in mmWave channels with limited scattering, the resulting precoder and combiner is shown through simulations to outperform beam-training and approach the throughput of an optimal capacity-achieving unconstrained precoder with water-filling.

Other precoding/combining methods proposed in the existing literature including those for wide-band multi-user scenarios can be found in \cite{116, 115, 117, 118, 119}.

### 5.6 Channel Tracking in mmWave MIMO

mmWave channels are time-varying owing to user-mobility as well as movement of scatterers in the environment. Since large arrays with narrow
beams are used at both ends of the communication link, user mobility results in pointing errors which have to be compensated for in order to maintain sufficient link SNR. Fading in conjunction with penetration losses from blockage mandate frequent re-training.

Moreover, channel estimation with mmWave communication links requires a large overhead since mmWave communication links operate with pre-beamforming SNRs in the range $-20$ dB to $5$ dB owing to the higher path-loss and larger operating bandwidth. This results in a large overhead for initial channel estimation since additional samples are necessary to reduce the noise power.

However, after initial channel estimation and precoder/combiner design, the array gain after beamforming ensures a higher SNR for data transmission. For instance, with perfect CSI, $M_T = 64$ and $M_R = 32$ critically spaced antenna elements at the transmitter and receiver provide an array gain of $10 \log_{10} (M_T M_R) \approx 33$ dB resulting in post-beamforming SNRs of upwards of $13$ dB. Consequently, very few samples are required for tracking variations in the channel in comparison with initial channel estimation since the former operation is performed in the post-beamforming high-SNR region [118].

With beam-training, channel tracking boils down to selecting a set of candidate beams around the estimated AoD/AoA and testing them for the received signal strength. This channel tracking can be performed during data transmission without a separate training interval since beam-training is an energy-based method, meaning that only the received signal strength is used for selecting the beam.

If the channel variation can be modeled, the set of optimal beams can be obtained through dynamic programming by treating the channel estimates as the output of a partially observable Markov decision process [120]. An alternative approach would be to use these models in a Kalman filter to track the parameters of individual channel paths [121, 122, 123]. Models for user mobility can also be used to reduce the overhead needed to track the channel beamspace [124].

In the absence of such models, if the new channel AoD/AoA can be restricted to within a confidence interval around the estimated value, a set of training precoders for channel estimation/tracking can be obtained by minimizing the CRLB of the AoAs or AoDs [118]. Since channel tracking is performed in the high SNR region, the ML estimates of the AoAs and AoDs are close to their CRLB.

With compressive sensing and sparse recovery, a heuristic method for channel tracking has been proposed in [125].

We have proposed two model-free methods for tracking the channel coefficients in Publication VII and Publication VIII. In Publication VII, we overlay training and data information to track abrupt changes in the channel and in Publication VIII we consider receiver-side channel changes.
These contributions are described in greater detail in Section 5.7.

5.7 Contribution

In Publication VII and Publication VIII (henceforth written as [126] and [127], respectively), we propose two sparse-recovery-based methods for channel tracking that do not require a dedicated pilot interval for estimating the support of the channel.

5.7.1 Method Proposed in Publication VII

In [126], we assume that a few of the $M^{RF}$ RF chains are reserved for tracking the channel. Then, the pilot and training are transmitted alongside each other, at different powers, for estimating the channel. Unlike the concept of SP described in Section 3.4 and proposed for use in sub-6 GHz systems, the difference in [126] is that the data is transmitted in the signal space and the pilots are transmitted in the null space of the estimated channel matrix. In the presence of estimation errors, the training sequence transmitted in the null-space of the channel estimate results in a non-zero received signal which can then be used to reconstruct the channel estimation error.

For a large number of antennas $M$, we have the relation [11]

$$\frac{1}{M} a^H(\theta_1) a(\theta_2) \approx \delta(\theta_1, \theta_2).$$

(5.17)

As a result of this asymptotic orthogonality, steering vectors corresponding to paths with AoDs/AoAs that are not in the estimated support of the channel matrix are in the null-space of the channel matrix. This property effectively decouples the data and channel estimation problems when the changed/newly-appeared paths are outside the range space of the channel estimate.

Transmitting energy in the null-space of the channel matrix allows for estimating both changes the parameters of paths in $\hat{H}$ as well as any new paths that may have appeared after initial channel training. The receiver then jointly estimates the channel and data using a SBL-based approach [99, 128].

In [129], a similar approach to channel tracking was developed independently in parallel with our work in [126]. While both methods involve transmitting energy in the range and null-spaces of the channel estimate, an important difference between [129] and [126] is that the former method transmits pilots both in the range and null-spaces of the channel at different powers whereas the latter method involves transmitting data in the range space. In [126], the transmitted data is jointly estimated along with the channel at the receiver using the EM method so as to maximize
the likelihood. Consequently, the choice of the training precoder in [129] is a particular case of [126] when the data is assumed to be known. A side-by-side comparison of the quality of the channel estimate from both methods is an interesting problem for future research.

5.7.2 Method Proposed in Publication VIII

In [127], we propose a blind subspace estimation/tracking algorithm for tracking the channel at the receiver assuming that the AoD at the transmitter is constant whereas the AoA at the receiver can change. Such an assumption is valid in scenarios such as a fixed BS communicating with a mobile terminal (hand-held receiver or a drone) in the DL.

Assuming that the receiver reserves an RF chain for the duration of the training interval, it is shown that estimating the basis of the left singular-subspace from the received observations during data transmission can be cast as a sparse-recovery problem and that the basis can be recovered provided that the transmitter probes all the eigenmodes of the channel. Since two bases for a vector space are related to each other through a non-singular matrix, the ZF combiner can be obtained by estimating this non-singular matrix using $M_{RF}$ pilots. Note that estimating both the subspace and the ZF combiner does not require any knowledge of the transmit precoder.

In Figs. 5.5 and 5.6, the performance of the proposed channel tracking method is compared with the combiner design method proposed in [118]. The algorithms were compared with a transmitter and receiver equipped with $M_T = 64$ and $M_R = 32$ antennas. Both the transmitter and receiver used a hybrid architecture with $M_{RF}^T = M_{RF}^R = 4$ RF chains. At the receiver $M_{RF}^{R-T} = 1$ RF chain out of the $M_{RF}^R$ RF chains is used for channel tracking for the duration of the training interval. The training interval is assumed to be 5120 symbols and the difference in AoA and AoD between subsequent training intervals is assumed to be distributed as $CN(0, \sigma_\theta)$ and $CN(0, \sigma_\psi)$, respectively. Additional details of the simulation setup can be found in [127]. The proposed method performs similar to the method in [118] despite not requiring the knowledge of the precoding matrix at the receiver.

The proposed method can be viewed as a generalization of energy-detection-based channel tracking algorithms such as beam-training (described in Section 5.6) for the hybrid architecture with compressive sensing and sparse recovery. The implementation discussed in [127] has two drawbacks, namely, (i) an additional RF chain is required for channel tracking and (ii) the training combiners are random and therefore do not use the prior information of the channel. We are currently working on designing the training combiners to track the channel, obviate the need for a dedicated RF chain, and obtain the updated combiner without using training.
Figure 5.5. Plot of the average achievable rate vs block index at SNR = 0 dB, $\sigma_\theta = 2^\circ$ and $\sigma_\psi = 0$.

Copyright 2018 IEEE.

Figure 5.6. Plot of the average achievable rate vs $\sigma_\theta$ at SNR = 0 dB and $\sigma_\psi = 0$ at the $n = 50$th block.

Copyright 2018 IEEE.
6. Conclusion and Future Work

In this thesis, we have considered channel estimation in sub-6 GHz and mmWave MIMO.

In the first and second parts of this thesis (Chapters 3 and 4, respectively), we focus on the detrimental phenomenon in massive MIMO systems known as pilot contamination. For mitigating/avoiding pilot contamination, we have proposed SP and its variants namely, staggered pilots and the hybrid system in Chapter 3.

SP, through the hybrid system, has the potential to increase the throughput of a network that consists of users transmitting only RP. Staggered pilots, on the other hand, eliminate the intra-cell component of the interference resulting from transmitting pilots alongside the data, and consequently, offers a significantly higher DL throughput in comparison with RP while requiring the same overhead.

We have also discussed using semi-blind methods to remove a portion of the inter and intra-cell interference resulting from using SP. However, the performance analysis of these methods in the multi-cell scenario with pilot contamination (both for RP and SP) is an unsolved problem.

In Chapter 4, we have considered the problem of estimating the individual user SCMs in the presence of pilot contamination. We have proposed a method for SCM estimation that is capable of providing asymptotically contamination-free covariance matrices without requiring any synchronization between BSs or user terminals. A theoretical analysis of the ergodic achievable spectral efficiency of the proposed method in comparison with other state-of-the-art methods is a topic of interest and is currently being investigated by us.

In the third part of this thesis (Chapter 5), we focus on channel tracking for mmWave MIMO. We show that semi-blind methods have the potential to reduce the overhead as well as the latency due to channel tracking. However, the proposed semi-blind methods require dedicated RF chains for training, and obviating this requirement is also a future research direction.
Conclusion and Future Work
Bibliography


Bibliography


Errata

Publication IV

Equation (52) should contain an ‘=’ sign after the term $\frac{1}{2} Y_j b_{j,m}^{TP}$.

Publication VII

$A_{BS}$ and $A_{MS}$ should be interchanged in Equation (4).

Publication VII

Equation (20) should be $(\Sigma^{(r)})^{-1}$.

Publication VIII

Fig. 1 is not plotted against $\ell$. It is plotted against the number of coherence blocks over which the channel is tracked.
Data generated at a sensing node is valuable only when there is a fast and reliable communication link to ferry it to a computing node for inference. The data generated by machines, with the advent of technologies such as the internet of things, virtual/augmented reality, and self-driving vehicles, and the need for high-speed internet access for conventional mobile users is fuelling the demand for higher communication throughputs. This demand is expected to be fulfilled by 5G and future networks through the use of two main technologies, namely, large-scale antenna arrays and communication at millimeter-wave (mmWave) frequencies. Channel state information (CSI) is essential for utilizing the benefits of large-scale antenna arrays, and acquiring CSI has its associated challenges. In this thesis, we address the problem of CSI acquisition at sub-6 GHz and mmWave frequencies and propose novel pilot structures and algorithms that improve the throughput and reliability of these future networks.