Department of Signal Processing and Acoustics

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

Novel Pilot Structures and Algorithms

Karthik Upadhya

A! Aalto University

DOCTORAL DISSERTATIONS

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

Novel Pilot Structures and Algorithms

Karthik Upadhya

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

Supervising professor
Professor Sergiy A. Vorobyov, Aalto University, Finland

Thesis advisor
Professor Sergiy A. Vorobyov, Aalto University, Finland

Preliminary examiners
Professor David J. Love, Purdue University, USA
Professor Osvaldo Simeone, King's College London, UK

Opponents
Professor A. Lee Swindlehurst, University of California, Irvine, USA
Professor Erik G. Larsson, Linköping University, Sweden

NORDIC ECOLABEL

441   697
Printed matter

Author
Karthik Upadhya

Name of the doctoral dissertation
Channel Estimation in Large-Scale Multi-Antenna Systems for 5G and Beyond

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.

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

Espoo, August 30, 2018,

Karthik Upadhya

Contents

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.

I K. Upadhya, S. A. Vorobyov, M. Vehkapera. Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO. *in Proc. IEEE Int. Conf. Acoustics Speech Signal Process.*, Shanghai, China, pp. 3366-3370, Mar. 2016.

II K. Upadhya, S. A. Vorobyov, M. Vehkapera. Downlink performance of superimposed pilots in massive MIMO systems in the presence of pilot contamination. *Proc. IEEE Global Conf. Signal Inform. Process.*, Washington D.C., pp. 665-669, Dec. 2016.

III K. Upadhya, S. A. Vorobyov, M. Vehkapera. Time-multiplexed / superimposed pilot selection for massive MIMO pilot decontamination. *in Proc. IEEE Int. Conf. Acoustics Speech Signal Process.*, New Orleans, LO, pp. 3459-3463, Mar. 2017.

IV K. Upadhya, S. A. Vorobyov, M. Vehkapera. Superimposed pilots are superior for mitigating pilot contamination in massive MIMO. *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917-2932, Jun. 2017.

V K. Upadhya, S. A. Vorobyov, M. Vehkaperä. Downlink performance of superimposed pilots in massive MIMO systems. Submitted to *IEEE Trans. Wireless Commun.*, Aug. 2017.

VI K. Upadhya, S. A. Vorobyov. Covariance matrix estimation for massive MIMO. *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546-550, Apr. 2018.

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

ing channels in mmWave systems using overlaid data and training. *IEEE 7th Int. Workshop Computational Advances Multi-Sensor Adaptive Process.*, pp. 1-5, Dec. 2017.

VIII K. Upadhya, S. A. Vorobyov, R. W. Heath Jr. Low-overhead receiver-side channel tracking for mmWave MIMO. *in Proc. IEEE Int. Conf. Acoustics Speech Signal Process.*, pp. 3859–3863, Apr. 2018.

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.

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

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

IoT	Internet of things
LDPC	Low-density parity check
LMMSE	Linear minimum mean-squared error
LOS	Line-of-sight
LP	Linear programming
LS	Least-squares
LTE	Long-term evolution
M-ZF	Multi-cell zero-forcing
MAP	Maximum a posteriori
MIMO	Multiple-input multiple-output
ML	Maximum-likelihood
MM	Majorization-minimization
MMSE	Minimum mean-squared error
mMTC	Massive MTC
MMV	Multiple measurement-vector
mmWave	Millimeter-wave
MR	Maximum-ratio
MRC	Maximum-ratio combiner
MSE	Mean-squared error
MTC	Machine-type communication
MU	Multi-user
MUSIC	Multiple signal classification
NLOS	Non-line-of-sight
OFDM	Orthogonal frequency-division multiplexing
OMP	Orthogonal matching pursuit
PARAFAC	Parallel factor analysis
QAM	Quadrature amplitude modulation
RF	Radio frequency
RMS	Root mean-square
RP	Regular pilot

RZF	Regularized zero-forcing
SBL	Sparse Bayesian learning
SCM	Spatial covariance matrix
SINR	Signal-to-interference-plus-noise ratio
SNR	Signal-to-noise ratio
SP	Superimposed pilot
TDD	Time-division duplexing
UL	Uplink
ULA	Uniform linear array
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
WLAN	Wireless local area network
ZF	Zero-forcing

List of Abbreviations

List of Symbols

Latin Letters

$\bar{\boldsymbol{A}}_{\mathrm{R}}$	Matrix with steering vectors corresponding to AoAs as columns
$\bar{\boldsymbol{A}}_{\mathrm{T}}$	Matrix with steering vectors corresponding to AoDs as columns
A_R	Effective aperture of receive antenna
A_T	Effective aperture of transmit antenna
B_c	Coherence bandwidth.
C_{d}	Number of symbols in the DL time slot.
C_{dl}	DL channel capacity.
$\mathbb{C}^{m \times n}$	Set of $m \times n$ complex valued matrices.
C_{u}	Number of symbols in the UL time slot.
C_{ul}	UL channel capacity.
d	Distance between transmitter and receiver.
$\boldsymbol{d}$	Vector of symbols transmitted in the DL.
$\operatorname{diag}\{\cdot\}$	diag operator.
$\mathbb{E}\{\cdot\}$	Expectation operator.
$\boldsymbol{F}$	Precoding matrix
$\boldsymbol{F}_{\mathrm{BB}}$	Baseband Precoding matrix
$\boldsymbol{F}_{\mathrm{RF}}$	RF Precoding matrix
$\boldsymbol{g}_{j\ell k}$	Component of the channel vector corresponding to small-scale fading between user (ℓ, k) and BS j.
G_R	Gain of receive antenna
G_T	Gain of transmit antenna
$\boldsymbol{H}$	Channel matrix between transmitter and receiver.
$\boldsymbol{h}_{j\ell k}$	Channel vector between user (ℓ, k) and BS j.
$I(\mathcal{U}_{\mathrm{RP}}, \mathcal{U}_{\mathrm{SP}})$	Total weighted interference in the hybrid system with user partitions $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$.
$I^{\mathrm{RP-dl}}$	Interference power caused by user (ℓ, k) in the DL if assigned to $\mathcal{U}_{\mathrm{RP}}$.
$I^{\mathrm{RP-ul}}$	Interference power caused by user (ℓ, k) in the UL if as-

	signed to $\mathcal{U}_{\text{RP}}$.
$I^{\text{SP}-\text{dl}}$	Interference power caused by user (ℓ, k) in the DL if assigned to $\mathcal{U}_{\text{SP}}$.
$I^{\text{SP}-\text{ul}}$	Interference power caused by user (ℓ, k) in the UL if assigned to $\mathcal{U}_{\text{SP}}$.
K	Number of users per cell.
L	Number of cells in the network.
M	Number of BS antennas.
$\boldsymbol{n}_{\text{dl}}$	AWGN vector in the DL.
N_s	Number of data streams.
$\boldsymbol{n}_{\text{ul}}$	AWGN vector in the UL.
P	Number of channel paths.
$\boldsymbol{p}_{\ell k}$	Superimposed pilot transmitted by user k of cell ℓ.
P_{dl}	Transmit power in the DL.
P_T	Transmit power
P_{ul}	Transmit power in the UL.
$\boldsymbol{Q}_{jm}$	Spatial covariance matrix of channel estimate of user (j, m).
r	Pilot reuse factor.
$\mathbb{R}_+$	Set of positive real-valued scalars.
$\boldsymbol{R}_{j\ell k}$	Spatial covariance matrix of user (ℓ, k) at BS j.
r^{RP}	Pilot reuse factor for users transmitting RP.
r^{SP}	Pilot reuse factor for users transmitting SP.
$\boldsymbol{s}$	Vector of symbols transmitted in the UL.
$\boldsymbol{s}_{\ell k}$	Vector of symbols transmitted in the UL by user k of cell ℓ.
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.
$\boldsymbol{W}$	Combining matrix
$\boldsymbol{W}_{\text{BB}}$	Baseband Combining matrix
$\boldsymbol{W}_{\text{RF}}$	RF Combining matrix
$\boldsymbol{x}_{\ell k}$	Vector containing UL data transmitted by user k of cell ℓ.
$\boldsymbol{y}_{\text{dl}}$	Received vector in downlink.
$\boldsymbol{y}_{\text{ul}}$	Received vector in uplink.

Greek Letters

α	Complex path-gain
$\beta_{j\ell k}$	Large-scale path-loss coefficient between user (ℓ, k) and BS j.
γ	Shrinkage coefficient.
δ	Antenna spacing
$\delta_{\text{Kr}}(\cdot)$	Kronecker-delta function
θ	Angle of arrival.
λ	Wavelength of the radio signal.

μ	μ^2 is the fraction of UL power allocated to pilot.
ξ^{dl}	Weight for the interference power in the DL.
ξ^{ul}	Weight for the interference power in the UL.
$\boldsymbol{\Pi}[n]$	Pilot allocation matrix in coherence block n
ρ	ρ^2 is the fraction of UL power allocated to data.
σ_k	kth singular value of the channel.
τ	RP pilot length.
$\boldsymbol{\Phi}$	Scaled unitary matrix with columns used as RP.
$\boldsymbol{\phi}_{\ell k}$	RP transmitted in the UL by user (ℓ, k).
ψ	Angle of departure.
$\Omega_p[n]$	Set of users assigned pilot sequence p in coherence block n

Notation

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

17

List of Symbols

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 $\boldsymbol{H} \in \mathbb{C}^{M \times K}$ 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

$$\boldsymbol{y}_{\mathrm{ul}} = \sqrt{\frac{P_{\mathrm{ul}}}{K}} \boldsymbol{H}\boldsymbol{s} + \boldsymbol{n}_{\mathrm{ul}} \tag{2.1}$$

$$\boldsymbol{y}_{\mathrm{dl}} = \sqrt{\frac{P_{\mathrm{dl}}}{M}} \boldsymbol{H}^T \boldsymbol{d} + \boldsymbol{n}_{\mathrm{dl}} \tag{2.2}$$

where P_{ul} and P_{dl} are the powers with which symbols $\boldsymbol{s} \in \mathbb{C}^K$ and $\boldsymbol{d} \in \mathbb{C}^M$ are transmitted in the UL and DL, respectively, and $(\cdot)^T$ denotes the matrix transpose. The vectors $\boldsymbol{s}$ and $\boldsymbol{d}$ are assumed to be such that $\mathbb{E}\left\{\|\boldsymbol{s}\|^2\right\} = K$ and $\mathbb{E}\left\{\|\boldsymbol{d}\|^2\right\} = M$, where $\mathbb{E}\left\{\cdot\right\}$ is the expectation operator and $\|\cdot\|$ is the ℓ^2 norm of a vector. The additive noise vectors $\boldsymbol{n}_{\mathrm{ul}}$ and $\boldsymbol{n}_{\mathrm{dl}}$ 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,

(a) Point-to-point MIMO in the UL

(b) Point-to-point MIMO in the DL
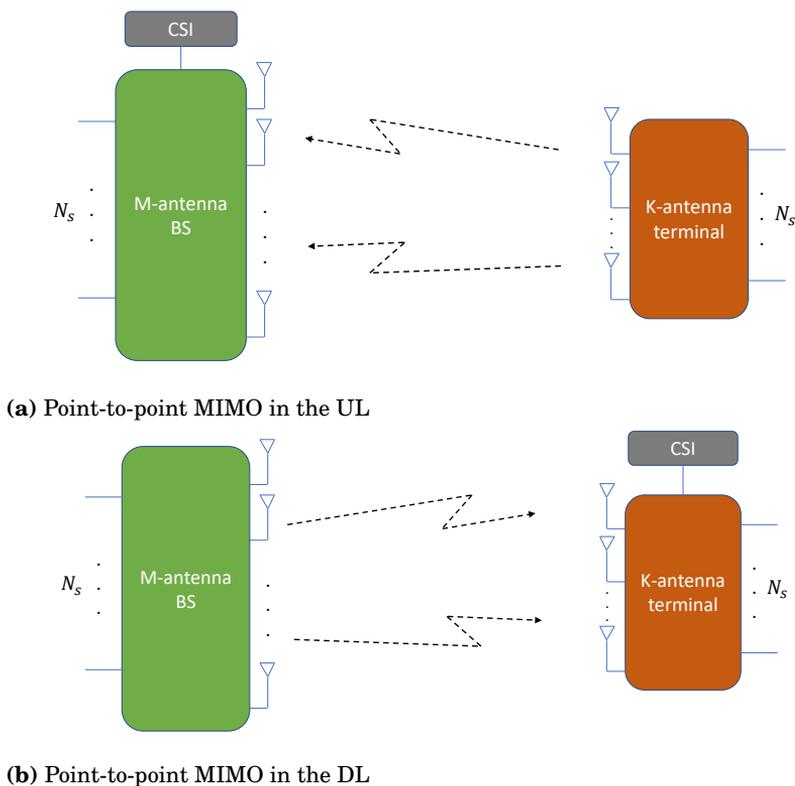
Figure 2.1. Representation of a point-to-point MIMO communication link in (a) the UL; (b) the DL.

the channel capacity in the UL and DL can then be obtained as [17]

$$C_{\text{ul}} = \log_2 \left| \boldsymbol{I}_K + \frac{P_{\text{ul}}}{K} \boldsymbol{H}^H \boldsymbol{H} \right| = \sum_{k=1}^{\min\{M,K\}} \log_2 \left(1 + \frac{P_{\text{ul}}}{K} \sigma_k^2 \right) \qquad (2.3)$$

$$C_{\text{dl}} = \log_2 \left| \boldsymbol{I}_M + \frac{P_{\text{dl}}}{M} \boldsymbol{H} \boldsymbol{H}^H \right| = \sum_{k=1}^{\min\{M,K\}} \log_2 \left(1 + \frac{P_{\text{dl}}}{M} \sigma_k^2 \right) \qquad (2.4)$$

where σ_k is the kth singular value of $\boldsymbol{H}$ and $(\cdot)^H$ denotes conjugate transpose.

The channel capacities C_{ul} and C_{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 $\boldsymbol{h}_k \in \mathbb{C}^M$ as the channel between user k and the BS, the received symbols in the UL can be written as

$$\boldsymbol{y}_{\mathrm{ul}} = \sqrt{P_{\mathrm{ul}}} \sum_{k=1}^{K} \boldsymbol{h}_k s_k + \boldsymbol{n}_{\mathrm{ul}} = \sqrt{P_{\mathrm{ul}}} \boldsymbol{H} \boldsymbol{s} + \boldsymbol{n}_{\mathrm{ul}} \tag{2.5}$$

where $\boldsymbol{H} \triangleq [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]$ and $\boldsymbol{s} \triangleq [s_1, \ldots, s_K]^T$. The UL symbols s_k are normalized such that $\mathbb{E}\left\{|s_k|^2\right\} = 1$. The received signal by user k in the DL can be written as

$$y_{\mathrm{dl},k} = \sqrt{P_{\mathrm{dl}}} \boldsymbol{h}_k^T \boldsymbol{d} + n_{\mathrm{dl},k} \;. \tag{2.6}$$

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

$$\boldsymbol{y}_{\mathrm{dl}} = \sqrt{P_{\mathrm{dl}}} \boldsymbol{H}^T \boldsymbol{d} + \boldsymbol{n}_{\mathrm{dl}} \tag{2.7}$$

where $\boldsymbol{n}_{\mathrm{dl}} \triangleq [n_{\mathrm{dl},1}, \ldots, n_{\mathrm{dl},K}]^T$. 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.
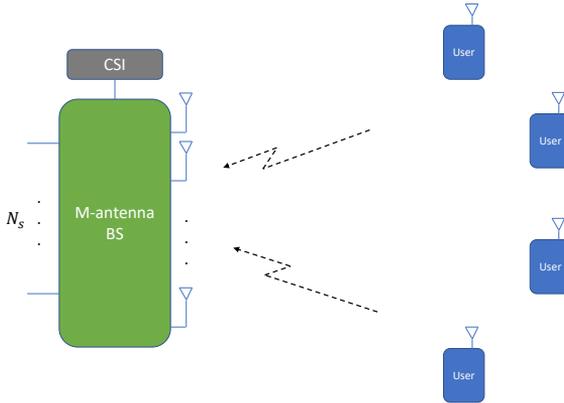
(a) MU-MIMO in the UL.
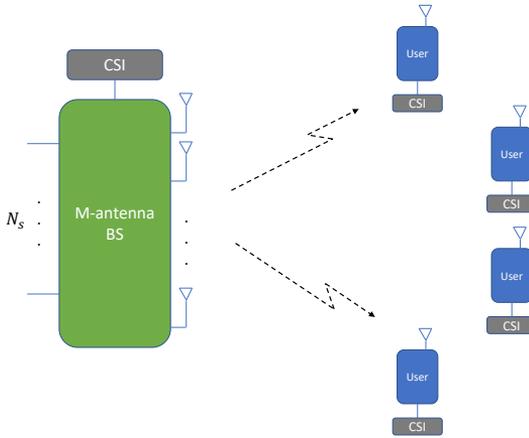
(b) MU-MIMO in the DL.

Figure 2.2. Representation of an MU-MIMO system in (a) the UL; (b) the DL.

capacity in the UL and DL is then given as [18]

$$C_{\mathrm{ul}} = \log_2 \left| \boldsymbol{I}_K + P_{\mathrm{ul}} \boldsymbol{H}^H \boldsymbol{H} \right| \tag{2.8}$$

$$C_{\mathrm{dl}} = \max_{\substack{\boldsymbol{p}^T \boldsymbol{1} = 1 \\ \boldsymbol{p} \geq \boldsymbol{0}}} \log_2 \left| \boldsymbol{I}_M + P_{\mathrm{dl}} \boldsymbol{H} \boldsymbol{P} \boldsymbol{H}^H \right| \tag{2.9}$$

where $\boldsymbol{P} = \mathrm{diag}\{\boldsymbol{p}\}$ is a diagonal matrix with the elements of $\boldsymbol{p}$ 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 non-linear 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 $\boldsymbol{h}_k \in \mathbb{C}^M$ 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.,

$$\frac{1}{M}\boldsymbol{h}_k^H \boldsymbol{h}_{k'} \to 0, \ \forall \ k \neq k' \ . \tag{2.10}$$

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., $\boldsymbol{h}_k \sim \mathcal{CN}(0, \beta_k \boldsymbol{I}_M)$, 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

$$\frac{1}{M}\boldsymbol{h}_k^H \boldsymbol{h}_{k'} \xrightarrow{\text{a.s.}} \frac{1}{M}\mathbb{E}\left\{\boldsymbol{h}_k^H \boldsymbol{h}_{k'}\right\} = 0, \ \forall \ k \neq k' \tag{2.11}$$

$$\frac{1}{M}\|\boldsymbol{h}_k\|^2 \xrightarrow{\text{a.s.}} \frac{1}{M}\mathbb{E}\left\{\|\boldsymbol{h}_k\|^2\right\} = \beta_k, \ \forall \ k \tag{2.12}$$

where $\xrightarrow{\text{a.s.}}$ 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* [6, 10] and results in the inner product of the channel vector $\boldsymbol{h}$ and a beamforming vector $\boldsymbol{v}$, which is obtained from $\boldsymbol{h}$, becoming equal to its average value asymptotically, i.e., $\boldsymbol{v}^H \boldsymbol{h} \xrightarrow{a.s.} \mathbb{E}\left\{\boldsymbol{v}^H \boldsymbol{h}\right\}$.

In addition, consider the narrow-band LOS channel with AoA θ 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 θ can be defined as $\boldsymbol{a}(\theta) \triangleq \left[1, e^{\mathrm{j}\pi \sin(\theta)}, e^{\mathrm{j}\pi 2 \sin(\theta)}, \ldots, e^{\mathrm{j}\pi(M-1)\sin(\theta)}\right]^T$. Then the channel vector of user k with AoA θ_k can be written as [19]

$$\boldsymbol{h}_k = \sqrt{\beta_k} e^{-\mathrm{j}\frac{2\pi d_k}{\lambda}} \boldsymbol{a}(\theta_k) \tag{2.13}$$

where d_k is the distance between user k and the reference BS and λ is the wavelength of the carrier. For large antenna arrays, we have the property [11]

$$\frac{1}{M} \boldsymbol{a}^H(\theta_k) \boldsymbol{a}(\theta_{k'}) \to \delta_{\mathrm{Kr}}(\theta_k, \theta_{k'}) \tag{2.14}$$

where $\delta_{\mathrm{Kr}}(\cdot)$ 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 $\boldsymbol{h}_k$ can be written as

$$\boldsymbol{h}_k = \sqrt{\beta_k} \boldsymbol{g}_k \tag{2.15}$$

where $\boldsymbol{g}_k \in \mathbb{C}^M$ accounts for the small-scale fading or the array response. Setting $\boldsymbol{D} \triangleq \mathrm{diag}\left\{\beta_1, \ldots, \beta_K\right\}$ and $\boldsymbol{G} \triangleq [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K]$, the channel matrix $\boldsymbol{H}$ can be written as $\boldsymbol{H} = \boldsymbol{G}\boldsymbol{D}^{1/2}$. As a consequence of asymptotic orthogonality, we have [7]

$$\boldsymbol{H}^H \boldsymbol{H} = \boldsymbol{D}^{1/2} \boldsymbol{G}^H \boldsymbol{G} \boldsymbol{D}^{1/2} \approx M\boldsymbol{D} \tag{2.16}$$

Substituting (2.16) into (2.8) and (2.9), the asymptotic capacity can be obtained as [7]

$$C_{\mathrm{ul}} = \log_2 \left| \boldsymbol{I}_K + P_{\mathrm{ul}} \boldsymbol{H}^H \boldsymbol{H} \right|$$

$$\approx \log_2 \left| \boldsymbol{I}_K + P_{\mathrm{ul}} M \boldsymbol{D} \right|$$

$$= \sum_{k=1}^{K} \log_2 \left(1 + P_{\mathrm{ul}} M \beta_k \right) \tag{2.17}$$

$$C_{\mathrm{dl}} = \max_{\substack{\boldsymbol{p}^T \boldsymbol{1} = 1 \\ \boldsymbol{p} \geq \boldsymbol{0}}} \log_2 \left| \boldsymbol{I}_M + P_{\mathrm{dl}} \boldsymbol{H} \boldsymbol{P} \boldsymbol{H}^H \right|$$

$$\approx \max_{\substack{\boldsymbol{p}^T \boldsymbol{1} = 1 \\ \boldsymbol{p} \geq \boldsymbol{0}}} \log_2 \left| \boldsymbol{I}_K + P_{\mathrm{dl}} M \boldsymbol{P} \boldsymbol{D} \right| \ . \tag{2.18}$$

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]

$$\widehat{\boldsymbol{s}} = \boldsymbol{H}^H \boldsymbol{y}_{\mathrm{ul}} = \sqrt{P_{\mathrm{ul}}} \boldsymbol{H}^H \boldsymbol{H} \boldsymbol{s} + \boldsymbol{H}^H \boldsymbol{n}_{\mathrm{ul}} \approx \sqrt{P_{\mathrm{ul}}} M \boldsymbol{D} \boldsymbol{s} + \boldsymbol{H}^H \boldsymbol{n}_{\mathrm{ul}} \ . \tag{2.19}$$

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

$$\boldsymbol{y}_{\mathrm{dl}} = \sqrt{\frac{P_{\mathrm{dl}}}{M}} \boldsymbol{H}^T \boldsymbol{H}^* \boldsymbol{D}^{-1/2} \boldsymbol{P}^{1/2} \boldsymbol{x} + \boldsymbol{n}_{\mathrm{dl}} \approx \sqrt{P_{\mathrm{dl}} M} \boldsymbol{D}^{1/2} \boldsymbol{P}^{1/2} \boldsymbol{x} + \boldsymbol{n}_{\mathrm{dl}} \ . \tag{2.20}$$

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 $\times$ 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 ℓ using the tuple (ℓ, 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 $\boldsymbol{\Phi} \in \mathbb{C}^{\tau \times \tau}$ be a scaled unitary matrix such that $\boldsymbol{\Phi}^H \boldsymbol{\Phi} = \tau \boldsymbol{I}_\tau$ 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 kth column of $\boldsymbol{\Phi}$, i.e., $\boldsymbol{\phi}_k$ as its pilot sequence. Denoting the channel vector between user (ℓ, k) and BS j as $\boldsymbol{h}_{j\ell k} \in \mathbb{C}^M$, if all the pilot transmissions are synchronized, the received observations at BS j during pilot transmission, denoted as $\boldsymbol{Y}_j^{(p)} \in \mathbb{C}^{M \times \tau}$, can be written as

$$\boldsymbol{Y}_j^{(p)} = \sqrt{P_{\mathrm{ul}}} \sum_{\ell=1}^{L} \sum_{k=1}^{K} \boldsymbol{h}_{j\ell k} \boldsymbol{\phi}_k^T + \boldsymbol{N}^{(p)} \qquad (2.21)$$

where $\boldsymbol{N}^{(p)} \in \mathbb{C}^{M \times \tau}$ 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

$$\widehat{\boldsymbol{h}}_{jjm} = \frac{1}{\tau\sqrt{P_{\mathrm{ul}}}}\boldsymbol{Y}_j^{(p)}\boldsymbol{\phi}_m^* = \underbrace{\boldsymbol{h}_{jjm}}_{\text{Desired Channel}} + \underbrace{\sum_{\ell \neq j}\boldsymbol{h}_{j\ell m}}_{\text{Interfering Channels}} + \frac{1}{\tau\sqrt{P_{\mathrm{ul}}}}\boldsymbol{N}^{(p)}\boldsymbol{\phi}_m^* \,.$$

(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

$$\mathrm{SINR}_{jjm}^{\mathrm{ul}} = \frac{\beta_{jjm}^2}{\sum\limits_{\ell \neq j}\beta_{j\ell m}^2}$$

(2.23)

$$\mathrm{SINR}_{\mathrm{dl}} = \frac{\nu_{jm}\beta_{jjm}^2}{\sum\limits_{\ell \neq j}\nu_{\ell m}\beta_{\ell jm}^2}$$

(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 $Y \in \mathbb{C}^{M \times C_\mathrm{u}}$ be the received observations at the M BS antennas in the C_u symbols in the UL time-slot. Let $X \in \mathbb{C}^{LK \times C_\mathrm{u}}$ 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

$$Y = \sqrt{P_\mathrm{ul}} H X + N \tag{3.1}$$

where, $N \in \mathbb{C}^{M \times C_\mathrm{u}}$ 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

$$R_Y \triangleq \mathbb{E}_{X,N}\left\{ Y Y^H \right\} = P_\mathrm{ul} H H^H + I = P_\mathrm{ul} G D G^H + I \tag{3.2}$$

where $\mathbb{E}_{X,N}\left\{\cdot\right\}$ 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 $\widehat{R}_Y \triangleq Y Y^H / C_\mathrm{u}$ 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 $\widehat{R}_Y$ 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 $\boldsymbol{R}_Y$. From (3.2), the eigenvalues corresponding to the signal subspace are approximately of the form $MP_{\mathrm{ul}}\beta_k + 1$. 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 θ, the channel vector of user k comprised of P paths can be written as

$$\boldsymbol{h}_k = \frac{1}{\sqrt{P}} \sum_{p=0}^{P-1} \alpha_{kp} \boldsymbol{a}\left(\theta_{kp}\right) \tag{3.3}$$

where α_{kp} is the channel coefficient of the pth path and is assumed to be distributed as $\mathcal{CN}\left(0, \beta_k\right)$.

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 $\boldsymbol{R}_k$ can be written as

$$\boldsymbol{R}_k = \mathbb{E}\left\{\boldsymbol{h}_k \boldsymbol{h}_k^H\right\} = \beta_k \int_0^{\pi} \boldsymbol{a}\left(\theta\right) \boldsymbol{a}\left(\theta\right)^H p_{\theta_k}\left(\theta\right) d\theta \,. \tag{3.4}$$

Let $\widehat{\boldsymbol{h}}_k$ be the LS channel estimate of user k, $\widetilde{\boldsymbol{h}}_k \triangleq \boldsymbol{h}_k - \widehat{\boldsymbol{h}}_k$ be the estimation error, and $\boldsymbol{R}_k^{\mathrm{i}} \triangleq \mathbb{E}\left\{\widetilde{\boldsymbol{h}}_k \widetilde{\boldsymbol{h}}_k^H\right\}$ be the SCM of the estimation errors in the LS channel estimate of user k. $\boldsymbol{R}_k^{\mathrm{i}}$ 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

$$\widehat{\boldsymbol{h}}_k^{\mathrm{LMMSE}} = \boldsymbol{R}_k \left(\boldsymbol{R}_k + \boldsymbol{R}_k^{\mathrm{i}}\right)^{-1} \widehat{\boldsymbol{h}}_k \,. \tag{3.5}$$

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 $p_\theta\left(\cdot\right)$ corresponding to the reference and interfering users do not overlap, it is shown in [11] that the signal spaces of $\boldsymbol{R}_k$ and $\boldsymbol{R}_k^{\mathrm{i}}$ 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 be 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 ℓ, the SCM at BS j is such that $\liminf_M \frac{1}{M} \text{trace}(\boldsymbol{R}_{j\ell k}) > 0$ and $\limsup_M \|\boldsymbol{R}_{j\ell k}\|_2 < \infty$ as $M \to \infty$.

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

$$
\liminf_M \inf_{\{\lambda_{jk}:\lambda_{j\ell'k}=1\}} \frac{1}{M} \left\| \sum_{\ell=1}^{L} \lambda_{j\ell k} \boldsymbol{R}_{j\ell k} \right\|_F^2 > 0 \, . \tag{3.6}
$$

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 τ pilots are shared. When $r > 1$, only a subset of the τ pilots are used in each cell. Let γ_τ be the SINR for a given τ, then a lower bound on the UL capacity R is given as

$$R = \left(1 - \frac{rK}{C_{\mathrm{u}}}\right) \log_2\left(1 + \gamma_\tau\right) \ . \tag{3.7}$$

For a particular K, utilizing a larger r lowers pilot contamination, thereby leading to a higher SINR γ_τ. 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 γ_τ.

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 Γ subsets and the users in a given subset A_γ 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_γ transmit their pilots, vanishes asymptotically in M, and the only remaining interference is the pilot contamination from the interfering users within A_γ. 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 timeslots 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 (ℓ, k) is of the form

$$s_{\ell k} = \rho x_{\ell k} + \mu p_{\ell k} \tag{3.8}$$

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 ρ^2 and μ^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 $s_{\ell k} = \left[\phi_{\ell k}^T, x_{\ell k}^T \right]^T$, 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 $C_{\mathrm{u}} - 1$ 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 ρ^2 and μ^2 are obtained by maximizing a lower bound on the UL rate as [56]

$$\rho_{\mathrm{opt}}^2 = \left(1 + \sqrt{M}\kappa\right)^{-1} \tag{3.9}$$

$$\mu_{\mathrm{opt}}^2 = \left(1 + \frac{1}{\kappa\sqrt{M}}\right)^{-1} \tag{3.10}$$

where κ is a constant dependent on the large-scale path-loss coefficients and is defined in [56]. Note that in (3.9), ρ_{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 $C_{\mathrm{u}} \geq LK$. 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 $C_{\mathrm{u}} \leq LK$ asymptotically achieves the MSE of the channel estimate obtained from RP with reuse factor $r^{\mathrm{RP}} = r^{\mathrm{SP}}$.

Since a lower ρ^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^{SP} is the same as that of RP with the same reuse factor.

For finite M, with the aforementioned values of ρ^2 and μ^2, SPs in general, offer a higher UL and DL achievable rate in comparison with RP when $r^{\mathrm{RP}} = 1$ [55, 56]. However, SP and RP are shown to be comparable when the pilot reuse factor r^{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 $C_{\mathrm{u}} - \tau$ 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 $C_{\mathrm{u}} - \tau$ 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 $r^{\mathrm{RP}} = 1$, 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
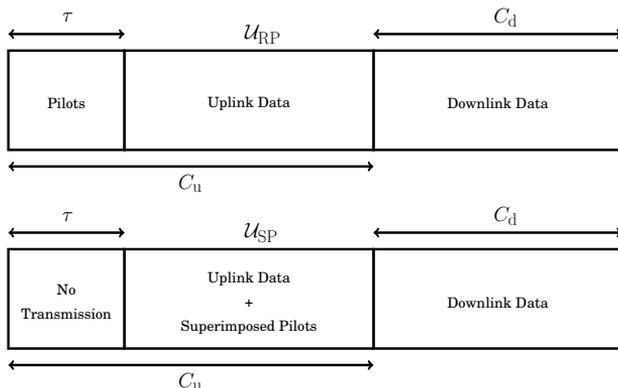
Figure 3.1. Frame structure of a hybrid system with users employing time-multiplexed and superimposed pilots.

framework and algorithm to partition the set of users $\mathcal{U}$ into two disjoint sets $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$ 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 ξ^{ul} and ξ^{dl} be non-negative weights for the UL and DL interference powers such that $\xi^{\mathrm{ul}} + \xi^{\mathrm{dl}} = 1$. Let $I^{\mathrm{RP-ul}}$ and $I^{\mathrm{RP-dl}}$ be user (ℓ, k)'s contribution to the interference power in the UL and DL when assigned to $\mathcal{U}_{\mathrm{RP}}$, and similarly, let $I^{\mathrm{SP-ul}}$ and $I^{\mathrm{SP-dl}}$ be the corresponding interference powers when user (ℓ, k) is assigned to $\mathcal{U}_{\mathrm{SP}}$. Then, the weighted interference power in the UL and DL, which is given as

$$T_{\ell k}^{\mathrm{RP}} \triangleq \xi^{\mathrm{ul}} I_{\ell k}^{\mathrm{RP-ul}} + \xi^{\mathrm{dl}} I_{\ell k}^{\mathrm{RP-dl}} \tag{3.11}$$

$$T_{\ell k}^{\mathrm{SP}} \triangleq \xi^{\mathrm{ul}} I_{\ell k}^{\mathrm{SP-ul}} + \xi^{\mathrm{dl}} I_{\ell k}^{\mathrm{SP-dl}} \,, \tag{3.12}$$

is used to construct the objective function. For a given partition $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$ 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

$$I\left(\mathcal{U}_{\mathrm{RP}}, \mathcal{U}_{\mathrm{SP}}\right) = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \left(T_{\ell k}^{\mathrm{RP}} \mathbf{1}_{\{(\ell, k) \in \mathcal{U}_{\mathrm{RP}}\}} + T_{\ell k}^{\mathrm{SP}} \mathbf{1}_{\{(\ell, k) \in \mathcal{U}_{\mathrm{SP}}\}} \right) \,. \tag{3.13}$$

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

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

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

$$\mathcal{U}_{\mathrm{RP}} \cap \mathcal{U}_{\mathrm{SP}} = \varnothing \tag{3.14}$$

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 $I(\mathcal{U}_{\mathrm{RP}}, \mathcal{U}_{\mathrm{SP}})$ 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 $r^{\mathrm{SP}} = r^{\mathrm{RP}}$ (c.f. Figs. 3.2-3.4). In Fig. 3.2, we see that no two cells that share the $\tau = r^{\mathrm{RP}} K$ 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 ρ^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 $C_{\mathrm{u}} = 35$ 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 $r^{\mathrm{SP}} = 7$.

It can be seen in Fig. 3.5 that the DL throughput of staggered pilots is very close to that of RP with $r^{\mathrm{RP}} = 7$ 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 $r^{\mathrm{RP}} = 7$. 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
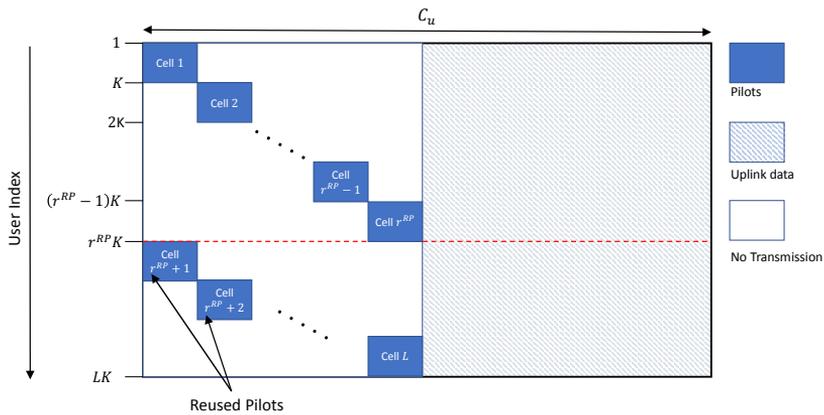
Figure 3.2. Pilot structure of RP. When the users in cell j transmits pilots, the users in the remaining $r^{\mathrm{RP}} - 1$ cells that share the set of τ orthogonal pilots effectively transmit zeros. Only users in the cells that reuse the same set of pilots as those of cell j transmit their pilots.
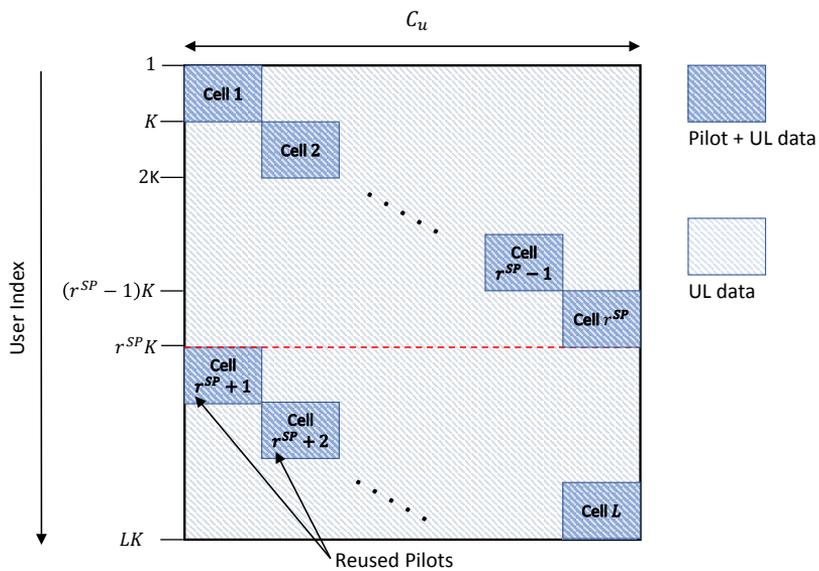
Figure 3.3. Pilot structure of superimposed. The pilot and data transmissions cover the entire UL slot.
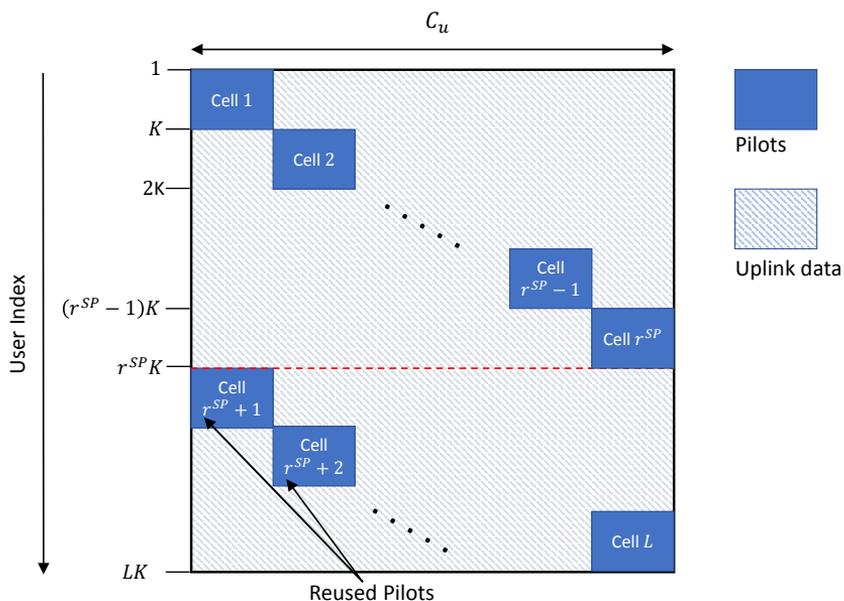
Figure 3.4. Pilot structure of staggered pilots. In contrast with RP, the pilot transmissions cover the entire UL slot, as with SP. However, in contrast with SP, the users in a cell do not transmit pilots and data simultaneously.

attractive alternative to pilot-reuse.

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 $\widehat{x}$ to maximize the likelihood by penalizing the MSE of $\widehat{h}$. 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
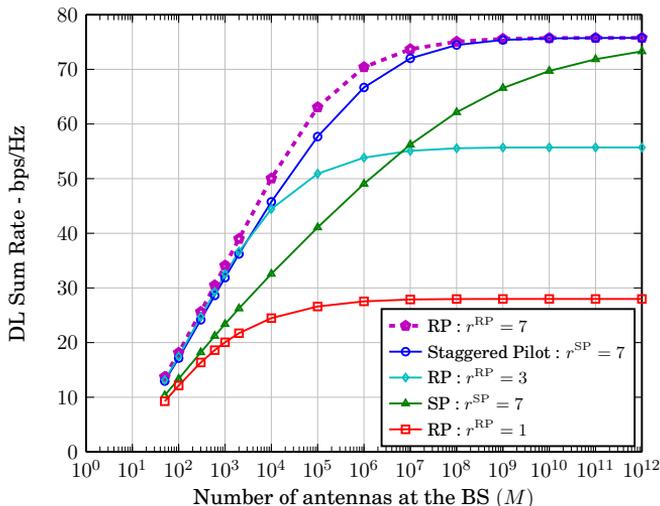
Figure 3.5. DL Sum Rate vs M with MR precoder. SP and Staggered pilots offer an asymptotic DL throughput equivalent to that of RP with $r^{\mathrm{RP}} = 7$, even though the UL overhead is only as much as that of RP with $r^{\mathrm{RP}} = 1$.
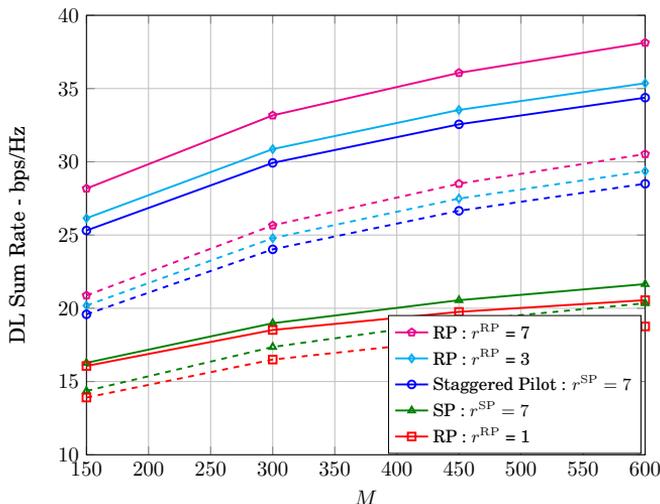
Figure 3.6. DL Sum Rate vs M. The solid lines are obtained using the ZF precoder and the dashed lines are obtained with the MF precoder.

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 $B_c = 200kHz$ and a coherence time of $T_c = 1$ms, which allows for a user communicating at 2 GHz carrier frequency with a 5μs 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 $B_cT_c = 200$ 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 $\tau_{\mathrm{s}} = (B_sT_s) / (B_cT_c) = 25000$ 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

(a) Time-frequency grid

(b) Frame structure of an individual coherence block
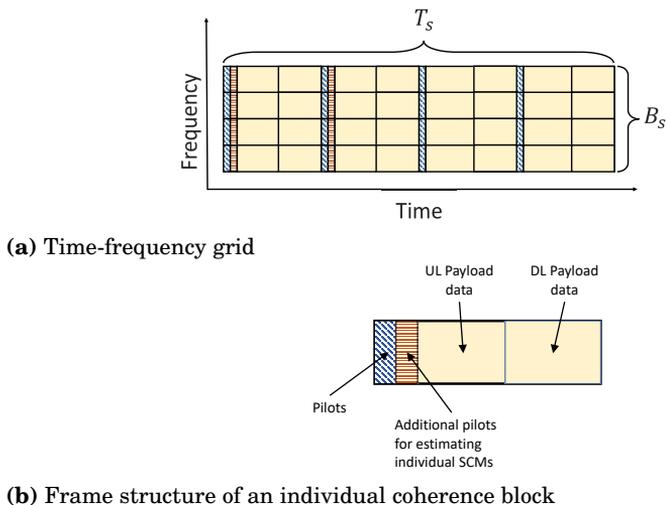
Figure 4.1. Coherence blocks across the time-frequency grid over which the channel statistics is fixed [38]. Some of the coherence blocks have additional pilots for estimating the SCM.

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.

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

$$\widehat{h}_1 = h_1 + h_2 + h_3 + w_1$$

$$\mathbb{E}\left\{\widehat{h}_1 \widehat{h}_1^H\right\} = \sum_{k=1}^{3} R_k + I$$
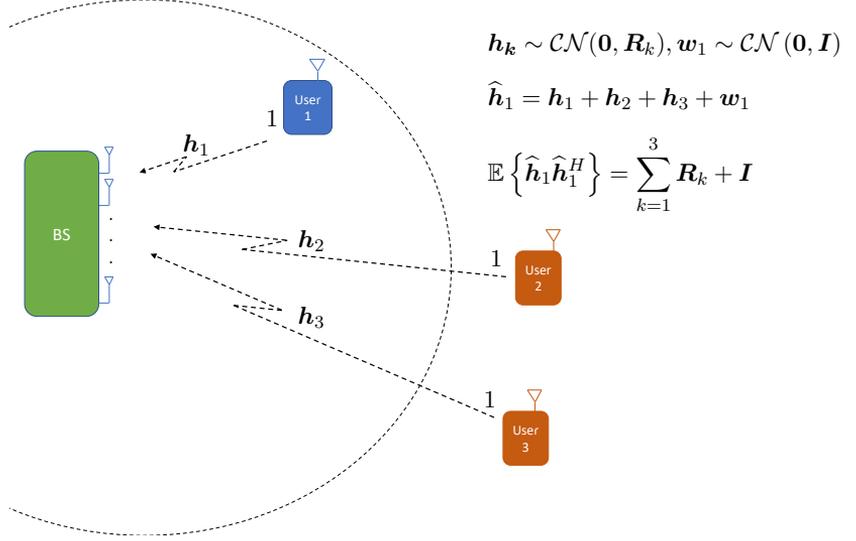
Figure 4.2. Illustration of the challenge in acquiring the individual user SCM in the presence of pilot contamination. Here, User 1 is the reference user and Users 2 and 3 are interfering users.

4.2 Challenges in Estimating the Spatial Covariance Matrices

We denote the SCM of user (ℓ, 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 $\widehat{h}_{j\ell k}$ be the LS estimate of $h_{j\ell k}$ at BS j. If $Q_{jm} \triangleq \mathbb{E}\left\{\widehat{h}_{jjm}\widehat{h}_{jjm}^H\right\}$ and $R_{j\ell k} \triangleq \mathbb{E}\left\{h_{j\ell k}h_{j\ell k}^H\right\}$ are the covariance matrices of the channel estimate and the channel vectors, respectively, the LMMSE estimate of the channel of user (ℓ, k) is given as

$$\widehat{h}_{j\ell k}^{\mathrm{LMMSE}} = R_{j\ell k}Q_{jm}^{-1}\widehat{h}_{j\ell k} . \tag{4.1}$$

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 $\widehat{h}_1$ 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

approach. 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 $\widehat{Q}_{jm}$ 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

$$\widehat{Q}_{jm}^{(\text{sample})} = \frac{1}{N_Q} \sum_{n=1}^{N_Q} \widehat{h}_{jjm}^{(n)} \left(\widehat{h}_{jjm}^{(n)} \right)^H \tag{4.2}$$

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

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

$$\widehat{Q}_{jm} = \gamma \widehat{Q}_{jm}^{(\text{sample})} + (1 - \gamma)\widehat{Q}_{jm}^{(\text{diag})} \tag{4.3}$$

where $\gamma \in [0, 1]$ is the shrinkage coefficient and $\widehat{Q}_{jm}^{(\text{diag})}$ is a diagonal matrix containing only the diagonal entries of $\widehat{Q}_{jm}^{(\text{sample})}$. Note that $\widehat{Q}_{jm}$ 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, $\widehat{R}_{j\ell k}$ 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 (ℓ, k) transmit pilots simultaneously in N_R coherence blocks, while the user (ℓ, k) maintains radio silence. The sample SCM obtained from these observations, denoted $\widehat{Q}_{j,-\ell k}$, contains the sum of the SCMs of all the interfering users. $\widehat{R}_{j\ell k}$ is then obtained by subtracting $\widehat{Q}_{j,-\ell k}$ from $\widehat{Q}_{jm}$ as

$$\widehat{R}_{j\ell k}^{(\text{diff})} = \widehat{Q}_{jm}^{(\text{sample})} - \widehat{Q}_{j,-\ell k}^{(\text{sample})} \ . \tag{4.4}$$

In both the aforementioned methods, $\widehat{R}_{j\ell k}^{(\text{sample})}$ and $\widehat{R}_{j\ell k}^{(\text{diff})}$ converges almost surely to $R_{j\ell k}$ as $N_R \to \infty$. When $N_R \leq M$, the estimated $\widehat{R}_{j\ell k}$ is regularized as in (4.3). However, note that this type of regularization for $\widehat{R}_{j\ell k}^{(\text{diff})}$ 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

$$Q_p[n] = P_{\text{ul}} \sum_{k \in \Omega_p[n]} R_k + I \ . \tag{4.5}$$

Stacking the vectorized versions of $Q_p[n]$ and R_k into the matrices B_R and $B_{Q[n]}$ defined as

$$B_R \triangleq [\text{vec}(R_1), \ldots, \text{vec}(R_{LK})] \tag{4.6}$$

$$B_{Q[n]} \triangleq [\text{vec}(Q_1[n]), \ldots, \text{vec}(Q_\tau[n])] \tag{4.7}$$

we have the relation

$$B_{Q[n]} = B_R \Pi[n] + \text{vec}(I)\mathbf{1}^T \ . \tag{4.8}$$

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., $\Pi \triangleq [\Pi_1, \ldots, \Pi_N]$, we can write

$$B_Q \triangleq \left[B_{Q[1]}, \ldots, B_{Q[N]} \right] = B_R \Pi + \text{vec}(I)\mathbf{1}^T \ . \tag{4.9}$$

If Π is chosen such that it has full row-rank, B_R can be recovered as

$$B_R = \left(B_Q - \text{vec}(I)\mathbf{1}^T \right) \Pi^\dagger \tag{4.10}$$

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.
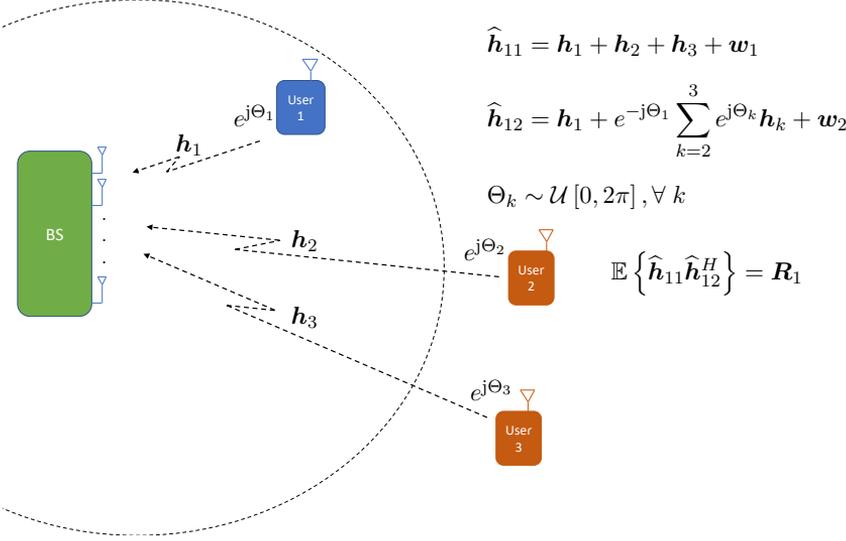
Figure 4.3. Illustration of contamination-free SCM estimation with randomly phase-shifted pilots. Each user transmits a pilot with a random phase-shift $e^{j\Theta_k}$. The realizations of Θ_1 are assumed to be known at the BS.

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 $\mathbb{E}\left\{e^{j\Theta_t}\right\} = 0$, $\forall\, t$ 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

$$\widehat{\boldsymbol{h}}_{jcu}^{(1)} = \boldsymbol{h}_{jcu} + \boldsymbol{\alpha}_{cu}^{(1)} + \boldsymbol{\epsilon}_{cu}^{(1)} + \boldsymbol{w}_j^{(1)} \tag{4.11}$$

$$\widehat{\boldsymbol{h}}_{jcu}^{(2)} = \boldsymbol{h}_{jcu} + e^{-j\Theta_t}\boldsymbol{\alpha}_{cu}^{(2)} + e^{-j\Theta_t}\boldsymbol{\epsilon}_{cu}^{(2)} + e^{-j\Theta_t}\boldsymbol{w}_j^{(2)} \tag{4.12}$$

where, for $p \in \{1, 2\}$, $\boldsymbol{\alpha}_{cu}^{(p)}$ corresponds to the interference from the subset of cells that are synchronized with cell c, $\boldsymbol{\epsilon}_{cu}^{(p)}$ corresponds to asynchronous UL or DL transmissions from the cells outside the subset, and $\boldsymbol{w}_j^{(p)}$ 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 Θ_t is independent of the channel vectors and

Method	Supports asynchronous networks	Pilot overhead for estimating $\boldsymbol{R}_{j\ell k}, \ \forall \, (\ell, k)$	Offers theoretical guarantees
[38]	-	$N_R L K$	+
[69]	-	0	+
[40]	-	0	-
Proposed method	+	$N K$	+

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

$$\boldsymbol{R}_{j\widehat{\boldsymbol{h}}_{jcu}^{(1)}, \widehat{\boldsymbol{h}}_{jcu}^{(2)}} = \mathbb{E}\left\{ \widehat{\boldsymbol{h}}_{jcu}^{(1)} \left(\widehat{\boldsymbol{h}}_{jcu}^{(2)} \right)^H \right\} = \boldsymbol{R}_{jcu} \ . \tag{4.13}$$

In the absence of the random phase-shift, the channel estimation errors in $\widehat{h}_{jcu}^{(1)}$ and $\widehat{h}_{jcu}^{(2)}$ are correlated, since $\boldsymbol{\alpha}_{cu}^{(1)}$ and $\boldsymbol{\alpha}_{cu}^{(2)}$ 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 Θ_1 are assumed to be known at the BS and therefore, the phase-shift on $\boldsymbol{h}_1$ can be compensated. It is then straightforward to see that the estimation errors in $\widehat{\boldsymbol{h}}_{11}$ and $\widehat{\boldsymbol{h}}_{12}$ 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 θ_{tn} of the random variable Θ_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 $\tau_s = 25000$
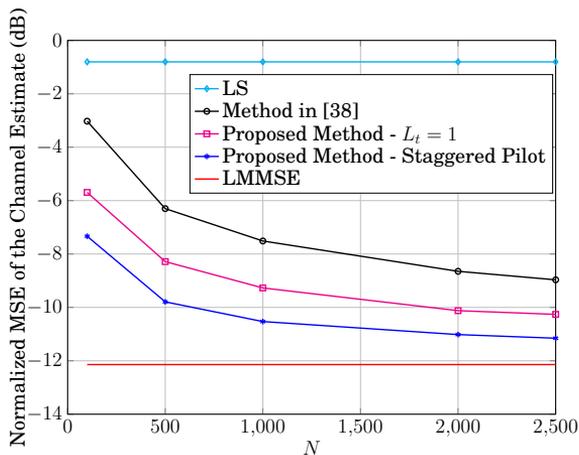
Figure 4.4. Normalized MSE of channel estimate of an average user vs. N.
Copyright 2018 IEEE.
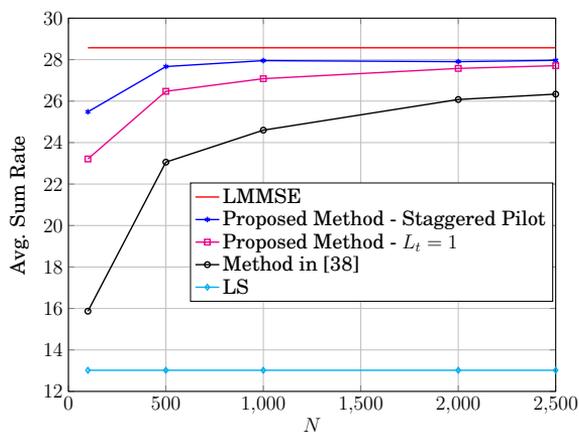
Figure 4.5. Achievable rate in the UL vs. N.
Copyright 2018 IEEE.

coherence 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 $20°$ 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 [83]

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

where λ 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

$$G_T = \frac{4\pi}{\lambda^2} A_T \ . \tag{5.2}$$

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 θ_p, ψ_p, α_p be the AoA, AoD and complex channel coefficient of path p, respectively. For a ULA with M antenna elements with spacing δ between the elements, the steering vector corresponding to angle θ is given as

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

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

$$\boldsymbol{H} = \sum_{p=0}^{P-1} \alpha_p \boldsymbol{a}_{\mathrm{R}}\left(\theta_p\right) \boldsymbol{a}_{\mathrm{T}}\left(\psi_p\right)^{H} = \bar{\boldsymbol{A}}_{\mathrm{R}}\bar{\boldsymbol{D}}\bar{\boldsymbol{A}}_{\mathrm{T}}^{H} \tag{5.4}$$

where $\bar{\boldsymbol{A}}_{\mathrm{R}} \triangleq \left[\boldsymbol{a}_{\mathrm{R}}\left(\theta_0\right), \ldots, \boldsymbol{a}_{\mathrm{R}}\left(\theta_{P-1}\right)\right]$, $\bar{\boldsymbol{A}}_{\mathrm{T}} \triangleq \left[\boldsymbol{a}_{\mathrm{T}}\left(\psi_0\right), \ldots, \boldsymbol{a}_{\mathrm{T}}\left(\psi_{P-1}\right)\right]$, and $\bar{\boldsymbol{D}} \triangleq \mathrm{diag}\left\{\alpha_0, \ldots, \alpha_{P-1}\right\}$.

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 $\bar{\boldsymbol{D}}$ 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 $\{\theta_q\}_{q=1}^{G_{\mathrm{R}}}$ and $\{\psi_q\}_{q=1}^{G_{\mathrm{T}}}$ be the set of G_{R} and G_{T} quantized AoAs and AoDs, respectively. Then, (5.4) can be written as

$$\boldsymbol{H} = \bar{\boldsymbol{A}}_{\mathrm{R}}\bar{\boldsymbol{D}}\bar{\boldsymbol{A}}_{\mathrm{T}}^{H} \approx \boldsymbol{A}_{\mathrm{R}}\boldsymbol{D}\boldsymbol{A}_{\mathrm{T}}^{H} \tag{5.5}$$

where the columns of $\boldsymbol{A}_{\mathrm{R}} \in \mathbb{C}^{M_R \times G_{\mathrm{R}}}$ and $\boldsymbol{A}_{\mathrm{T}} \in \mathbb{C}^{M_T \times G_{\mathrm{T}}}$ are steering vectors corresponding to the quantized AoAs and AoDs, and $\boldsymbol{D}$ is a sparse
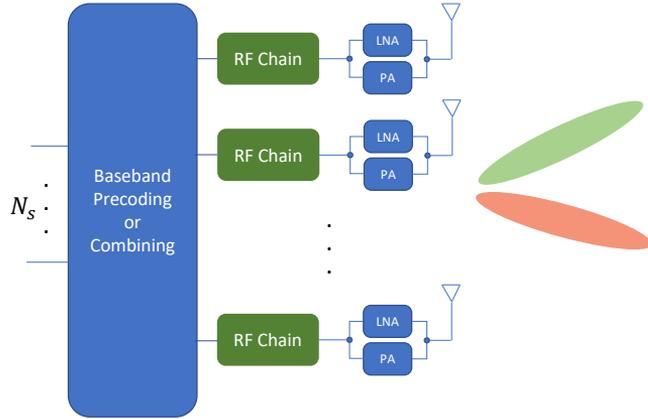
Figure 5.1. Architecture of fully-digital transceivers typically used in sub-6 GHz transceivers.
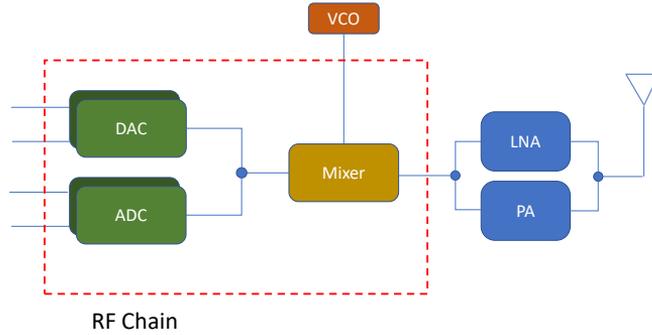
Figure 5.2. RF elements in a sub-6 GHz transceiver. For the rest of the chapter, the term 'RF chain' will be used to refer to the elements inside the dashed box.

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

$$H = V_R H_b V_T^H .$$ (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.
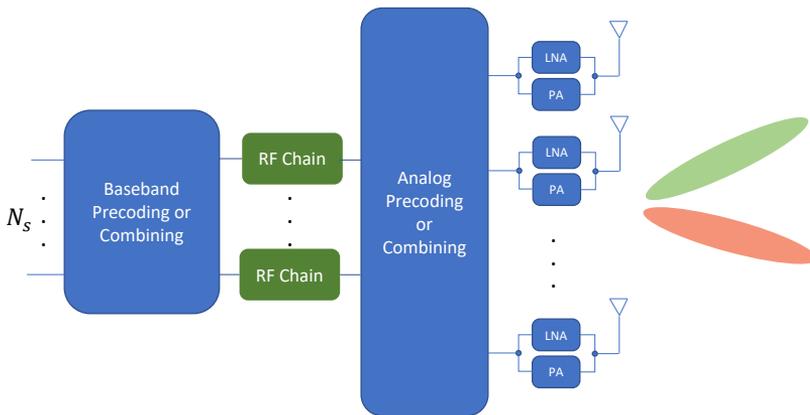
Figure 5.3. mmWave MIMO architecture based on hybrid precoding/combining.

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 posses unit-amplitude with quantized phase-shifts.
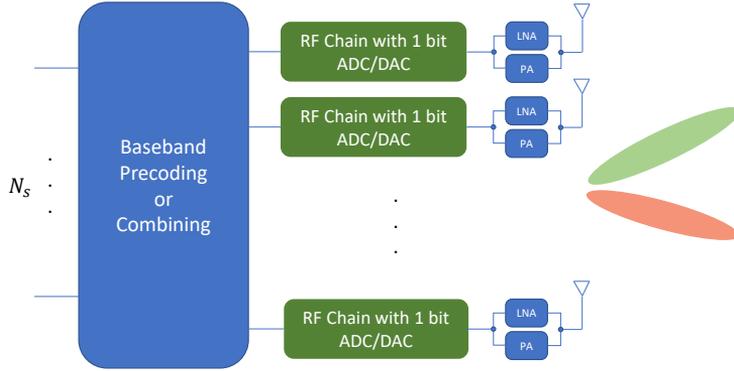
Figure 5.4. mmWave MIMO architecture based on low-resolution ADCs.

Switch networks, on the other hand, perform antenna selection and connect the M^{RF} RF chains to M^{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 $M^{\mathrm{RF}} \ll M$ 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 $\boldsymbol{F} \triangleq [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{N_T}]$ and $\boldsymbol{W} \triangleq [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{N_R}]$, the received observations during training is given as

$$\boldsymbol{Y} = \sqrt{P_T} \boldsymbol{W}^H \boldsymbol{H} \boldsymbol{F} + \boldsymbol{Q} \tag{5.7}$$

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

$$\boldsymbol{y} \triangleq \mathrm{vec}(\boldsymbol{Y}) = \sqrt{P_T} \left(\boldsymbol{F}^T \bar{\boldsymbol{A}}_{\mathrm{T}}^* \circ \boldsymbol{W}^H \bar{\boldsymbol{A}}_{\mathrm{R}} \right) \mathrm{diag} \left\{ \bar{\boldsymbol{D}} \right\} + \boldsymbol{q}$$

$$\approx \sqrt{P_T} \left(\boldsymbol{F}^T \boldsymbol{A}_{\mathrm{T}}^* \otimes \boldsymbol{W}^H \boldsymbol{A}_{\mathrm{R}} \right) \boldsymbol{d} + \boldsymbol{q} \tag{5.8}$$

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

With the measurement matrix defined as $\boldsymbol{\Psi} \triangleq \sqrt{P_T} \left(\boldsymbol{F}^T \boldsymbol{A}_{\mathrm{T}}^* \otimes \boldsymbol{W}^H \boldsymbol{A}_{\mathrm{R}} \right)$, an estimate of the sparse vector $\widehat{\boldsymbol{d}}$ can be recovered from $\boldsymbol{y}$ as

$$\widehat{\boldsymbol{d}} = \arg \min_{\boldsymbol{d}} \|\boldsymbol{d}\|_0$$

$$\text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{\Psi} \boldsymbol{d}\| < \epsilon \tag{5.9}$$

where ϵ is a design parameter. The non-convex optimization problem in (5.9) can be solved by relaxing the $\| \cdot \|_0$ operator with the ℓ_1 norm or using a sparse recovery algorithm such as orthogonal matching pursuit (OMP) or sparse Bayesian learning (SBL) [99]. The channel estimate $\widehat{H}$ is obtained from $\widehat{d}$ as

$$\widehat{H} = A_{\mathrm{R}} \widehat{D} A_{\mathrm{T}}^{H} \ . \tag{5.10}$$

The left and right singular vectors of $\widehat{H}$ 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 $\widehat{d}$ 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

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 y 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 [105] and using a Newton-type method to obtain the off-grid components [98]. 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, $\widehat{H}$ can be recovered from Y in (5.7) by solving the following optimization problem

$$\widehat{H} = \arg\min_{H} \quad \text{rank}\,(H)$$

$$\text{subject to} \quad \|Y - \sqrt{P_T}W^H H F\| < \epsilon . \tag{5.11}$$

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 $\boldsymbol{F}$ and $\boldsymbol{W}$ 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 $\boldsymbol{F}$ and $\boldsymbol{W}$ 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 $\boldsymbol{F}$ and $\boldsymbol{W}$ 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 $\boldsymbol{F} = \boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}$ and $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{RF}}\boldsymbol{W}_{\mathrm{BB}}$. When $\boldsymbol{H}$ is available at the receiver, the matrices $\boldsymbol{F}_{\mathrm{RF}}$, $\boldsymbol{F}_{\mathrm{BB}}$, $\boldsymbol{W}_{\mathrm{RF}}$, and $\boldsymbol{W}_{\mathrm{BB}}$ can be designed to maximize the mutual information which, when Gaussian symbols are transmitted over the mmWave channel, is given by [17]

$$\mathcal{I}\left(\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{F}_{\mathrm{BB}}, \boldsymbol{W}_{\mathrm{RF}}, \boldsymbol{W}_{\mathrm{BB}}\right) = \log_2 \left| \boldsymbol{I}_{N_s} + \frac{P_T}{N_s} \boldsymbol{R}_n^{-1} \boldsymbol{W}_{\mathrm{BB}}^H \boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{H} \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \right.$$

$$\left. \times \boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{F}_{\mathrm{RF}}^H \boldsymbol{H}^H \boldsymbol{W}_{\mathrm{RF}} \boldsymbol{W}_{\mathrm{BB}} \right| \qquad (5.12)$$

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

$$\mathcal{I}\left(\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{F}_{\mathrm{BB}}\right) = \log_2 \left| \boldsymbol{I}_{N_s} + \frac{P_T}{N_s} \boldsymbol{R}_n^{-1} \boldsymbol{H} \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \boldsymbol{F}_{\mathrm{BB}}^H \boldsymbol{F}_{\mathrm{RF}}^H \boldsymbol{H}^H \right|$$

$$\approx \log_2 \left| \boldsymbol{I}_{N_s} + \frac{P_T}{N_s} \boldsymbol{\Sigma}_s^2 \right| - \left(N_s - \|\boldsymbol{V}_s^H \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}}\|_F^2 \right) \qquad (5.13)$$

where $\boldsymbol{\Sigma}_s$ is a diagonal matrix containing the N_s largest singular values and $\boldsymbol{V}_s$ contains the corresponding right singular vectors of $\boldsymbol{H}$. Setting $\boldsymbol{F}_{\mathrm{opt}} \triangleq \boldsymbol{V}_s$, (5.13) can be approximately maximized by solving the following

optimization problem [100]

$$\left(\boldsymbol{F}_{\mathrm{RF}}^{\mathrm{opt}}, \boldsymbol{F}_{\mathrm{BB}}^{\mathrm{opt}} \right) = \arg \min_{\boldsymbol{F}_{\mathrm{RF}}, \boldsymbol{F}_{\mathrm{BB}}} \| \boldsymbol{F}_{\mathrm{opt}} - \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \|_F$$

$$\text{subject to } \boldsymbol{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}$$

$$\| \boldsymbol{F}_{\mathrm{RF}} \boldsymbol{F}_{\mathrm{BB}} \|_F^2 = N_s \qquad (5.14)$$

where $\mathcal{F}_{\mathrm{RF}}$ is the set of feasible RF precoders. For analog phase-shifters, $\mathcal{F}_{\mathrm{RF}}$ is the set of $M_T \times M_T^{\mathrm{RF}}$ matrices with elements that are unit-amplitude and have quantized phase-shifts. Similarly, for switches, $\mathcal{F}_{\mathrm{RF}}$ is such that each column of $\boldsymbol{F}_{\mathrm{RF}}$ is a binary vector with a single one and zeros elsewhere [101].

With a constraint set $\mathcal{W}_{\mathrm{RF}}$ on $\boldsymbol{W}_{\mathrm{RF}}$, $\boldsymbol{W}$ can be designed by minimizing the MSE of the data s at the output of the combiner.

$$\left(\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{opt}}, \boldsymbol{W}_{\mathrm{BB}}^{\mathrm{opt}} \right) = \arg \min_{\boldsymbol{W}_{\mathrm{RF}}, \boldsymbol{W}_{\mathrm{BB}}} \mathbb{E} \left\{ \left\| s - \boldsymbol{W}_{\mathrm{BB}}^H \boldsymbol{W}_{\mathrm{RF}}^H \boldsymbol{y} \right\|_2^2 \right\}$$

$$\text{subject to } \boldsymbol{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}} . \qquad (5.15)$$

Equation (5.15) can be shown to be equivalent to minimizing the weighted Frobenius norm of the difference between the MMSE combiner $\boldsymbol{W}_{\mathrm{MMSE}}$ and $\boldsymbol{W}_{\mathrm{RF}} \boldsymbol{W}_{\mathrm{BB}}$ [100]. The latter problem is given as

$$\left(\boldsymbol{W}_{\mathrm{RF}}^{\mathrm{opt}}, \boldsymbol{W}_{\mathrm{BB}}^{\mathrm{opt}} \right) = \arg \min_{\boldsymbol{W}_{\mathrm{RF}}, \boldsymbol{W}_{\mathrm{BB}}} \left\| \mathbb{E} \left[\boldsymbol{y} \boldsymbol{y}^H \right]^{\frac{1}{2}} \left(\boldsymbol{W}_{\mathrm{MMSE}} - \boldsymbol{W}_{\mathrm{RF}} \boldsymbol{W}_{\mathrm{BB}} \right) \right\|_F$$

$$\text{subject to } \boldsymbol{W}_{\mathrm{RF}} \in \mathcal{W}_{\mathrm{RF}} . \qquad (5.16)$$

With analog phase-shifters capable of continuous phase-shifts, $\boldsymbol{F}_{\mathrm{opt}}$ and $\boldsymbol{W}_{\mathrm{MMSE}}$ can be exactly realized with the hybrid architecture if $M^{\mathrm{RF}} \geq 2N_s$ [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, [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 [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]

$$\frac{1}{M} \boldsymbol{a}^H \left(\theta_1\right) \boldsymbol{a} \left(\theta_2\right) \approx \delta \left(\theta_1, \theta_2\right) \ . \tag{5.17}$$

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 $\widehat{\boldsymbol{H}}$ 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 $M_T^{\mathrm{RF}} = M_R^{\mathrm{RF}} = 4$ RF chains. At the receiver $M_R^{\mathrm{RF-T}} = 1$ RF chain out of the M_R^{RF} 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 $\mathcal{CN}(0, \sigma_\theta)$ and $\mathcal{CN}(0, \sigma_\psi)$, 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.
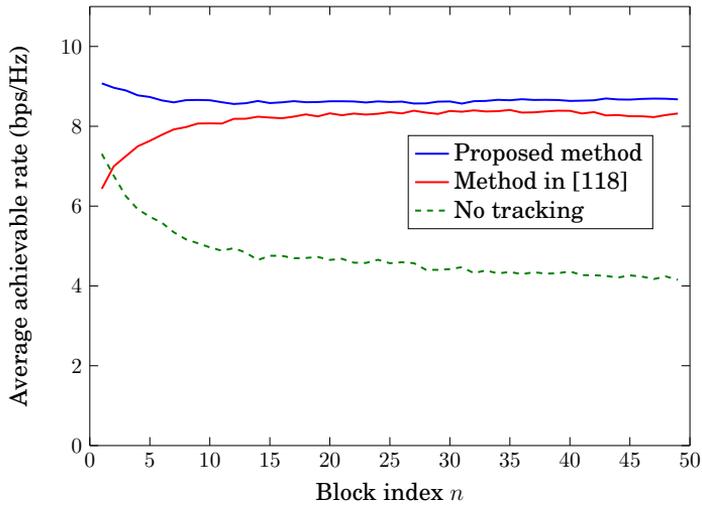
Figure 5.5. Plot of the average achievable rate vs block index at SNR = 0 dB, $\sigma_\theta = 2°$ and $\sigma_\psi = 0$.
Copyright 2018 IEEE.
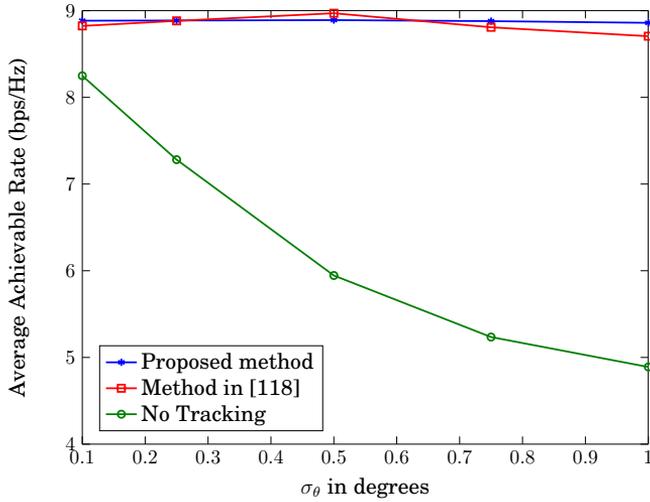
Figure 5.6. Plot of the average achievable rate vs σ_θ at SNR = 0 dB and $\sigma_\psi = 0$ at the $n = 50$th block.
Copyright 2018 IEEE.

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

[1] V. Va, T. Shimizu, G. Bansal, and R. W. Heath Jr., "Millimeter wave vehicular communications: A survey," *Foundations and Trends in Networking*, vol. 10, no. 1, pp. 1–113, 2016.

[2] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016.

[3] C. Bockelmann *et al.*, "Massive machine-type communications in 5G: physical and mac-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[4] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[5] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 GHz band," *IEEE Std 802.11ad-2012*, pp. 1–628, Dec. 2012.

[6] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[7] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[8] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[9] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct. 2006, pp. 359–363.

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

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

[12] E. Björnson, J. Hoydis, and L. Sanguinetti, "Pilot contamination is not a fundamental asymptotic limitation in massive MIMO," in *IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[13] ——, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018. [Online]. Available: https://arxiv.org/abs/1705.00538

[14] J. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[15] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[16] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[17] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[18] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[20] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 60–67, Oct. 2004.

[21] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[22] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[23] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2178–2181, Dec. 2015.

[24] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.

[25] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing — the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[26] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO — 5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, Secondquarter 2016.

[27] H. Q. Ngo and E. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Kyoto*, Mar. 2012, pp. 3249–3252.

[28] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[29] J. Vinogradova, E. Björnson, and E. G. Larsson, "On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 3421–3425.

[30] D. Neumann, M. Joham, and W. Utschick, "Suppression of pilot-contamination in massive mimo systems," in *Proc. IEEE 15th Int. Workshop Signal Process. Advances Wireless Commun, (SPAWC)*, Jun. 2014, pp. 11–15.

[31] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 540–553, Jan. 2018.

[32] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Jun. 2014.

[33] K. Mawatwal, D. Sen, and R. Roy, "A semi-blind channel estimation algorithm for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 70–73, Feb. 2017.

[34] A. Zaib, M. Masood, A. Ali, W. Xu, and T. Y. Al-Naffouri, "Distributed channel estimation and pilot contamination analysis for massive MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4607–4621, Nov. 2016.

[35] K. Wang and Z. Ding, "FEC code anchored robust design of massive MIMO receivers," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8223–8235, Dec. 2016.

[36] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *the 5th Meeting of COST 273*, Lisbon, Portugal, Sep. 2002.

[37] ——, "Validity of spatial covariance matrices over time and frequency," in *Proc. Global Telecommun. Conf.*, Taipei, Taiwan, Nov. 2002, pp. 851–855.

[38] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2016, pp. 974–978.

[39] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, Jun. 2016.

[40] S. Haghighatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8316–8332, Dec. 2017. [Online]. Available: http://arxiv.org/abs/1702.07207

[41] Y. Li, Y.-H. Nam, B. L. Ng, and J. Zhang, "A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Anaheim, CA, Dec. 2012, pp. 4500–4504.

[42] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[43] I. Atzeni, J. Arnau, and M. Debbah, "Fractional pilot reuse in massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, London, UK, Jun. 2015, pp. 1030–1035.

[44] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 2010, pp. 1–5.

[45] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[46] W. A. W. M. Mahyiddin, P. A. Martin, and P. J. Smith, "Performance of synchronized and unsynchronized pilots in finite massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6763–6776, Dec. 2015.

[47] D. Kong, D. Qu, K. Luo, and T. Jiang, "Channel estimation under staggered frame structure for massive MIMO system," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1469–1479, Feb. 2016.

[48] L. S. Muppirisetty, T. Charalambous, J. Karout, G. Fodor, and H. Wymeersch, "Location-aided pilot contamination avoidance for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2662–2674, Apr. 2018.

[49] N. Akbar, S. Yan, N. Yang, and J. Yuan, "Mitigating pilot contamination through location-aware pilot assignment in massive MIMO networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Washington, D.C., Dec. 2016, pp. 1–6.

[50] X. Luo and X. Zhang, "Flexible pilot contamination mitigation with doppler PSD alignment," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1449–1453, Oct. 2016.

[51] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.

[52] X. Luo, X. Zhang, P. Cai, and H. Qian, "Aligning power in multiple domains for pilot decontamination in massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7919–7935, Dec. 2017.

[53] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symposium Inform. Theory Proc.*, Jul. 2012, pp. 1137–1141.

[54] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[55] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917–2932, Jun. 2017.

[56] K. Upadhya, S. A. Vorobyov, and M. Vehkaperä, "Downlink performance of superimposed pilots in massive MIMO systems," *Submitted to IEEE Trans. Wireless Commun.* [Online]. Available: https://arxiv.org/pdf/1606.04476.pdf

[57] E. Alameda-Hernandez, D. C. McLernon, A. G. Orozco-Lugo, M. M. Lara, and M. Ghogho, "Frame/training sequence synchronization and DC-offset removal for (data-dependent) superimposed training based channel estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2557–2569, Jun. 2007.

[58] T. Cui and C. Tellambura, "Pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM), St. Louis*, vol. 4, Dec. 2005, pp. 5 pp.–2233.

[59] M. Coldrey and P. Bohlin, "Training-based MIMO systems – part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[60] A. K. Jagannatham and B. D. Rao, "Superimposed pilots vs. conventional pilots for channel estimation," in *Proc. 40th Asilomar Conf. on Signals, Syst. and Comput.*, Oct. 2006, pp. 767–771.

[61] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, Mar. 2016.

[62] D. Verenzuela, E. Björnson, and L. Sanguinetti, "Spectral and energy efficiency of superimposed pilots in uplink massive MIMO," *arXiv Preprint*, 2017. [Online]. Available: http://arxiv.org/abs/1709.07722

[63] D. Verenzuela, E. Björnson, and L. Sanguinetti, "Joint UL and DL spectral efficiency optimization of superimposed pilots in massive MIMO," in *IEEE Globecom Workshops (GC Wkshps)*, Singapore, Dec. 2017, pp. 1–7.

[64] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Time-multiplexed / superimposed pilot selection for massive MIMO pilot decontamination," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LO, Mar. 2017, pp. 3459 – 3463.

[65] T. Liu, J. Tong, Q. Guo, J. Xi, Y. Yu, and Z. Xiao, "Energy efficiency of uplink massive MIMO systems with successive interference cancellation," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 668–671, Mar. 2017.

[66] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365 – 411, 2004.

[67] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, Oct. 2014. [Online]. Available: https://arxiv.org/abs/1401.5703

[68] M. Shaghaghi and S. A. Vorobyov, "Subspace leakage analysis and improved DOA estimation with small sample size," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3251–3265, Jun. 2015.

[69] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *arXiv preprint arXiv:1705.02895*, 2017.

[70] D. Neumann, K. Shibli, M. Joham, and W. Utschick, "Joint covariance matrix estimation and pilot allocation in massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[71] A. Decurninge and M. Guillaud, "Covariance estimation with projected data: Applications to CSI covariance acquisition and tracking," in *Proc. 25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, Aug. 2017, pp. 628–632.

[72] K. Upadhya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, no. 4, pp. 546–550, Apr. 2018. [Online]. Available: https://arxiv.org/abs/1802.01513

[73] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[74] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[75] R. W. Heath Jr., *Introduction to Wireless Digital Communication: A Signal Processing Perspective*, 1st ed. Pearson Education, 2017.

[76] S. Salous *et al.*, "Millimeter-wave propagation: Characterization and modeling toward fifth-generation systems." *IEEE Antennas Propag. Mag.*, vol. 58, no. 6, pp. 115–127, Dec. 2016.

[77] D. Solomitckii *et al.*, "Characterizing the impact of diffuse scattering in urban millimeter-wave deployments," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 432–435, Aug. 2016.

[78] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks — with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.

[79] S. Rajagopal, S. Abu-Surra, and M. Malmirchegini, "Channel feasibility for outdoor non-line-of-sight mmWave mobile communication," in *Proc. IEEE Veh. Tech. Conf. (VTC Fall)*, Sep. 2012, pp. 1–6.

[80] T. S. Rappaport, R. W. Heath Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Prentice Hall, 2014.

[81] F. Fuschini *et al.*, "Item level characterization of mm-wave indoor propagation," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, p. 4, Dec. 2016.

[82] G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter-wave human blockage at 73 ghz with a simple double knife-edge diffraction model and extension for directional antennas," in *Proc. IEEE 84th Veh. Tech. Conf.*, Sep. 2016, pp. 1–6.

[83] H. T. Friis, "A note on a simple transmission formula," *Proc. IRE*, vol. 34, no. 5, pp. 254–256, May 1946.

[84] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[85] H. S. Lee and C. G. Sodini, "Analog-to-digital converters: Digitizing the analog world," *Proceedings of the IEEE*, vol. 96, no. 2, pp. 323–334, Feb. 2008.

[86] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. Inform. Theory Applications Workshop (ITA)*, Feb. 2015, pp. 191–198.

[87] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[88] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.

[89] J. Wang, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[90] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[91] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[92] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[93] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Single-sided adaptive estimation of multi-path millimeter wave channels," in *Proc. IEEE 15th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Jun. 2014, pp. 125–129.

[94] M. E. Eltayeb, A. Alkhateeb, R. W. Heath, and T. Y. Al-Naffouri, "Opportunistic beam training with hybrid analog/digital codebooks for mmWave systems," in *Proc. IEEE Global Conf. Signal Inform. Process. (GlobalSIP)*, Dec. 2015, pp. 315–319.

[95] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmWave communication system," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2011, pp. 1–6.

[96] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, "Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 601–616, Feb. 2017.

[97] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1859–1872, Apr. 2017.

[98] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Proc. Inform. Theory Appl. Workshop*, Feb. 2012, pp. 234–239.

[99] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[100] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[101] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[102] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, South Brisbane, Australia, Apr. 2015, pp. 2909–2913.

[103] A. Ali and R. W. Heath, "Compressed beam-selection in millimeterwave systems with out-of-band partial support information," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 3499–3503.

[104] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

[105] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 273–277.

[106] M. Kokshoorn, H. Chen, Y. Li, and B. Vucetic, "Beam-on-graph: Simultaneous channel estimation for mmWave MIMO systems with multiple users," *IEEE Trans. Commun.*, to be published.

[107] J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, to be published.

[108] K. Venugopal, N. González-Prelcic, and R. W. Heath, "Optimality of frequency flat precoding in frequency selective millimeter wave channels," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 330–333, Jun. 2017.

[109] J. Deng, O. Tirkkonen, and C. Studer, "MmWave channel estimation via atomic norm minimization for multi-user hybrid precoding," in *Proc. Wireless Commun. Networking Conf.*, to be published.

[110] S. Park, J. Park, A. Yazdan, and R. W. Heath, "Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 14, pp. 3818–3832, Jul. 2017.

[111] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[112] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 303–318, Jan. 2017.

[113] S. Park and R. W. Heath, "Spatial channel covariance estimation for mmWave hybrid MIMO architecture," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 1424–1428.

[114] S. Park and R. W. Heath, Jr, "Spatial channel covariance estimation for the hybrid MIMO architecture: A compressive sensing based approach," *unpublished*. [Online]. Available: https://arxiv.org/abs/1711.04207

[115] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[116] A. F. Molisch *et al.*, "Hybrid beamforming for massive mimo: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[117] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[118] N. Garcia, H. Wymeersch, and D. T. M. Slock, "Optimal robust precoders for tracking the AoD and AoA of a mm-Wave path," *ArXiv*, 2017. [Online]. Available: http://arxiv.org/abs/1703.10978

[119] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[120] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[121] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," in *Proc. IEEE Global Conf. Signal Inform. Process. (GlobalSIP)*, Dec. 2016, pp. 743–747.

[122] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[123] S. G. Larew and D. J. Love, "Adaptive beam tracking with the unscented kalman filter for millimeter wave communication," *unpublished*. [Online]. Available: https://arxiv.org/abs/1804.08640

[124] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5689–5696, Jul. 2017.

[125] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," vol. 10, no. 3, pp. 514–527, Apr. 2016.

[126] K. Upadhya, R. W. Heath Jr., and S. A. Vorobyov, "Tracking abruptly changing channels in mmWave systems using overlaid data and training," in *Proc. IEEE 7th Int. Workshop Computational Advances in Multi-Sensor Adaptive Processing*, Curacao, Dutch Antilles, Dec. 2017.

[127] K. Upadhya, S. A. Vorobyov, and R. W. Heath Jr., "Low-overhead receiver-side channel tracking for mmwave mimo," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Alberta, Apr. 2018, p. 3859–3863.

[128] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, Jul. 2014.

[129] J. Rodríguez-Fernández, N. Gonzáiez-Prelcic, and R. W. Heath, "Frequency-domain wideband channel estimation and tracking for hybrid MIMO systems," in *Proc. 51st Asilomar Conf. Signals, Syst., and Comput.*, Oct. 2017, pp. 1829–1833.

Bibliography

Errata

Publication IV

Equation (52) should contain an '=' sign after the term $\frac{1}{\tau}\boldsymbol{Y}_j\boldsymbol{b}_{j,m}^{\mathrm{TP}}$.

Publication VII

$\boldsymbol{A}_{\mathrm{BS}}$ and $\boldsymbol{A}_{\mathrm{MS}}$ should be interchanged in Equation (4).

Publication VII

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

Publication VIII

Fig. 1 is not plotted against ℓ. It is plotted against the number of coherence blocks over which the channel is tracked.

Errata

Publication I

K. Upadhya, S. A. Vorobyov, M. Vehkapera. Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO. *in Proc. IEEE Int. Conf. Acoustics Speech Signal Process.*, Shanghai, China, pp. 3366-3370, Mar. 2016.

SUPERIMPOSED PILOTS: AN ALTERNATIVE PILOT STRUCTURE TO MITIGATE PILOT CONTAMINATION IN MASSIVE MIMO

Karthik Upadhya[⋆] *Sergiy A. Vorobyov*[⋆] *Mikko Vehkapera*[†]

[⋆] Dept. of Signal Processing and Acoustics , Aalto University, Espoo, Finland
[†]EEE Department, University of Sheffield, Sheffield, United Kingdom
Emails:karthik.upadhya@aalto.fi, svor@ieee.org, m.vehkapera@sheffield.ac.uk

ABSTRACT

Superimposed pilots are proposed as an alternative to time-multiplexed pilot and data symbols for mitigating pilot contamination in massive multiple-input multiple-output systems. Provided that the uplink duration is larger than the total number of users in the system, superimposed pilots enable each user to be assigned a unique pilot sequence, thereby allowing for a significant reduction in pilot contamination. Channel estimation performance in the uplink is further improved using an iterative data-aided algorithm. Based on approximate expressions for the uplink signal-to-interference-plus-noise ratio, it is shown that superimposed pilots provide a better performance when compared with methods that use time-multiplexed data and pilots. Numerical simulations are used to validate the approximations and the improved performance of the proposed method.

Index Terms— Massive MIMO, pilot decontamination, superimposed pilots

1. INTRODUCTION

Existing methods to mitigate pilot contamination for massive multiple-input multiple-output (MIMO) [1] systems are designed for the case where the pilots are time-multiplexed with the data, henceforth referred to as conventional time-multiplexed pilots. To mitigate interference, methods that employ conventional time-multiplexed pilots utilize the array gain offered by the large antenna array, asymptotic orthogonality between channel vectors and coordination between base stations (BS) [2–4].

On the other hand, superimposed pilots have been extensively studied for channel estimation in MIMO systems [5–8] in the context of minimizing the loss in data-transmission rate by foregoing the need for separate time slot for pilot-symbol transmission. However, in multi-cell massive MIMO systems, superimposed pilots allow each user in the system to be assigned a unique pilot sequence enabling the BS to estimate the channel vectors of both the desired and interfering users. Most recently, superimposed pilots have been used in

the context of multi-cell multiuser MIMO systems [9] without realizing that it provides a superior solution to the pilot decontamination problem in massive MIMO systems.

In this paper, we show that superimposed pilots are superior to conventional time multiplexed pilots in the context of pilot contamination in massive MIMO systems. We propose a novel low-complexity iterative matched-filter based channel estimation scheme for superimposed pilots and demonstrate its effectiveness by deriving a closed-form approximation for the uplink (UL) signal-to-interference-plus-noise ratio (SINR) of the system. Although the use of superimposed pilots requires some coordination between the BSs in assigning pilot sequences to the users and estimating their path-loss coefficients, these are minor impediments compared to the performance improvements provided by the proposed scheme.

2. SYSTEM MODEL

For channel estimation, the massive MIMO uplink with L cells and K users per cell is considered. Each cell has a BS with $M \gg K$ antennas. Assuming that the channel is constant for C_u symbols in the uplink, the matrix of received values $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$ at the jth BS can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \mathbf{x}_{\ell,k}^T + \mathbf{W}_j \qquad (1)$$

where $\mathbf{h}_{j,\ell,k} \in \mathbb{C}^{M \times 1}$ is the channel vector from the kth user in the ℓth cell to the jth BS, $\mathbf{x}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$ is the sequence of symbols transmitted by the kth user in the ℓth cell, $(\cdot)^T$ denotes the transpose, $\mathbf{W}_j \in \mathbb{C}^{M \times C_u}$ is the additive white Gaussian noise at the jth BS with each column distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ and mutually independent of the other columns. The nth element of $\mathbf{h}_{j,\ell,k}$ is given as $[\mathbf{h}_{j,\ell,k}]_n = g_{j,\ell,k,n} \sqrt{\beta_{j,\ell,k}}$, where $g_{j,\ell,k,n} \sim \mathcal{CN}(0,1)$ and $\beta_{j,\ell,k}$ are the fast-fading and large-scale fading components, respectively. The elements $g_{j,\ell,k,n}$ are mutually independent of each other for all n and $\beta_{j,\ell,k}$ is assumed to be known at the BS. Therefore, the vectors $\mathbf{h}_{j,\ell,k}$ are complex normal

distributed and are assumed to be mutually independent of each other $\forall j, \ell, k$.

3. CONVENTIONAL TIME-MULTIPLEXED PILOTS AND THE PILOT CONTAMINATION PROBLEM

With conventional time-multiplexed pilots, each user in a cell transmits a $\tau \geq K$ length orthogonal pilot sequence for channel estimation and the same pilot sequences are transmitted by the users in the $L-1$ interfering cells. The received signal at the jth BS during pilot training $\mathbf{Y}_j^{(p)} \in \mathbb{C}^{M \times \tau}$ can be written as

$$\mathbf{Y}_j^{(p)} = \sum_{\ell=0}^{L-1} \mathbf{H}_{j,\ell} \boldsymbol{\Phi}^T + \mathbf{W}_j \quad (2)$$

where $\mathbf{H}_{j,\ell} \triangleq [\mathbf{h}_{j,\ell,0}, \ldots, \mathbf{h}_{j,\ell,K-1}]$, $\boldsymbol{\Phi} \in \mathbb{C}^{\tau \times K}$ is the matrix of pilot sequences with mutually orthogonal columns such that $\boldsymbol{\Phi}^T \boldsymbol{\Phi}^* = \tau \mathbf{I}_K$, $(\cdot)^*$ denotes conjugation, and $\mathbf{I}_N$ represents the $N \times N$ identity matrix. The least-squares (LS) estimate of the channel of the mth user in the jth BS can be written as

$$\widehat{\mathbf{h}}_{j,j,m}^{\text{CP}} \triangleq \frac{1}{\tau} \mathbf{Y}_j^{(p)} \boldsymbol{\Phi}_m^* = \mathbf{h}_{j,j,m} + \sum_{\substack{\ell=0 \\ \ell \neq j}}^{L-1} \mathbf{h}_{j,\ell,m} + \frac{1}{\tau} \mathbf{W}_j \boldsymbol{\Phi}_m^* \quad (3)$$

where $\boldsymbol{\Phi}_m$ is the mth column of $\boldsymbol{\Phi}$ and it is the pilot sequence transmitted by the mth user in each cell. It can be seen that the estimates of the channel vectors of the users in the jth cell are contaminated by the channel vectors of the users in the remaining $L-1$ cells. The MSE of the channel estimate of the mth user at the jth BS can be shown to be

$$\text{MSE}_{j,j,m}^{\text{CP}} \triangleq \frac{1}{M} \mathbb{E}\left[\|\mathbf{h}_{j,j,m} - \widehat{\mathbf{h}}_{j,j,m}\|^2\right] = \sum_{\substack{\ell=0 \\ \ell \neq j}}^{L-1} \beta_{j,\ell,m} + \frac{\sigma^2}{\tau}. \quad (4)$$

4. SUPERIMPOSED PILOTS

When using superimposed pilots, the pilot symbols are transmitted at a reduced power alongside the data symbols for the entire duration of the uplink data slot C_u. If the total number of users in the system is less than the uplink pilot duration, i.e., $KL \leq C_u$, each user can be assigned a unique orthogonal pilot sequence $\mathbf{p}_{j,m} \in \mathbb{C}^{C_u \times 1}$ taken from the columns of an orthogonal matrix $\mathbf{P} \in \mathbb{C}^{C_u \times C_u}$. The received signal at the jth BS $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$, when using the superimposed pilot scheme, can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \left(\rho_d \mathbf{x}_{\ell,k} + \rho_p \mathbf{p}_{\ell,k}\right)^T + \mathbf{W}_j \quad (5)$$

where ρ_d^2 and ρ_p^2 are the fractions of the transmit powers reserved for the pilot and data symbols, respectively. The total transmitted power p_u is given as $p_u = \rho_p^2 + \rho_d^2$.

4.1. Non-Iterative Channel Estimation

The LS estimate of the channel of the mth user at the jth BS can be written as

$$\widehat{\mathbf{h}}_{j,j,m} \triangleq \frac{1}{C_u \rho_p} \mathbf{Y}_j \mathbf{p}_{j,m}^* = \mathbf{h}_{j,j,m} + \frac{\rho_d}{C_u \rho_p} \mathbf{W}_j \mathbf{p}_{j,m}^* \quad (6)$$

$$+ \frac{\rho_d}{C_u \rho_p} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \mathbf{x}_{\ell,k}^T \mathbf{p}_{j,m}^*.$$

The estimated channel can be used to obtain the data vector using matched filtering (MF) as

$$(\widehat{\mathbf{x}}_{j,m})^T = \eta \left\{ \frac{1}{M \rho_d} \widehat{\mathbf{h}}_{j,j,m}^H \left(\mathbf{Y}_j - \rho_p \widehat{\mathbf{h}}_{j,j,m} \mathbf{p}_{j,m}^T\right) \right\} \quad (7)$$

where $\eta\{\cdot\}$ is an hard-slicing function that replaces each element of the input vector with the constellation point that is closest in Euclidean distance to that element and $(\cdot)^H$ denotes the Hermitian transpose. The mean-squared error (MSE) of the channel estimate in (6) can be obtained as

$$\text{MSE}_{j,j,m}^{\text{SP}} = \frac{\rho_d^2}{C_u \rho_p^2} \left(\sum_{k=0}^{K-1} \beta_{j,j,k} + \sum_{\substack{\ell=0 \\ \ell \neq j}}^{L-1} \sum_{k=0}^{K-1} \beta_{j,\ell,k} \right) + \frac{\sigma^2}{C_u \rho_p^2}. \quad (8)$$

The details of all derivations are given in [10, 11] which are our associated journal papers. Large coherence times, and in turn large values of C_u, reduce the MSE in (8) by a significant amount. Moreover, by comparing (4) and (8), it can be seen that in scenarios with little or no pilot contamination, i.e., $(\beta_{j,\ell,k} \approx 0, \ \forall \ \ell \neq j)$, the first term in (8), which is due to the overlapping data symbols, dominates and leads to a higher MSE than for the case of time-multiplexed pilots. In addition, since this scheme allows for estimating the channels of the interfering users, iterative data-aided methods can be used to remove interference from both the desired and interfering users.

4.2. Iterative Data-Aided Channel Estimation

For the sake of clarity and without loss of generality, we replace the two indices k, ℓ with a single index k that is used to index the users in all the L cells. Assuming that the path-loss coefficients of all the users are available at the jth BS, they are arranged in the decreasing order, i.e., $\beta_{j,0} \geq \beta_{j,1} \geq \ldots \geq \beta_{j,N-1}$, where $N \triangleq KL$. It is assumed that the BSs have access to the exact values of the path-loss coefficients $\beta_{j,k}$ and that there is no false-ordering. This assumption is reasonable since for large M, the path-loss coefficients can be computed

$$\mathcal{I}_{j,m}^{\mathrm{ul}}(i) \approx \frac{p_u}{M\rho_d^2} \sum_{\substack{k=0 \\ k \neq m}}^{N-1} \beta_{j,k}\beta_{j,m} + \frac{p_u}{C_u\rho_p^2} \left\{ \sum_{\substack{k=0 \\ k \in \mathcal{U}_j}}^{m-1} \beta_{j,k}^2 \alpha_{j,k}^{(i)} + \sum_{\substack{k=m \\ k \in \mathcal{U}_j}}^{N-1} \beta_{j,k}^2 \alpha_{j,k}^{(i-1)} + \sum_{\substack{k=0 \\ k \notin \mathcal{U}_j}}^{N-1} \beta_{j,k}^2 \right\} + \frac{\sigma^2 \beta_{j,m}}{M\rho_d^2}$$

$$+ \frac{p_u}{MC_u\rho_p^2} \left\{ \sum_{\substack{k=0 \\ k \in \mathcal{U}_j}}^{m-1} \sum_{\substack{n=0 \\ n \neq k}}^{N-1} \beta_{j,n}\beta_{j,k}\alpha_{j,k}^{(i)} + \sum_{\substack{k=m \\ k \in \mathcal{U}_j}}^{N-1} \sum_{\substack{n=0 \\ n \neq k}}^{N-1} \beta_{j,n}\beta_{j,k}\alpha_{j,k}^{(i-1)} + \sum_{\substack{k=0 \\ k \notin \mathcal{U}_j}}^{N-1} \sum_{\substack{n=0 \\ n \neq k}}^{N-1} \beta_{j,n}\beta_{j,k} + \sum_{k=0}^{N-1} \beta_{j,k}\sigma^2 \right\} \quad (13)$$

at the BS with negligible error by averaging the power of the channel coefficients over the entire array.

The channel vector for the mth user can be estimated similar to (6) and can be written as

$$\widehat{\mathbf{h}}_{j,m}^{(1)} = \mathbf{h}_{j,m} + \frac{\rho_d}{C_u\rho_p} \sum_{k=0}^{N-1} \mathbf{h}_{j,k}\mathbf{x}_k^T \mathbf{p}_m^* + \frac{1}{C_u\rho_p} \mathbf{W}_j \mathbf{p}_m^* \quad (9)$$

where the superscript (1) is the iteration index and it denotes the first iteration. The estimate for the mth user is contaminated by the product of the channel and data vectors of all the users projected onto the mth orthogonal pilot. Therefore, to improve the quality of the channel estimate, it is necessary to estimate the channel and data vectors of the remaining $N-1$ users and subtract them from the channel estimate. The decision-aided estimation is started with the 0th user who incidentally has the highest SINR, owing to the ordering of the users. The hard-sliced data from the 0th user is then used to correct the channel estimate of the user with the next highest SINR, which is the 1st user. This can be written as

$$\left(\widetilde{\mathbf{x}}_{j,0}^{(1)}\right)^T = \frac{1}{M\rho_d} \left(\widehat{\mathbf{h}}_{j,0}^{(1)}\right)^H \left(\mathbf{Y}_j - \rho_p \widehat{\mathbf{h}}_{j,0}^{(1)} \mathbf{p}_0^T\right) \quad (10)$$

$$\widehat{\mathbf{x}}_{j,0}^{(1)} = \eta\left(\widetilde{\mathbf{x}}_{j,0}^{(1)}\right) \quad (11)$$

$$\widehat{\mathbf{h}}_{j,1}^{(1)} = \frac{1}{C_u\rho_p} \left(\mathbf{Y}_j - \rho_d \widehat{\mathbf{h}}_{j,0}^{(1)} \left(\widehat{\mathbf{x}}_{j,0}^{(1)}\right)^T\right) \mathbf{p}_1^* \quad (12)$$

where the subscript j in $\widehat{\mathbf{x}}_{j,0}$ and $\widetilde{\mathbf{x}}_{j,0}^{(1)}$ indicates that these estimates are computed at the jth BS. To prevent error propagation, it is necessary to exclude users with poor SINR from the feedback loop. This exclusion can be accomplished by analyzing the impact of symbol errors on the MSE.

To derive the MSE, let $\mathcal{U}_j$ be the set of users whose estimated data is fed back to improve the channel estimate. It is assumed that the estimation errors of both the channel and data vectors in the feedback loop are uncorrelated in every iteration and that the estimation error in each element of $\widetilde{\mathbf{x}}_{j,m}$ is zero-mean circular complex-Gaussian with variance equal to the uplink interference power $\mathcal{I}_{j,m}^{\mathrm{ul}}(i)$. The expression for $\mathcal{I}_{j,m}^{\mathrm{ul}}(i)$, whose derivation is omitted here and will appear in [10], is given in (13) at the top of the page. The parameter $\alpha_{j,m}^{(i)}$ in (13) is the variance of each element of the symbol error vector $\Delta\mathbf{x}_{j,m}^{(\mathrm{i})} \triangleq \mathbf{x}_m - \widehat{\mathbf{x}}_{j,m}^{(i)}$. For example, when $\mathbf{x}_m$ takes

values from a quadrature phase-shift keying (QPSK) constellation, $\alpha_{j,m}^{(i)}$ is given as

$$\alpha_{j,m}^{(i)} = \begin{cases} \mathbb{E}\left\{\left|\left[\Delta\mathbf{x}_{j,m}^{(i)}\right]_n\right|^2\right\} = 4Q\left(\frac{\beta_{j,m}}{\sqrt{\mathcal{I}_{j,m}^{\mathrm{ul}}(i)}}\right), & i \geq 1 \\ 1, & i = 0 \end{cases} \quad (14)$$

where $Q(x)$ is the Q-function. Using (13) and (14), the MSE (derived in [11]) can be written as

$$\mathrm{MSE}_{j,m}^{\mathrm{SP-iter}}(i) \approx \frac{1}{C_u\rho_p^2} \left[\rho_d^2 \sum_{\substack{k=0 \\ k \in \mathcal{U}_j}}^{m-1} \beta_{j,k}\alpha_{j,k}^{(i)} + \rho_d^2 \sum_{\substack{k=m \\ k \in \mathcal{U}_j}}^{N-1} \beta_{j,k}\alpha_{j,k}^{(i-1)} \right.$$
$$\left. + \rho_d^2 \sum_{\substack{k=0 \\ k \notin \mathcal{U}_j}}^{N-1} \beta_{j,k} + \sigma^2 \right]. \quad (15)$$

Let us now assume that the decoded data vector of only the mth user is used in the feedback loop, so that the MSE after the first iteration is obtained from (15) by setting $\mathcal{U}_j = \{m\}$ as

$$\mathrm{MSE}_{j,m}^{\mathrm{SP-iter}}(2) = \frac{\rho_d^2}{C_u\rho_p^2} \left(\beta_{j,m}\alpha_{j,m}^{(1)} + \sum_{\substack{k=0 \\ k \neq m}}^{N-1} \beta_{j,k} \right) + \frac{\sigma^2}{C_u\rho_p^2}. \quad (16)$$

On the other hand, the MSE for the non-iterative scheme, given by equation (8), can be rewritten as

$$\mathrm{MSE}_{j,m}^{\mathrm{SP}} = \frac{\rho_d^2}{C_u\rho_p^2} \left(\sum_{k=0}^{N-1} \beta_{j,k} \right) + \frac{\sigma^2}{C_u\rho_p^2}. \quad (17)$$

Then, a reasonable approach for the inclusion of the mth user in the feedback loop would be to include the user only if it does not increase the MSE. The impact of including the mth user in the feedback loop is clarified by comparing (16) and (17). Therefore, with $\mathcal{U}_j = \{m\}$, the estimated data of the mth user is included in the feedback loop if

$$4Q\left(\frac{\beta_{j,m}}{\sqrt{\mathcal{I}_{j,m}^{(\mathrm{ul})}(1)}}\right) < 1. \quad (18)$$
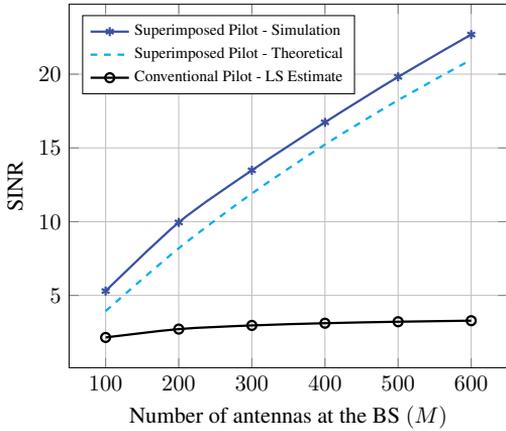
Fig. 1. Comparison of the SINR of a user in the reference BS for the LS-based and the proposed schemes at 10dB SNR.
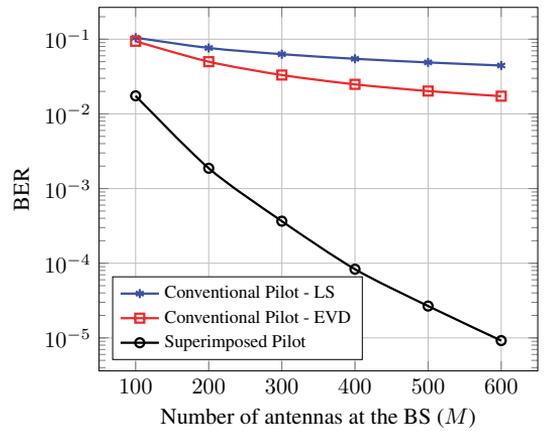
Fig. 2. Comparison of the BER for the EVD-based, the LS-based, and the proposed scheme at 10dB SNR.

If γ_j is a threshold on $\beta_{j,m}$ obtained from (18) and if $\mathcal{U}_j \triangleq \{m \mid \beta_{j,m} > \gamma_j\}$, then the estimate of the channel and data vectors for the mth user at the ith iteration can be written as

$$\widehat{\mathbf{h}}_{j,m}^{(i)} = \frac{1}{C_u \rho_p} \left[\mathbf{Y}_j - \rho_d \sum_{\substack{k=0 \\ k \in \mathcal{U}_j}}^{m-1} \widehat{\mathbf{h}}_{j,k}^{(i)} \left(\widehat{\mathbf{x}}_k^{(i)} \right)^T \right. $$

$$\left. - \rho_d \sum_{\substack{k=m \\ k \in \mathcal{U}_j}}^{N-1} \widehat{\mathbf{h}}_{j,k}^{(i-1)} \left(\widehat{\mathbf{x}}_k^{(i-1)} \right)^T \right] \mathbf{P}_{j,m}^* \quad (19)$$

$$\left(\widetilde{\mathbf{x}}_{j,m}^{(i)} \right)^T = \frac{1}{M \rho_d} \left(\widehat{\mathbf{h}}_{j,m}^{(i)} \right)^H \left(\mathbf{Y}_j - \rho_p \widehat{\mathbf{h}}_{j,m}^{(i)} \mathbf{p}_{j,m}^T \right) \quad (20)$$

$$\widehat{\mathbf{x}}_{j,m}^{(i)} = \eta \left(\widetilde{\mathbf{x}}_{j,m}^{(i)} \right) \quad (21)$$

where $\widehat{\mathbf{h}}_{j,m}^{(0)} = \mathbf{0}, \widehat{\mathbf{x}}_{j,m}^{(0)} = \mathbf{0} \ \forall \ m = 0, \ldots, N-1$. It has to be noted that the threshold is computed in the first iteration and is fixed for the subsequent iterations. Ideally, the threshold should decrease at each iteration since the improvement in the channel estimate would allow users with lower SINR to be included in $\mathcal{U}_j$. However, this is not feasible since the expression for $\mathcal{I}_{j,m}^{\text{ul}}(i)$ is only an approximation and a conservative threshold is used instead. The SINR for the iterative scheme, given the above assumptions, is written then as

$$\text{SINR}_{j,m}^{\text{SP-ul}}(i) = \frac{\beta_{j,m}^2}{\mathcal{I}_{j,m}^{\text{ul}}(i)} \quad (22)$$

5. SIMULATION RESULTS

The bit-error rate (BER) and SINR of the proposed superimposed pilot scheme is compared with that of the LS-

based channel estimate [1] and the eigenvalue decomposition (EVD)-based method [3] that use time-multiplexed pilots. The methods are tested with $L = 7$ cells and $K = 5$ users per cell. The users are assumed to be spaced at equal intervals on a circle of radius 800m from the BS in hexagonal cells with radius 1km. The path-loss coefficient is assumed to be 3 and C_u is assumed to be 100 symbols. The power of the transmitted symbols p_u is set to 1 and for simplicity, $\rho_p = \rho_d = 1/\sqrt{2}$. The simulations for the proposed superimposed pilot-based method are performed for 4 iterations.

It can be seen from the SINR plot presented in Fig. 1 that as M increases, the SINR saturates for the conventional-pilot scheme due to pilot contamination. However, the SINR increases almost linearly in M for the proposed superimposed pilots scheme and this trajectory can be potentially maintained using techniques such as adaptive modulation and coding, thereby implying that the pilot contamination effect can be eliminated. From Fig. 2, it can be seen that the proposed pilot structure and the iterative channel estimation algorithm offer a significant improvement in BER over both the LS-based and EVD-based schemes.

6. CONCLUSION

We have proposed using superimposed pilots as a superior alternative to time-multiplexed data and pilots for uplink channel estimation in massive MIMO systems. In addition, an iterative data-aided channel estimation scheme is developed. This scheme uses data symbols from both the desired and interfering users in the feedback loop, provided the SINR of these users exceeds a threshold. Theoretical expressions and numerical simulations show that the proposed method significantly alleviates the pilot contamination problem.

7. REFERENCES

[1] T.L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

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

[3] R.R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

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

[5] S. He, J.K. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3007–3021, June 2007.

[6] M. Coldrey and P. Bohlin, "Training-based MIMO systems – part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[7] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.

[8] T. Cui and C. Tellambura, "Pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE Global Telecommun. Conf.*, St. Louis, Mo, Dec. 2005, pp. 2229–2233.

[9] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol., to be published*.

[10] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO systems part I: Theory and channel estimation," *IEEE Trans. Signal Process.*, Submitted, Nov. 2015.

[11] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO systems part II: Performance analysis and optimization," *IEEE Trans. Signal Process.*, Submitted, Nov. 2015.

Publication II

K. Upadhya, S. A. Vorobyov, M. Vehkapera. Downlink performance of superimposed pilots in massive MIMO systems in the presence of pilot contamination. *Proc. IEEE Global Conf. Signal Inform. Process.*, Washington D.C., pp. 665-669, Dec. 2016.

DOWNLINK PERFORMANCE OF SUPERIMPOSED PILOTS IN MASSIVE MIMO SYSTEMS IN THE PRESENCE OF PILOT CONTAMINATION

Karthik Upadhya[*]   *Sergiy A. Vorobyov*[*]   *Mikko Vehkapera*[†]

[*] Dept. of Signal Processing and Acoustics , Aalto University, Espoo, Finland
[†]EEE Department, University of Sheffield, Sheffield, United Kingdom
Emails:karthik.upadhya@aalto.fi, svor@ieee.org, m.vehkapera@sheffield.ac.uk

ABSTRACT

In this paper, we analyze the downlink (DL) performance of superimposed pilots in time division duplexing massive multiple-input multiple-output (MIMO) systems, and show that superimposed pilots offer an increased resilience to pilot contamination with respect to time-multiplexed pilots and data. Based on a closed form expression for the DL signal-to-interference-plus-noise ratio (SINR) at the user terminal, we show that the DL SINR increases without bound with an increasing number of antennas at the base station (BS). In addition, we derive the Cramér-Rao lower bound (CRLB) for the channel estimator that uses superimposed pilots. The CRLB is compared with the mean-squared error and we show that the estimator achieves the CRLB asymptotically in the number of antennas at the BS. Simulation results validate our closed-form expressions and the performance of the proposed method.

Index Terms— Massive MIMO, pilot decontamination, superimposed pilots

1. INTRODUCTION

Existing methods for channel estimation in time-division duplexing (TDD) massive multiple-input multiple-output (MIMO) systems employ time-multiplexed pilots and data [1, 2]. Since time-multiplexed pilots and data (henceforth referred to as time-multiplexed (TM) pilots), require dedicated symbols for pilot transmission, assigning a unique orthogonal pilot to each user in the system will reduce the transmission efficiency, thereby necessitating the sharing of pilots across cells [1]. This pilot sharing results in a phenomenon referred to as pilot contamination and leads to inter-cell interference in both the uplink (UL) and downlink (DL) [1, 2]. Several methods have been proposed to decontaminate pilots by exploiting the asymptotic orthogonality between channel vectors, angular separation between the users, forward error correction (FEC) code diversity, pilot allocation, and coordination between BSs [3–11]. These methods, with the exception of [7–10], focus on separating the channels of desired and interfering users at the BS. However, there is

scope for mitigating interference from pilot contamination by focusing on the user terminals as well.

Superimposed (SP) pilots and embedded training have been studied extensively for use in MIMO systems [12–14], especially when it is impractical to reserve dedicated symbols for pilot transmission [12]. SP pilots have been proposed for massive MIMO in our previous papers [15, 16] in which we describe a non-iterative and an iterative data-aided channel estimation scheme, and show that the UL performance offered by SP pilots is resilient to pilot contamination in scenarios with high inter-cell interference. The UL performance of SP and TM pilots have been compared in greater detail in [15, 16].

In this paper,[1] we find a closed-form expression for the DL signal-to-interference-plus-noise ratio (SINR) at the user terminal when the channel estimate based on SP pilots is used in a matched-filter (MF) precoder at the BS. In addition, we compare the mean-squared error (MSE) of the channel estimated from SP pilots with its Cramér-Rao lower bound (CRLB). Simulations demonstrate the validity of these expressions and the superiority of the proposed method with respect to channel estimates based on TM pilots.

2. SYSTEM MODEL

We consider a TDD massive MIMO system with L cells and K single-antenna users per cell. Each cell has a BS with $M \gg K$ antennas. The number of symbols C, for which the channel is coherent, is assumed to be divided into C_u and C_d symbols corresponding to the UL and DL time slots, respectively. Using the tuple (ℓ, k) to denote user k in cell ℓ, the matrix of received symbols $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$ at the j'th BS in the UL time slot can be written as[2]

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \mathbf{s}_{\ell,k}^T + \mathbf{W}_j \qquad (1)$$

---

[1]This paper is an overview-type version of a relevant part of our submitted journal paper [17].

[2]We have considered narrow-band flat fading channels in this paper. The analysis would also hold in frequency-selective channels, provided modulation schemes such as orthogonal frequency division multiplexing are used.

where $\mu_{\ell,k}$ is the power with which user (ℓ, k) transmits the vector of symbols $\mathbf{s}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$, $(\cdot)^T$ denotes the transpose, $\mathbf{h}_{j,\ell,k} \in \mathbb{C}^{M \times 1}$ is the channel vector between user (ℓ, k) and the j'th BS, and $\mathbf{W}_j \in \mathbb{C}^{M \times C_u}$ is the matrix of additive white Gaussian noise at the BS, whose columns are independent and identically distributed (i.i.d) as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. The channel vector $\mathbf{h}_{j,\ell,k} \in \mathbb{C}^{M \times 1}$ is assumed to be distributed as $\mathcal{CN}\left(\mathbf{0}, \beta_{j,\ell,k}\mathbf{I}\right)$ where $\beta_{j,\ell,k}$ is the large-scale path-loss coefficient.[3] In addition, the channel vectors are assumed to be asymptotically orthogonal [1]. The transmit power $\mu_{\ell,k}$ is set according to the statistics-aware power control scheme in [8], i.e., $\mu_{\ell,k} \triangleq \omega / \beta_{\ell,\ell,n}$, where ω is a design parameter chosen so that the transmit powers of all users satisfy a power constraint. Users whose channels are severely attenuated are denied service. Assuming reciprocity at the BS, for DL transmission, the BS employs $\widehat{\mathbf{h}}_{\ell,\ell,k}$, which is an estimate of $\mathbf{h}_{\ell,\ell,k}$, in an MF precoder. If $d_{\ell,k} \in \mathbb{C}$ is the data symbol transmitted to user (ℓ, k), the received symbol at user (j, m) can be written as

$$\widehat{d}_{j,m} = \frac{1}{M}\left(\sum_{\ell=0}^{L-1}\mathbf{h}_{\ell,j,m}^T\sum_{k=0}^{K-1}\widehat{\mathbf{h}}_{\ell,\ell,k}^* d_{\ell,k} + \eta_{j,m}\right) \quad (2)$$

where $(\cdot)^*$ denotes the complex conjugate and $\eta_{j,m}$ is zero-mean additive white Gaussian noise at the user terminal with variance σ^2. The data symbols $d_{\ell,k}$ are assumed to have zero mean, unit variance, and take values from a given constellation.

3. IMPACT OF PILOT CONTAMINATION IN THE DOWNLINK

We compare here the MSE performance of channel estimates obtained from TM and SP pilots, and also compare the latter with its CRLB. In addition, we evaluate the DL SINR at the user terminal when these channel estimates are employed in an MF precoder at the BS.

3.1. Time-Multiplexed Pilots

In the case of TM pilots, each user transmits a pilot of length $\tau \geq K$ symbols for channel estimation followed by UL data. Assuming that all the pilot transmissions are synchronized, the LS estimate of the channel of user (j, m) can be easily found as [1, 2]

$$\widehat{\mathbf{h}}_{j,j,m}^{\mathrm{TM}} = \mathbf{h}_{j,j,m} + \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \mathbf{h}_{j,\ell,m} + \mathbf{w}_{j,m} \quad (3)$$

where r is the pilot reuse factor in [8], $\mathcal{L}_j(r)$ is the subset of the L cells that use the same pilots as cell j, $\mathbf{w}_{j,m} \triangleq$

$\mathbf{W}_j \boldsymbol{\phi}_{j,m}^* / \tau$, and $\boldsymbol{\phi}_{j,m}$ is the orthogonal pilot sequence transmitted by user (j, m) such that $\boldsymbol{\phi}_{j,m}^H \boldsymbol{\phi}_{\ell,k} = \tau \delta_{m,k}$ where $\delta_{m,m}$ denotes the Kronecker delta function. It can be seen from (3) that the channel estimate of user (j, m) is contaminated by the channel vectors of users in neighboring cells that use the same set of pilots. The normalized MSE of the channel estimate in (3) can then be obtained as

$$\mathrm{MSE}_{j,m}^{\mathrm{TM}} \triangleq \frac{1}{M}\mathbb{E}\left\{\|\widehat{\mathbf{h}}_{j,j,m}^{\mathrm{TM}} - \mathbf{h}_{j,j,m}\|^2\right\} = \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}}\beta_{j,\ell,m} + \frac{\sigma^2}{\tau}$$
$$(4)$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector. The DL SINR at user (j, m), when the channel estimate in (3) is used in an MF precoder at the BS and when $M \to \infty$, can be written as [1]

$$\mathrm{SINR}_{j,m}^{\mathrm{TM-dl}} = \frac{\beta_{j,j,m}^2}{\displaystyle\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}}\beta_{\ell,j,m}^2} \ . \quad (5)$$

3.2. Superimposed Pilots

In the case of SP pilots, the users transmit the pilot symbol at a reduced power alongside the data [16], i.e., $\mathbf{s}_{\ell,k} = \rho\mathbf{x}_{\ell,k} + \lambda\mathbf{p}_{\ell,k}$ where $\mathbf{p}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$ is the pilot sequence transmitted by user (ℓ, k), and λ^2 and ρ^2 are the fractions of transmit powers allocated for transmitting the pilots and data, respectively, such that $\rho^2 + \lambda^2 = 1$.[4] The pilot sequences $\mathbf{p}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$ are taken from the columns of an orthogonal matrix $\mathbf{P}$, such that $\mathbf{P}^H\mathbf{P} = C_u\mathbf{I}_{C_u}$. For the rest of the paper, we assume that the number of users in the L cells is smaller than C_u, thereby allowing each user to be assigned a unique pilot sequence. In deriving the DL SINR, for the sake of clarity and simplicity, we restrict our attention to the non-iterative channel estimation method described in [15, 16]. The least-squares (LS) estimate[5] of the channel of user (j, m) can be found to be [15, 16]

$$\widehat{\mathbf{h}}_{j,\ell,k}^{\mathrm{SP}} = \mathbf{h}_{j,\ell,k} - \Delta\mathbf{h}_{j,\ell,k} \quad (6)$$

$$\Delta\mathbf{h}_{j,\ell,k} \triangleq -\frac{\rho}{C_u\lambda}\sum_{n=0}^{L-1}\sum_{p=0}^{K-1}\mathbf{h}_{j,n,p}\mathbf{x}_{n,p}^T\mathbf{p}_{\ell,k}^* - \frac{\mathbf{W}_j\mathbf{p}_{\ell,k}^*}{C_u\lambda} \quad (7)$$

In addition, each user in the system is assigned a different pilot sequence such that $\mathbf{p}_{\ell,k}^H\mathbf{p}_{n,p} = C_u\delta_{\ell,n}\delta_{k,p}$. From (7), the normalized MSE of the channel estimate in (6) can be

---

[3]In this paper, for the sake of clarity and simplicity, shadowing is not taken into account. However, the analysis and the algorithms hold if each user is associated with the strongest BS. Moreover, the conclusions made for i.i.d. channels are also applicable in the case of correlated channels.

[4]We assume that UL transmissions in the L cells are synchronized. For example, if $L = 7$, then the synchronization is limited to the first tier of interfering cells. This requirement on synchronization is equivalent to TM pilots that are reused with a pilot reuse factor $r = 3$ [8].

[5] We restrict ourselves to the LS method for the sake of convenience. While, the linear minimum mean-squared error criterion can be used to obtain the estimator, it does not offer any benefit in the asymptotic performance.

$$\text{SINR}_{j,m}^{\text{SP}-\text{dl}} = \beta_{j,j,m}^2 \left(\frac{\rho^2 K}{C_u \lambda^2} \sum_{\ell=0}^{L-1} \beta_{\ell,j,m}^2 + \frac{1}{M} \left(\frac{\rho^2 K}{C_u \lambda^2} \sum_{\ell=0}^{L-1} \sum_{n=0}^{L-1} \sum_{p=0}^{K-1} \beta_{\ell,j,m} \beta_{\ell,n,p} + \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \beta_{\ell,j,m} \beta_{\ell,\ell,k} + \sigma^2 \right) \right)^{-1} \quad (11)$$

obtained as

$$\text{MSE}_{j,m}^{\text{SP}} \triangleq \frac{1}{M} \mathbb{E} \left\{ \| \mathbf{h}_{j,j,m}^{\text{SP}} - \mathbf{h}_{j,j,m} \|^2 \right\}$$

$$= \frac{\rho^2}{C_u \lambda^2} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \beta_{j,\ell,k} + \frac{\sigma^2}{\lambda^2 C_u} . \quad (8)$$

The first term in (8) is due to the interference that results from transmitting data alongside the pilots and the second term is due to the additive noise at the receiver.

The CRLB[6] of the channel estimate is found in [17] to be

$$\text{CRLB} \left(\mathbf{h}_{j,j,m} \right) = \left(\beta_{j,j,m} + \frac{C_u}{\sigma^2} \right)^{-1} \approx \frac{M \sigma^2}{C_u} . \quad (9)$$

The approximation in (9) is valid when $\sigma^2 / C_u \ll \beta_{j,j,m}$. Therefore, we have the relation

$$\text{MSE}_{j,m}^{\text{SP}} \geq \frac{1}{M} \text{CRLB} \left(\mathbf{h}_{j,j,m} \right) = \frac{\sigma^2}{C_u} . \quad (10)$$

The DL SINR, when the channel estimate in (6) is used in an MF precoder, can be written as (11) (shown at the top of the page) and it is derived in [17]. The expression for the DL SINR, when $M \to \infty$, can then be written as +1

$$\text{SINR}_{j,m}^{\text{SP}-\text{dl}} = \frac{\beta_{j,j,m}^2}{\frac{\rho^2 K}{C_u \lambda^2} \sum_{\ell=0}^{L-1} \beta_{\ell,j,m}^2} . \quad (11)$$

In [16], we have shown that ρ^2 and λ^2 can be chosen to maximize a lower bound on the UL sum rate and obtained them as

$$\rho_{\text{opt}}^2 = \left(1 + \sqrt{\frac{M + LK}{C_u}} \right)^{-1} \quad (12)$$

$$\lambda_{\text{opt}}^2 = 1 - \rho_{\text{opt}}^2 = \left(1 + \sqrt{\frac{C_u}{M + LK}} \right)^{-1} . \quad (13)$$

Substituting (12) and (13) into (8), the normalized MSE of the channel estimate can be rewritten as

$$\text{MSE}_{j,m}^{\text{SP}} \Big|_{\rho_{\text{opt}}, \lambda_{\text{opt}}} = \frac{1}{\sqrt{(M + LK) C_u}} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \beta_{j,\ell,k} + \frac{\sigma^2}{\lambda_{\text{opt}}^2 C_u} . \quad (14)$$

Inspecting (14), it can be seen that unlike $\text{MSE}_{j,m}^{\text{TM}}$, $\text{MSE}_{j,m}^{\text{SP}}$ decreases at a rate proportional to $\sqrt{M}$ and achieves its CRLB

asymptotically, when optimized values of ρ^2 and λ^2 are employed. This reduction in MSE also results in an increasing SINR in the DL. Substituting (12) and (13) into (11), the DL SINR with optimized values of ρ^2 and λ^2 can be written as

$$\text{SINR}_{j,m}^{\text{SP}-\text{dl}} \Big|_{\rho_{\text{opt}}, \lambda_{\text{opt}}} \approx \frac{\sqrt{C_u (M + LK)} \beta_{j,j,m}^2}{K \sum_{\ell=0}^{L-1} \beta_{\ell,j,m}^2} . \quad (15)$$

It can be seen from the above equation that the DL SINR increases without bound at a rate proportional to $\sqrt{M}$.[7] This behavior is in contrast to $\text{SINR}_{j,m}^{\text{TM}-\text{dl}}$ which saturates when $M \to \infty$ due to pilot contamination.

The unbounded increase in DL SINR can be attributed to using progressively smaller values of ρ^2 with increasing antenna array gain. When M increases, the larger array gain allows for a user to employ a smaller value of ρ^2 without sacrificing the UL rate. Since, from the denominator of (11), the amount of interference caused by a user employing SP pilots is proportional to ρ^2, employing smaller values of ρ^2 leads to a corresponding reduction in the DL interference and an increase in the DL SINR.

4. SIMULATION RESULTS

In this section, we compare the MSE, DL bit-error rate (BER), DL throughput, and DL SINR of a massive MIMO system that employs the channel estimate obtained from SP pilots to a system that employs the channel estimate obtained from TM pilots. These methods are simulated in hexagonal cells of radius 1km in two scenarios. *Scenario* 1: Users are distributed uniformly in the cell. *Scenario* 2: Users in both the reference and interfering cells are equally spaced on a circle of radius 800m with the BS in the center. Our simulations are performed for $L = 7$ cells with $K = 5$ users per cell. The path-loss coefficient is set to 3. The signal-to-noise ratio (SNR), i.e., ω/σ^2 is set to 10dB, and ω is set to 1. The power allocation parameters ρ and λ are computed from (12) and (13), respectively. The UL and DL durations, i.e., C_u and C_d, respectively, are both chosen as 100 symbols. The simulations in Scenario 1 are performed for 2×10^4 realizations of user locations with 100 realizations of channel and data for each realization of user location. Similarly, the simulations in Scenario 2 are performed for 2×10^4 realizations of channel and data for fixed user location. The solid and dashed lines in

---

[6]In deriving the CRLB, the assumptions on the statistics of the channel and data vectors made in Section 2 are used.

[7]In practice, the ceiling on the DL SINR will depend on the reuse factor of SP pilots. For example, if $L = 7$, the ceiling on the DL SINR is equivalent to that obtained when TM pilots are reused with pilot reuse factor $r = 3$ [8].
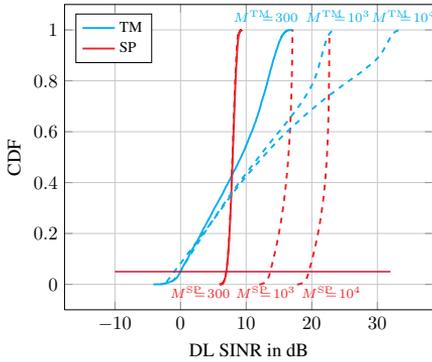
Fig. 1. Cumulative distribution of the DL SINR of an arbitrary user in the reference BS in Scenario 1. The threshold line indicates SINRs with probability ≥ 0.95. The values of M used for the TM and SP-based methods have superscripts TM and SP, respectively.
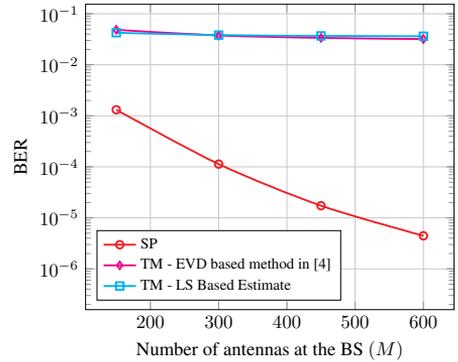
Fig. 2. BER in the DL vs. M in Scenario 1.
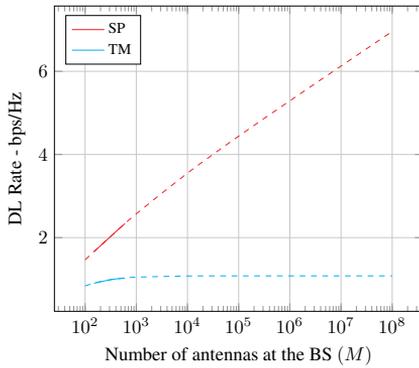
Fig. 3. DL rate vs. M in Scenario 2.
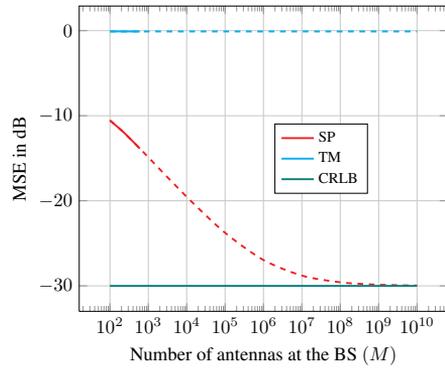
Fig. 4. MSE vs. M in Scenario 2.

the plots represent simulated and theoretical curves, respectively. The UL and DL throughput and BER is averaged over the users in the center cell. A 4-quadrature amplitude modulation (QAM) constellation is employed for computing the BER.

In Fig. 1, it can be seen that the proposed method offers a uniform quality of service over its users since $\text{SINR}_{j,m}^{\text{SP}-\text{dl}}$ is concentrated around its mean with very short tails, whereas $\text{SINR}_{j,m}^{\text{TM}-\text{dl}}$ has heavier tails, implying there are users with both very high and very low SINRs. In addition, it can be seen that the probability with which $\text{SINR}_{j,m}^{\text{SP}-\text{dl}}$ is higher than $\text{SINR}_{j,m}^{\text{TM}-\text{dl}}$ increases with M. As a result, it can be seen in Fig. 2 that the proposed method offers a significantly lower BER that decreases with M. In Fig. 3, the DL rate when SP pilots are employed is observed to be significantly higher than when TM pilots are employed. Furthermore, the DL rate

offered by the former method increases without bound in M whereas the latter method saturates at a significantly smaller value. In Fig. 4, it can be seen that $\text{MSE}_{j,m}^{\text{SP}}$ decreases with increasing M and achieves its CRLB asymptotically, whereas $\text{MSE}_{j,m}^{\text{TM}}$ remains constant with increasing M.

5. CONCLUSION

We have shown that SP pilots offer a superior performance in the DL when compared to TM pilots and data. It is shown that the fractions of transmit power allocated to pilots and data, when optimized to maximize the UL sum rate, results in a DL SINR that increases without bound at a rate proportional to $\sqrt{M}$, thereby significantly mitigating the impact of pilot contamination on the DL SINR performance. In addition, it is shown that the MSE of the channel estimate obtained from SP pilots achieves its CRLB asymptotically in M.

6. REFERENCES

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[3] H. Q. Ngo and E. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.

[4] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[5] J. Vinogradova, E. Björnson, and E. G. Larsson, "On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3421–3425.

[6] K. Upadhya and S. A. Vorobyov, "An array processing approach to pilot decontamination for massive MIMO," in *Proc. IEEE 6th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Mexico, Dec. 2015, pp. 453–456.

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

[8] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[9] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for massive MIMO," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1644–1647, Sept. 2015.

[10] V. Saxena, G. Fodor, and E. Karipidis, "Mitigating pilot contamination by pilot reuse and power control schemes for massive MIMO systems," in *Proc. IEEE 81st Vehicular Technology Conf. (VTC Spring)*, Glasgow, UK, May 2015, pp. 1–6.

[11] K. Wang and Z. Ding, "Robust receiver design based on FEC code diversity in pilot-contaminated multi-user massive MIMO systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3426–3430.

[12] M. Coldrey and P. Bohlin, "Training-based MIMO systems – part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[13] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.

[14] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.

[15] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3366–3370.

[16] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO – Part I: Theory and channel estimation," *IEEE Trans. Signal Process., Submitted*, June 2016, https://arxiv.org/abs/1603.00648.

[17] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO – Part II: Performance analysis and optimization," *IEEE Trans. Signal Process., Submitted*, June 2016, https://arxiv.org/abs/1606.04476.

Publication III

K. Upadhya, S. A. Vorobyov, M. Vehkapera. Time-multiplexed / superimposed pilot selection for massive MIMO pilot decontamination. *in Proc. IEEE Int. Conf. Acoustics Speech Signal Process.*, New Orleans, LO, pp. 3459-3463, Mar. 2017.

TIME-MULTIPLEXED / SUPERIMPOSED PILOT SELECTION FOR MASSIVE MIMO PILOT DECONTAMINATION

Karthik Upadhya[*]   *Sergiy A. Vorobyov*[*]   *Mikko Vehkapera*[†]

[*]Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland
[†]EEE Department, University of Sheffield, Sheffield, United Kingdom
E-mails: karthik.upadhya@aalto.fi, svor@ieee.org, m.vehkapera@sheffield.ac.uk

ABSTRACT

In massive multiple-input multiple-output (MIMO) systems, super-imposed (SP) and time-multiplexed (TM) pilots exhibit a complementary behavior, with the former and latter schemes offering a higher throughput in high and low inter-cell interference scenarios, respectively. Based on this observation, in this paper, we propose an algorithm for partitioning users into two disjoint sets comprising users that transmit TM and SP pilots. This selection of user sets is accomplished by minimizing the total inter-cell and intra-cell interference, and since this problem is found to be non-convex, a greedy approach is proposed to perform the partitioning. Based on simulations, it is shown that the proposed method is versatile and offers an improved performance in both high and low-interference scenarios.

Index Terms— Massive MIMO, pilot decontamination, super-imposed pilots, hybrid system, pilot selection

1. INTRODUCTION

Channel training and estimation is a critical component of any coherent transceiver. The same holds true for massive multiple-input multiple-output (MIMO) systems, which have been touted as a potential candidate for fifth generation wireless communication systems [1–3]. Existing schemes for channel training in a time-division duplexing (TDD) massive MIMO system employ time-multiplexed pilots and data (henceforth referred to as time-multiplexed (TM) pilots), wherein a subset of the symbols in the uplink (UL) time slot are reserved for pilot transmission. Maintaining high transmission efficiency necessitates the reuse of pilot sequences across cells, which leads to a phenomenon called 'pilot contamination' that limits the UL and downlink (DL) transmission efficiency [1,4,5]. Methods for mitigating pilot contamination utilize additional information about pilot transmissions, such as asymptotic orthogonality of user channels, non-overlapping user angle spread at the base station (BS), coordination between BSs, forward error correction (FEC) code diversity, pilot assignment, power control, and pilot reuse to assign unique signatures to users in order to improve their channel separability [6–15].

Superimposed (SP) pilots have recently been introduced as an alternative pilot structure for massive MIMO [16,17]. Since superimposed pilots do not require a separate set of symbols for transmitting pilots and data, they offer a larger set of orthogonal pilots and therefore, do not need to be reused as often as TM pilots. This allows SP pilots to offer superior UL and DL throughput in high interference scenarios, when compared to its TM counterpart [17,18]. However, in low inter-cell interference scenarios, TM pilots are superior since the interference from data that is transmitted alongside SP pilots results in a ceiling on its throughput.

In this paper[1], we utilize the complementary behavior of TM and SP pilots to develop an approach for selecting the type of pilot that is transmitted by a particular user. In order to perform this selection, we propose a novel framework that is based on minimizing the total inter-cell and intra-cell interference. Based on simulations, we show that the hybrid system offers a performance that is robust to interference when compared to systems that employ only TM or SP pilots.

2. SYSTEM MODEL

We consider a TDD massive MIMO system with L cells and K users per cell. Each cell has a BS with $M \gg K$ antennas. The number of symbols over which the channel is coherent C is divided into C_u and C_d symbols for the UL and DL time slots, respectively. The channel is assumed to be static within the coherence block and realizations are assumed to be independent between coherence blocks. Using the tuple (ℓ, k) to denote user k in cell ℓ, the received signal in the UL at the j'th BS $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$ can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \mathbf{s}_{\ell,k}^T + \mathbf{W}_j \qquad (1)$$

where $(\cdot)^T$ denotes the transpose, $\mathbf{s}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$ and $\mu_{\ell,k}$ are the transmitted symbol and the UL transmit power, respectively, of user (ℓ, k), $\mathbf{W}_j \in \mathbb{C}^{M \times C_u}$ is the matrix corresponding to additive white Gaussian noise at the BS with each column mutually independent of the other columns and distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$, $\mathbf{h}_{j,\ell,k} \in \mathbb{C}^{M \times 1}$ is the channel vector between user (ℓ, k) and BS j. The channel vector $\mathbf{h}_{j,\ell,k}$ is assumed to be distributed as $\mathcal{CN}(0, \beta_{j,\ell,k} \mathbf{I}_M)$ with $\beta_{j,\ell,k}$ denoting the large-scale path-loss coefficient.[2] The parameter $\mu_{\ell,k}$ is chosen using the statistics-aware power control scheme [11], i.e., $\mu_{\ell,k} \triangleq \omega / \beta_{j,\ell,k}$ where ω is a design parameter chosen such that each user satisfies its power constraint. When this power control scheme is employed, the effective path-loss coefficient between the user and the BS can be written as $\bar{\beta}_{j,\ell,k} \triangleq \mu_{\ell,k} \beta_{j,\ell,k}$. In the rest of the paper, we drop the over-bar in $\bar{\beta}_{j,\ell,k}$.

If a matched filter (MF) based precoder is used in the DL and if $d_{\ell,k}$ is the symbol transmitted by BS ℓ to its k'th user, the estimate

---

[1]This paper is a condensed version of our submitted journal paper [19].
[2]We do not consider shadowing in this paper. However, the framework and the algorithm are valid in the presence of shadowing, provided each user is associated with its strongest BS.

of the data at user (j, m) can be written as

$$\widehat{d}_{j,m} = \frac{1}{M} \left(\sum_{\ell=0}^{L-1} \mathbf{h}_{\ell,j,m}^T \sum_{k=0}^{K-1} \widehat{\mathbf{h}}_{\ell,\ell,k}^* d_{\ell,k} + \eta_{j,m} \right) \qquad (2)$$

where $\widehat{\mathbf{h}}_{\ell,\ell,k}$ is an estimate of $\mathbf{h}_{\ell,\ell,k}$, $(\cdot)^*$ denotes the complex conjugate, and $\eta_{j,m}$ is the additive noise at the user terminal that is distributed as $\mathcal{CN}\left(0, \sigma^2\right)$.

3. EXISTING CHANNEL TRAINING SCHEMES

In this section, we briefly review TM and SP pilots and their UL and DL signal-to-interference-plus-noise ratio (SINR) performance.

3.1. Time-multiplexed Pilots

When TM pilots are employed, each user in a cell transmits a $\tau \geq K$ length orthogonal pilot followed by UL data. Assuming that all pilot transmissions are synchronized, the least squares (LS) estimate of the channel in the UL can be found as [1, 17]

$$\widehat{\mathbf{h}}_{j,j,m}^{\mathrm{TM}} = \mathbf{h}_{j,j,m} + \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \mathbf{h}_{j,\ell,m} + \mathbf{w}_{j,m} \qquad (3)$$

where $\mathcal{L}_j(r)$ is the set of cells that use the same pilots as cell j, $\mathbf{w}_{j,m} = \mathbf{W}_j \boldsymbol{\phi}_{j,m}^* / \tau$, $\boldsymbol{\phi}_{j,m}$ is the pilot sequence transmitted by user (j, m), and $(\cdot)^*$ denotes the complex conjugate. From (3), it can be seen that the estimate of the channel is contaminated by the channel vectors of the users in the neighboring cells that use the same pilots. The UL and DL SINRs when this channel estimate is used in an MF-based detector and precoder, respectively, and when $M \to \infty$ can be obtained as [1]

$$\mathrm{SINR}_{j,m}^{\mathrm{TM-ul}} = \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,m}^2} \qquad (4)$$

$$\mathrm{SINR}_{j,m}^{\mathrm{TM-dl}} = \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{\ell,j,m}^2} . \qquad (5)$$

3.2. Superimposed Pilots

When SP pilots are employed, each user in the cell transmits pilots at a reduced power alongside the data, i.e., $\mathbf{s}_{j,m} = \rho \mathbf{x}_{j,m} + \lambda \mathbf{p}_{j,m}$, where $\mathbf{x}_{j,m} \in \mathbb{C}^{C_u \times 1}$ and $\mathbf{p}_{j,m} \in \mathbb{C}^{C_u \times 1}$ are the data and pilot vectors transmitted by user (j, m), respectively, with transmit powers ρ^2 and λ^2 chosen such that $\rho^2 + \lambda^2 = 1$. The pilot vectors are taken from the columns of an orthogonal matrix $\mathbf{P} \in \mathbb{C}^{C_u \times C_u}$. The LS estimate of the channel when the users transmit SP pilots can be written as [16, 17]

$$\widehat{\mathbf{h}}_{j,\ell,k}^{\mathrm{SP}} = \mathbf{h}_{j,\ell,k} + \frac{\rho}{C_u \lambda} \sum_{n=0}^{L-1} \sum_{p=0}^{K-1} \mathbf{h}_{j,n,p} \mathbf{x}_{n,p}^T \mathbf{p}_{\ell,k}^* - \frac{\mathbf{W}_j \mathbf{p}_{\ell,k}^*}{C_u \lambda} . \qquad (6)$$
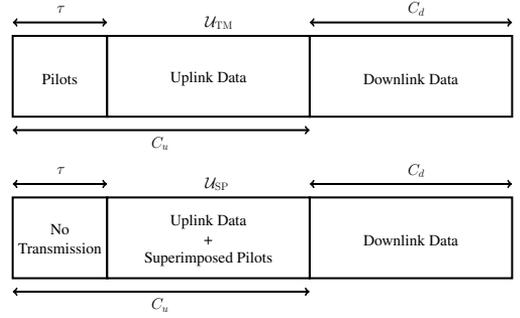
Fig. 1. Frame structure of a hybrid system with users employing TM and SP pilots.

It has been shown in [17] that the values of ρ and λ can be chosen to maximize a lower bound on the UL sum rate as

$$\rho^2 = \left(1 + \sqrt{\frac{M + LK}{C_u}}\right)^{-1} \qquad (7)$$

$$\lambda^2 = 1 - \rho^2 = \left(1 + \sqrt{\frac{C_u}{M + LK}}\right)^{-1} . \qquad (8)$$

Then, the UL and DL SINR of SP pilots, when $M \to \infty$, can be obtained as [17–19]

$$\mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} = \frac{\lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,j,m}^2}{\frac{1}{C_u} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \mu_{\ell,k} \beta_{j,\ell,k}^2} \qquad (9)$$

$$\mathrm{SINR}_{j,m}^{\mathrm{SP-dl}} = \frac{\sqrt{C_u \left(M + LK\right)} \beta_{j,j,m}^2}{\sum_{\ell=0}^{L-1} \sum_{k=0}^{L-1} \beta_{\ell,j,m}^2} \qquad (10)$$

and the corresponding rates in the UL and DL can be written as

$$R_{j,m}^{\mathrm{SP-ul}} = \frac{C_u}{C} \log_2 \left(1 + \mathrm{SINR}_{j,m}^{\mathrm{SP-ul}}\right) \qquad (11)$$

$$R_{j,m}^{\mathrm{TP-dl}} = \frac{C_d}{C} \log_2 \left(1 + \mathrm{SINR}_{j,m}^{\mathrm{SP-dl}}\right) . \qquad (12)$$

In the rest of the paper, for the sake of clarity and convenience, we use the non-iterative method for channel estimation described in [16, 17]. With suitable modifications, the approach can be extended to the case when iterative methods are used.

4. PILOT SELECTION

With the hybrid system, a user transmits either TM or SP pilots. As shown in Fig. 1, users in $\mathcal{U}_{\mathrm{TM}}$ transmit TM pilots for τ symbols followed by UL data. Users in $\mathcal{U}_{\mathrm{SP}}$ maintain radio silence for τ symbols and then transmit SP pilots and data. In this paper, given a set of K users per cell and the path-loss coefficients $\beta_{j,\ell,k}, \forall j, \ell, k$, we aim at partitioning the users into disjoint sets $\mathcal{U}_{\mathrm{TM}}$ and $\mathcal{U}_{\mathrm{SP}}$ by minimizing the total inter-cell and intra-cell interference.

The received signal at BS j for the proposed pilot system can be written as

$$\mathbf{Y}_j = \mathbf{Y}_j^{\mathrm{TM}} + \mathbf{Y}_j^{\mathrm{SP}} + \mathbf{W}_j \qquad (13)$$

where $\mathbf{Y}_j^{\text{TM}}$ and $\mathbf{Y}_j^{\text{SP}}$ are the received signals from the users in $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$, respectively. From Fig. 1, $\mathbf{Y}_j^{\text{TP}}$ and $\mathbf{Y}_j^{\text{SP}}$ can be written as

$$\mathbf{Y}_j^{\text{TM}} \triangleq \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\text{TM}}}}^{L-1}\sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k}\left[\boldsymbol{\phi}_{\ell,k}^T, \sqrt{p_u}\mathbf{x}_{\ell,k}^T\right] \tag{14}$$

$$\mathbf{Y}_j^{\text{SP}} \triangleq \sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\text{SP}}}}^{L-1}\sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k}\left[\mathbf{0}_{1\times\tau}, \rho\mathbf{x}_{\ell,k}^T + \lambda\mathbf{p}_{\ell,k}^T\right] \tag{15}$$

where the users in $\mathcal{U}_{\text{TM}}$ transmit data at power p_u. Using the LS estimates of the channels for the users in $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$, which can be obtained similar to (3) and (6), respectively, the UL and DL SINRs of users in $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$, when $M \to \infty$, can be obtained as [1, 17, 18]

$$\text{SINR}_{j,m}^{\text{TM-ul}} = \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,m)\in\mathcal{U}_{\text{TM}}}} \beta_{j,\ell,m}^2} \tag{16}$$

$$\text{SINR}_{j,m}^{\text{TM-dl}} = \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,m)\in\mathcal{U}_{\text{TM}}}} \beta_{\ell,j,m}^2} \tag{17}$$

$$\text{SINR}_{j,m}^{\text{SP-ul}} \approx \frac{\beta_{j,j,m}^2}{\frac{1}{(C_u-\tau)\lambda^2}\sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\text{SP}}}}^{L-1}\sum_{k=0}^{K-1} \beta_{j,\ell,k}^2} \tag{18}$$

$$\text{SINR}_{j,m}^{\text{SP-dl}} \approx \frac{\beta_{j,j,m}^2}{\frac{\rho^2}{(C_u-\tau)\lambda^2}\sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\text{SP}}}}^{L-1}\sum_{k=0}^{K-1} \beta_{\ell,j,m}^2} \tag{19}$$

where the approximations in (18) and (19) have been made assuming that the users in $\mathcal{U}_{\text{TM}}$ do not interfere with the users in $\mathcal{U}_{\text{SP}}$.[3][4] Since, by design, the users in $\mathcal{U}_{\text{SP}}$ do not interfere with the transmission of $\mathcal{U}_{\text{TM}}$, the transmissions of both sets of users can be considered to be independent of each other. In addition, for the sake of simplicity, it is assumed that M is large enough such that the above expressions are valid.

In order to obtain an approach to partition the users into $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$, we define the following terms. Let $I_{j,m}^{\text{TM-ul}}$ and $I_{j,m}^{\text{TM-dl}}$ be the total interference in the UL and DL, respectively, caused by user (j,m) when assigned to $\mathcal{U}_{\text{TM}}$. Similarly, let $I_{j,m}^{\text{SP-ul}}$ and $I_{j,m}^{\text{SP-dl}}$ be the total interference in the UL and DL, respectively, caused by user (j,m) when assigned to $\mathcal{U}_{\text{SP}}$. Then, from the denominators of (16), (17), (18), and (19), $I_{j,m}^{\text{TM-ul}}$, $I_{j,m}^{\text{TM-dl}}$, $I_{j,m}^{\text{SP-ul}}$, and $I_{j,m}^{\text{SP-dl}}$ can be obtained as [19, Section IV.A]

$$I_{j,m}^{\text{TM-ul}} = \sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,k)\in\mathcal{U}_{\text{TM}}}}\sum_{k=0}^{K-1} \beta_{\ell,j,k}^2\delta_{m,k} = \sum_{\substack{\ell\neq j \\ \ell\in\mathcal{L}_j(r) \\ (\ell,m)\in\mathcal{U}_{\text{TM}}}} \beta_{\ell,j,m}^2 \tag{20}$$

$$I_{j,m}^{\text{TM-dl}} = \sum_{\substack{n\neq j \\ n,\ell\in\mathcal{L}_j(r) \\ (n,k)\in\mathcal{U}_{\text{TM}}}}\sum_{\ell}\sum_{k=0}^{K-1} \beta_{n,\ell,k}^2\delta_{j,\ell}\delta_{m,k} = \sum_{\substack{n\neq j \\ n\in\mathcal{L}_j(r) \\ (n,m)\in\mathcal{U}_{\text{TM}}}} \beta_{n,j,m}^2 \tag{21}$$

$$I_{j,m}^{\text{SP-ul}} = \frac{1}{(C_u-\tau)\lambda^2}\sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\text{SP}}}}^{L-1}\sum_{k=0}^{K-1} \beta_{\ell,j,m}^2 \tag{22}$$

$$I_{j,m}^{\text{SP-dl}} = \frac{\rho^2}{(C_u-\tau)\lambda^2}\sum_{\substack{\ell=0 \\ (\ell,k)\in\mathcal{U}_{\text{SP}}}}^{L-1}\sum_{k=0}^{K-1} \beta_{\ell,j,m}^2 = \rho^2 I_{j,m}^{\text{SP-ul}}. \tag{23}$$

If ξ^{ul} and ξ^{dl} are weights such that $\xi^{\text{ul}} + \xi^{\text{dl}} = 1$, then the total cost due to inter-cell and intra-cell interference can be expressed as

$$I(\mathcal{U}_{\text{TM}}, \mathcal{U}_{\text{SP}}) \triangleq \sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\left(T_{\ell,k}^{\text{TM}}\mathbf{1}_{\{(\ell,k)\in\mathcal{U}_{\text{TM}}\}} + T_{\ell,k}^{\text{SP}}\mathbf{1}_{\{(\ell,k)\in\mathcal{U}_{\text{SP}}\}}\right) \tag{24}$$

where $T_{\ell,k}^{\text{TM}}$ and $T_{\ell,k}^{\text{SP}}$ are the total costs incurred when user (ℓ,k) is assigned to $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$, respectively, and can be written as

$$T_{\ell,k}^{\text{TM}} \triangleq \xi^{\text{ul}}I_{\ell,k}^{\text{TM-ul}} + \xi^{\text{dl}}I_{\ell,k}^{\text{TM-dl}} \tag{25}$$

$$T_{\ell,k}^{\text{SP}} \triangleq \xi^{\text{ul}}I_{\ell,k}^{\text{SP-ul}} + \xi^{\text{dl}}I_{\ell,k}^{\text{SP-dl}}. \tag{26}$$

Using (24) as the objective function, the sets $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$ can be obtained as the solution of the following optimization problem

$$(\mathcal{U}_{\text{TM}}, \mathcal{U}_{\text{SP}}) = \arg\min_{\substack{\mathcal{U}_{\text{TM}}\subseteq\mathcal{U} \\ \mathcal{U}_{\text{SP}}\subseteq\mathcal{U}}} I(\mathcal{U}_{\text{TM}}, \mathcal{U}_{\text{SP}})$$

$$\text{subject to} \quad \mathcal{U}_{\text{TM}} \cup \mathcal{U}_{\text{SP}} = \mathcal{U}$$

$$\mathcal{U}_{\text{TM}} \cap \mathcal{U}_{\text{SP}} = \varnothing \tag{27}$$

where $\mathcal{U}$ is the set of all users in the system and $\varnothing$ is the null set. However, this optimization problem is combinatorial in nature and requires a search over $2^{|\mathcal{U}|}$ combinations. Alternatively, a greedy approach can be used to partition $\mathcal{U}$ into $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$. At each step, given $\mathcal{U}_{\text{TM}}$ and $\mathcal{U}_{\text{SP}}$, a user $\left(\tilde{\ell}, \tilde{k}\right)$ in $\mathcal{U}_{\text{TM}}$ is chosen as

$$\left(\tilde{\ell}, \tilde{k}\right) = \arg\max_{(\ell,k)\in\mathcal{U}_{\text{TM}}} T_{\ell,k}^{\text{TM}}. \tag{28}$$

This user is added to $\mathcal{U}_{\text{SP}}$ if

$$I\left(\mathcal{U}_{\text{TM}}', \mathcal{U}_{\text{SP}}'\right) \leq I\left(\mathcal{U}_{\text{TM}}, \mathcal{U}_{\text{SP}}\right) \tag{29}$$

where $\mathcal{U}_{\text{TM}}' = \mathcal{U}_{\text{TM}}\setminus\left(\tilde{\ell}, \tilde{k}\right)$ and $\mathcal{U}_{\text{SP}}' = \mathcal{U}_{\text{SP}} \cup \left(\tilde{\ell}, \tilde{k}\right)$. The algorithm is initialized with $\mathcal{U}_{\text{TM}} = \mathcal{U}$ and is terminated when either $\mathcal{U}_{\text{TM}}$ is empty or when (29) is no longer satisfied. The approach detailed above is summarized in Algorithm 1.

5. SIMULATION RESULTS

In this section, we compare the bit error rate (BER) and throughput of the hybrid system with systems employing TM and SP pilots. The simulations are performed with hexagonal cells of 1km diameter in two scenarios (i) *Scenario* 1: The users are uniformly distributed in the cells; (ii) *Scenario* 2: the users in both the reference and interfering cells are in a fixed configuration and are equally spaced on a

---

[3]In this paper, we assume for the sake of simplicity that p_u is small enough with respect to the transmit powers of the users in $\mathcal{U}_{\text{SP}}$. This does not affect the throughput of the users in $\mathcal{U}_{\text{TM}}$, since their UL and DL SINRs are independent of p_u. In the absence of this assumption, the BS will have to estimate and remove $\mathbf{Y}_j^{\text{TP}}$ before estimating the channels of users in $\mathcal{U}_{\text{TM}}$.

[4]Moreover, it can be seen from the simulation results that the hybrid system outperforms the existing schemes despite using these approximate expressions when selecting the pilots.

Fig. 2. Sum Rate in the UL over users in the first tier of cells vs. user radius in Scenario 2

Fig. 3. Sum Rate in the DL over users in the first tier of cells vs. user radius in Scenario 2

Algorithm 1 Greedy algorithm to select $\mathcal{U}_{\mathrm{TM}}$ and $\mathcal{U}_{\mathrm{SP}}$

Data: $\beta_{j,\ell,k}, \ \forall j, \ell = 0, \ldots, L-1, \ \ k = 0, \ldots, K-1$
Initialize: $\mathcal{U}_{\mathrm{TM}} \leftarrow \mathcal{U}, \mathcal{U}_{\mathrm{SP}} \leftarrow \varnothing$
1: Compute $\left(\tilde{\ell}, \tilde{k}\right)$ as in (28)
2: Set $\mathcal{U}'_{\mathrm{TM}} \leftarrow \mathcal{U}_{\mathrm{TM}} \setminus \left(\tilde{\ell}, \tilde{k}\right)$ and $\mathcal{U}'_{\mathrm{SP}} \leftarrow \mathcal{U}_{\mathrm{SP}} \cup \left(\tilde{\ell}, \tilde{k}\right)$
3: **if** $\mathcal{U}_{\mathrm{TM}} \neq \varnothing$ **and if** $I\left(\mathcal{U}'_{\mathrm{TM}}, \mathcal{U}'_{\mathrm{SP}}\right) \leq I\left(\mathcal{U}_{\mathrm{TM}}, \mathcal{U}_{\mathrm{SP}}\right)$ **then**
4: $\quad$ $\mathcal{U}_{\mathrm{TM}} := \mathcal{U}'_{\mathrm{TM}}, \mathcal{U}_{\mathrm{SP}} := \mathcal{U}'_{\mathrm{SP}}$
5: $\quad$ Return to Step (1).
6: **else**
7: $\quad$ STOP
8: **end if**

Table 1. UL and DL performance of TM, SP, and hybrid systems in Scenario 1

	UL Sum Rate	DL Sum Rate	Total Rate	BER in the UL	BER in the DL
Hybrid System	47.72	86.46	134.19	1.31×10^{-2}	1.52×10^{-5}
TM Pilots	51.06	66.30	117.36	2.96×10^{-2}	3.66×10^{-2}
SP Pilots	35.40	75.60	111.00	2.69×10^{-2}	4.77×10^{-5}

circle with the BS in the center. The radius of this circle is varied from 0.2 to 0.9km in the simulation. The number of cells in the system is set to $L = 19$ cells with $M = 600$ antennas and $K = 5$ users per cell. However, the optimization is performed over 7 cells which consists of the central and first tier of cells. In addition, the BER and throughput is measured over the users in the central and first tier of cells. The number of symbols in the UL and DL, i.e., C_u and C_d are set to 40 symbols. The values of ρ and λ are computed from (7) and (8), respectively. The signal-to-noise ratios (SNRs) in the UL and DL, i.e., ω/σ^2 are set to 10dB, where ω is the design parameter in the statistics-aware power control scheme and is set to 1. In addition p_u for the users in $\mathcal{U}_{\mathrm{TM}}$ is set to 0.1. For the hybrid system, the parameters ξ^{ul} and ξ^{dl} are both set to 0.5. The results in Scenario 1 are generated by averaging over 10^3 realizations of user locations. For each realization of user locations, the throughput and
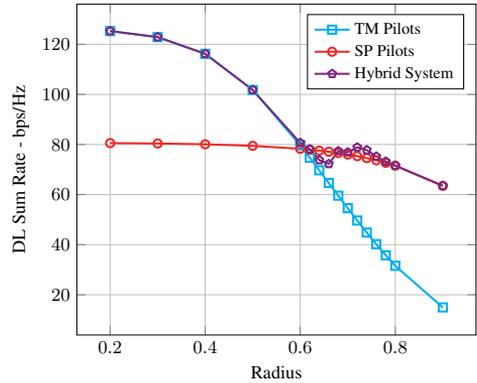
BER is averaged over 100 realizations of channel and data vectors. The results in Scenario 2 are obtained by averaging over 10^4 realizations of channel and data. Gaussian signaling and 4-quadrature amplitude modulation (QAM) are used to compute the throughput and BER, respectively.

In Figs. 2 and 3, the UL and DL sum rates, respectively, for the systems employing TM and SP pilots, and the hybrid system are plotted against the user radius in the cell. As can be observed from the figures, the UL and DL throughputs of TM pilots are higher than that of SP pilots in the range of radius $[0.2, 0.6]$. Similarly, the UL and DL throughputs of SP pilots are higher than that of TM pilots in the range $[0.8, 1]$. Therefore, in these two ranges, there is a clear choice of $\mathcal{U}_{\mathrm{TM}}$ and $\mathcal{U}_{\mathrm{SP}}$ for the partitioning algorithm. However, in the range $[0.6, 0.8]$, the behavior of the greedy algorithm is dependent on the parameters ξ^{ul} and ξ^{dl}, and since both parameters are equal, the algorithm attempts to strike a balance between the UL and DL sum rates and offers a performance that is in between TM and SP pilots. In addition, since Algorithm 1, is greedy, the UL and DL performance of the resulting partition is non-smooth across different user radius.

In Table 1, the throughput and BER in Scenario 1 are detailed for TM and SP pilot-based systems as well as the proposed hybrid system. The proposed hybrid system offers roughly 14.34% higher total throughput when compared to the existing methods. In addition, the proposed method offers a higher throughput in the DL than TM pilot-based methods. However, this improved DL performance comes at a cost of lower throughput in the UL, but the hybrid system allows the DL rate to be traded-off against the UL rate through the parameters ξ^{ul} and ξ^{dl}.

6. CONCLUSION

We have proposed an algorithm, for TDD massive MIMO systems, that minimizes the total inter-cell and intra-cell interference by selecting the type of pilot that a user transmits. By means of simulations, it is shown that the proposed scheme offers a performance that is robust with respect to the user location in the cell. However, the objective function, that is described in this paper, is non-convex and requires cooperation between BSs. Obtaining a distributed solution to solve this optimization problem is a potential direction for future research.

7. REFERENCES

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[5] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[6] K. Upadhya and S. A. Vorobyov, "An array processing approach to pilot decontamination for massive MIMO," in *Proc. IEEE 6th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Mexico, Dec. 2015, pp. 453–456.

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

[8] H. Q. Ngo and E. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.

[9] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[10] J. Vinogradova, E. Björnson, and E. G. Larsson, "On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3421–3425.

[11] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[12] V. Saxena, G. Fodor, and E. Karipidis, "Mitigating pilot contamination by pilot reuse and power control schemes for massive MIMO systems," in *Proc. IEEE 81st Vehicular Technology Conf. (VTC Spring)*, Glasgow, UK, May 2015, pp. 1–6.

[13] K. Wang and Z. Ding, "Robust receiver design based on FEC code diversity in pilot-contaminated multi-user massive MIMO systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3426–3430.

[14] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for massive MIMO," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1644–1647, Sept. 2015.

[15] I. Atzeni, J. Arnau, and M. Debbah, "Fractional pilot reuse in massive mimo systems," in *IEEE Int. Conf. on Communication Workshop (ICCW)*, London, UK, June 2015, pp. 1030–1035.

[16] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 3366–3370.

[17] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process., Submitted*, June 2016, `https://arxiv.org/abs/1603.00648`.

[18] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Downlink performance of superimposed pilots in massive MIMO systems in the presence of pilot contamination," in *Proc. IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, Greater Washington, D.C., USA, Dec. 2016, pp. 665–669.

[19] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO – Part II: Performance analysis and optimization," *IEEE Trans. Signal Process., Submitted*, June 2016, `https://arxiv.org/abs/1606.04476`.

Publication IV

K. Upadhya, S. A. Vorobyov, M. Vehkapera. Superimposed pilots are superior for mitigating pilot contamination in massive MIMO. *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917-2932, Jun. 2017.

Superimposed Pilots Are Superior for Mitigating Pilot Contamination in Massive MIMO

Karthik Upadhya, *Student Member, IEEE*, Sergiy A. Vorobyov, *Senior Member, IEEE*, and Mikko Vehkapera, *Member, IEEE*

Abstract—In this paper, superimposed pilots are introduced as an alternative to time-multiplexed pilot and data symbols for mitigating pilot contamination in massive multiple-input multiple-output (MIMO) systems. We propose a non-iterative scheme for uplink channel estimation based on superimposed pilots and derive an expression for the uplink signal-to-interference-plus-noise ratio (SINR) at the output of a matched filter employing this channel estimate. Based on this expression, we observe that power control is essential when superimposed pilots are employed. Moreover, the quality of the channel estimate can be improved by reducing the interference that results from transmitting data alongside the pilots, and an intuitive iterative data-aided scheme that reduces this component of interference is also proposed. Approximate expressions for the uplink SINR are provided for the iterative data-aided method as well. In addition, we show that a hybrid system with users utilizing both time-multiplexed and superimposed pilots is superior to an optimally designed system that employs only time-multiplexed pilots, even when the non-iterative channel estimate is used to build the detector and precoder. We also describe a simple approach to implement this hybrid system by minimizing the overall inter- and intracell interference. Numerical simulations demonstrating the performance of the proposed channel estimation schemes and the superiority of the hybrid system are also provided.

Index Terms—Massive MIMO, pilot contamination, superimposed pilots.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (MIMO) systems, proposed in [1], have received significant interest in recent years as a candidate for fifth-generation mobile communication technologies [2]–[4]. These systems are a variation of multi-user MIMO (MU-MIMO) and have a large number of base station (BS) antennas that result in an improved spectral efficiency through spatial multiplexing [5], [6]. Under favorable propagation conditions [1], significant gains in throughput can

be achieved by employing simple linear processing at the BS [7], [8]. In addition, large numbers of antennas result in an improved uplink (UL) energy-efficiency [9] and render the system performance resilient to hardware impairments [10].

However, all the above mentioned benefits of a massive MIMO communication system hinge on the assumption that the BS has access to accurate estimates of the channel state information (CSI). For systems that employ either frequency division duplexing (FDD) or time division duplexing (TDD), the channel estimates are obtained using orthogonal pilot sequences. In FDD systems, each antenna at the BS transmits a pilot sequence that is orthogonal to the pilot sequences transmitted by the other antennas. Since the number of orthogonal pilot sequences required becomes proportional to the number of BS antennas, FDD is considered impractical for channel estimation in massive MIMO [5], [11]. Moreover, since the CSI corresponding to each antenna is estimated by the users, it has to be feedback from the users to the BS, consuming additional bandwidth. Consequently, massive MIMO systems are typically assumed to employ TDD with full frequency-reuse and utilize channel reciprocity to obtain CSI. In these systems, each user in a cell is assigned a different pilot sequence and these pilots are time-multiplexed with data in each coherence block. When using time-multiplexed pilots and data, the requirement for high spectral efficiency necessitates sharing of pilot sequences between adjacent cells, resulting in the channel estimates of the users in a cell being corrupted by the channel vectors of users in the adjacent cells. This phenomenon called 'pilot contamination' [12] introduces interference in both the UL and downlink (DL), degrading the overall performance of the system.

Existing methods to mitigate pilot contamination for massive MIMO are designed for the case wherein the pilots are time-multiplexed with the data, henceforth referred to as time-multiplexed pilots. In [13], it has been observed that the eigenvectors of the autocorrelation matrix of the received data correspond to the channel vectors of the desired and interfering users, and a method for channel estimation has been developed based on this observation. Pilot decontamination has been performed in [14] by projecting the contaminated channel estimate on an interference-free subspace spanned by the channel vectors of the desired users, whereas [15] derives asymptotic conditions for separability between the subspaces of the desired and interfering users. In [16], a coordinated method for pilot allocation has been proposed for separating desired and interfering users in correlated channels. A pilot decontamination method based on the array processing model has been proposed in [17] for use in parametric channels with a finite number of discrete paths. In [18], a resource allocation approach has been proposed for

optimizing the number of users scheduled in each cell in order to minimize the effect of pilot contamination. A common theme for the approaches described above, except for [16] and [18], is that they focus on decontaminating the channel estimate at the receiver, which in this case is the BS. However, since pilot contamination results from interfering pilot transmissions, there is a scope for *better separating the desired and interfering users by optimizing the pilot transmissions at the user terminal* as well.

In this paper, we propose using superimposed pilots as an alternative to, and in combination with, time-multiplexed pilots in massive MIMO systems. Methods for channel estimation based on pilots that are embedded in data, such as superimposed pilots, have been extensively studied for MIMO systems [19]–[24]. However, these papers have focused on embedded and superimposed pilots in the context of accommodating a loss in signal-to-noise ratio (SNR) in exchange for a reduced pilot transmission overhead [20], [21]. Particularly, scenarios with high user-mobility, wherein it is impractical to allocate dedicated symbols for training, have been of interest for employing superimposed pilots. In the context of multi-cell massive MIMO, provided that the number of users in the system is smaller than the number of symbols in the UL, superimposed pilots allow for each user in the system to be assigned a unique pilot sequence, enabling the receiver to estimate the channel vectors of both the desired and interfering users. In addition, these pilots mitigate pilot contamination by time-averaging over long sequences and offer a higher efficiency due to a reduced transmission overhead.

We obtain expressions for the signal-to-interference-plus-noise ratio (SINR) at the output of a matched filter (MF)-based detector when a non-iterative least-squares (LS)-based channel estimate is employed for channel estimation. Based on the SINR expression, we highlight the need for power control when superimposed pilots are employed in a massive MIMO system. Moreover, we discuss the shortcomings of the non-iterative channel estimator and propose an intuitive low-complexity iterative channel estimation scheme for superimposed pilots.[1] In addition, we introduce the concept of a hybrid system and show by means of theoretical arguments that the hybrid system is superior to its counterpart that employs only time-multiplexed pilots, even when the non-iterative channel estimator is used to obtain the channel estimate from superimposed pilots. A simple approach to design and implement this hybrid system is also detailed. Although the use of superimposed pilots requires some coordination between the BSs in assigning pilot sequences to the users and estimating their path-loss coefficients, these are minor impediments compared to the performance improvements provided by the proposed scheme.

The article in the existing literature that is closest to this paper is [26], wherein superimposed pilots have been employed in the context of multi-cell multiuser MIMO systems. However, unlike [26], the focus of our paper is to demonstrate the superiority of superimposed pilots when used in conjunction with time-multiplexed pilots in a hybrid system. The theoretical results and simulations that have been obtained are in line with this objective.

In Section II, the system model for the massive MIMO UL is described. In Section III, time-multiplexed pilots are

described and the pilot contamination problem is detailed. Section IV introduces the superimposed pilot scheme and describes the non-iterative method for channel estimation and Section V discusses the iterative data-aided scheme. In Section VI, the concept of a hybrid system that employs both time-multiplexed and superimposed pilots is introduced and in Section VII, a simple approach for implementing this hybrid system is discussed. Section VIII presents simulation results demonstrating the effectiveness of employing superimposed pilots for pilot decontamination. Section IX concludes the paper. Some of the proofs are given in Appendix.

Notation: Lower case and upper case boldface letters denote column vectors and matrices, respectively. The notations $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ represent the conjugate, transpose, Hermitian transpose, and inverse, respectively. The notation $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the complex normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $\mathbb{E}\{\cdot\}$ is used to denote the expectation operator. The notation $\mathbf{I}_N$ denotes an $N \times N$ identity matrix, and $\|\cdot\|$ and $\|\cdot\|_F$ denote the Euclidean norm of a vector and Frobenius norm of a matrix, respectively. Upper case calligraphic letters denote sets, and $\varnothing$ denotes the empty set. The notation $\mathbf{1}_{\{\mathcal{S}\}}$ represents the indicator function over the set $\mathcal{S}$, whereas $\mathrm{Card}\,(\mathcal{S})$ is used to represent its cardinality. The notation $\delta_{n,m}$ denotes the Kronecker delta function, and $\eta(\cdot)$ stands for an element-by-element decision function that replaces each element of the input vector with the constellation point that is closest in Euclidean distance to that element. The big O notation $f(x) = O(g(x))$ implies that $|f(x)|/|g(x)|$ is bounded as $x \to \infty$.

II. SYSTEM MODEL

We consider a TDD massive MIMO UL with L cells and K single-antenna[2] users per cell. Each cell has a BS with $M \gg K$ antennas. The number of symbols C, over which the channel is coherent, is assumed to be divided into C_u and C_d, which are the number of symbols in the UL and DL time slots, respectively. The matrix of received measurements $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$ at BS j can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \mathbf{s}_{\ell,k}^T + \mathbf{W}_j \tag{1}$$

where $\mu_{\ell,k}$ denotes the power with which the vector of symbols $\mathbf{s}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$ are transmitted by user k in cell ℓ, $\mathbf{W}_j \in \mathbb{C}^{M \times C_u}$ is the additive white Gaussian noise at BS j with each column distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$. Moreover, the columns of $\mathbf{W}_j$ are mutually independent of each other. The vector $\mathbf{h}_{j,\ell,k} \in \mathbb{C}^{M \times 1}$ represents the channel response between the antennas at BS j, and user k in cell ℓ, and is assumed to be distributed as[3]

$$\mathbf{h}_{j,\ell,k} \sim \mathcal{CN}(\mathbf{0}, \beta_{j,\ell,k} \mathbf{I}_M) \tag{2}$$

where $\beta_{j,\ell,k}$ denotes the large-scale path-loss coefficient which depends on the user location in the cell. In addition, the channel vectors $\mathbf{h}_{j,\ell,k}$ are assumed to be mutually independent of each other $\forall j, \ell, k$. The aforementioned statistics of the channel

---

[2]For training and channel estimation, users with $T > 1$ antennas can be treated as T separate single-antenna users.

[3]While, for the sake of simplicity, an environment with rich scattering is assumed, the conclusions made in this paper are independent of the channel distribution and only require the channel vectors of any pair of users to be asymptotically orthogonal.

vector correspond to the non-line-of-sight scenario with rich scattering [1]. By virtue of their zero mean and mutual independence, the channel vectors are asymptotically orthogonal and the following equation holds almost surely [1]

$$\lim_{M \to \infty} \frac{\mathbf{h}_{j,\ell,k}^H \mathbf{h}_{m,n,p}}{M} = \beta_{j,\ell,k}\, \delta_{j,m}\, \delta_{\ell,n}\, \delta_{k,p}, \ \forall\, j,k,\ell,m,n,p. \tag{3}$$

Moreover, $\mathbf{h}_{j,\ell,k}$ is assumed to be constant for the duration of C symbols, and $\beta_{j,\ell,k}$ is constant for a significantly longer duration which depends on the user mobility. For the sake of simplicity, the effects of shadowing are not taken into account in this paper.[4] The transmitted symbols $\mathbf{s}_{\ell,k}$ contain both pilots and data. The pilots could either be time-multiplexed or superimposed pilots, and the elements of the data vector $\mathbf{x}_{\ell,k}$ are assumed to be independent and identically distributed (i.i.d) random variables with zero-mean and unit variance and take values from an alphabet χ, which is a realistic assumption.

III. TIME-MULTIPLEXED PILOTS AND THE PILOT CONTAMINATION PROBLEM

With time-multiplexed pilots, each user in a cell transmits a $\tau \geq K$ length orthogonal pilot sequence for channel estimation followed by $C_u - \tau$ symbols of uplink data. In order to minimize the overhead incurred, it is necessary to reuse these pilot sequences in the adjacent cells. However, this pilot-reuse results in the channel estimates of the desired users being contaminated by the channel vectors of users in adjacent cells, causing interference and in turn, a loss in spectral efficiency.

It is assumed here that the transmission of the pilot sequences by the users in the L cells are synchronized, which corresponds to the worst-case scenario for pilot contamination.[5] Consider a matrix $\mathbf{\Phi} \in \mathbb{C}^{\tau \times \tau}$ whose columns $\{\phi_1, \ldots, \phi_\tau\}$ are the orthogonal pilot sequences that are transmitted by the users, i.e., $\phi_n^H \phi_p = \tau \delta_{n,p}$. If $\phi_{b_{\ell,k}}$ is the pilot sequence transmitted by user k of cell ℓ, where $b_{\ell,k} \in \{1, \ldots, \tau\}$ is the index of the transmitted pilot, and if each pilot sequence is reused once every $r \triangleq \tau/K$ cells [18], the LS estimate of the channel of user m in cell j can be obtained as [1]

$$\widehat{\mathbf{h}}_{j,j,m}^{\mathrm{TP}} \triangleq \frac{1}{\tau \sqrt{p_u}} \mathbf{Y}_j^{(p)} \phi_{b_{j,m}}^* = \mathbf{h}_{j,j,m} + \sum_{\substack{\ell=0 \\ \ell \neq j \\ \ell \in \mathcal{L}_j(r)}}^{L-1} \mathbf{h}_{j,\ell,m}$$

$$+ \frac{1}{\tau \sqrt{p_u}} \mathbf{W}_j^{(p)} \phi_{b_{j,m}}^* \tag{4}$$

where the superscript TP indicates that the estimates are computed when using time-multiplexed pilots, the superscript p indicates that the observations are made during pilot transmission, and $\mathcal{L}_j(r)$ is the subset of the L cells that use the same set of pilots as cell j. In addition, it is assumed in (4) without loss of generality that the transmit powers are same for all users employing time-multiplexed pilots, i.e., $\mu_{\ell,k} = p_u, \ \forall\, \ell, k$, and

any variation in the transmit power of an individual user is absorbed into the corresponding path-loss coefficient β. It can be observed from (4) that the estimates of the channel vectors of the users in cell j are contaminated by the channel vectors of the users in the remaining Card $(\mathcal{L}_j(r) - 1)$ cells. When $M \to \infty$, the UL SINR of user m in cell j, at the output of an MF that uses the channel estimate in (4) for detection, can be written as [1]

$$\mathrm{SINR}_{j,m}^{\mathrm{TP-ul}} = \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,m}^2}. \tag{5}$$

The corresponding throughput of the user using Gaussian signaling in the UL can then be expressed as [1]

$$\mathbf{R}_{j,m}^{\mathrm{TP-ul}} = \frac{(C_u - \tau)}{C} \log_2 \left(1 + \mathrm{SINR}_{j,m}^{\mathrm{TP-ul}}\right). \tag{6}$$

From the above equation, it can be observed that the rate per user is a function of both the overhead τ as well as the loss in SINR due to pilot contamination. A larger value of r would reduce the effect of pilot contamination and increase the SINR at the cost of a reduced transmission efficiency $(C_u - \tau)/C$.

IV. SUPERIMPOSED PILOTS

With superimposed pilots, the pilot symbols are transmitted at a reduced power alongside the data symbols, and in its simplest version, the pilot and data symbols are transmitted alongside each other for the entire duration of the uplink data slot C_u. If the total number of users in the system is smaller than the number of symbols in the uplink, i.e., $KL \leq C_u$,[6] then with superimposed pilots, each user can be assigned a unique orthogonal pilot $\mathbf{p}_{\ell,k} \in \mathbb{C}^{C_u \times 1}$. The pilots are taken from the columns of a matrix $\mathbf{P} \in \mathbb{C}^{C_u \times C_u}$ such that $\mathbf{P}^H \mathbf{P} = C_u \mathbf{I}_{C_u}$, and therefore $\mathbf{p}_{\ell,k}^H \mathbf{p}_{n,p} = C_u \delta_{\ell,n} \delta_{k,p}$. If $\rho_{\ell,k} \mathbf{x}_{\ell,k} + \lambda_{\ell,k} \mathbf{p}_{\ell,k}$ is the transmitted vector from user k in cell ℓ, then the received signal at the j'th BS $\mathbf{Y}_j \in \mathbb{C}^{M \times C_u}$, when using the superimposed pilot scheme, can be written as

$$\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \left(\rho_{\ell,k} \mathbf{x}_{\ell,k} + \lambda_{\ell,k} \mathbf{p}_{\ell,k}\right)^T + \mathbf{W}_j \tag{7}$$

where $\lambda_{\ell,k}^2$ and $\rho_{\ell,k}^2$ are the fractions of the transmit power reserved for the pilot and data symbols, respectively, and the total transmitted power $\mu_{\ell,k}$ is given as $\mu_{\ell,k} = \lambda_{\ell,k}^2 + \rho_{\ell,k}^2$.

A. Non-Iterative Channel Estimation

Treating the data symbols of all users as additive noise, the channel estimate of user k in cell ℓ can be obtained at the j'th BS using the LS criterion [26]

$$\widehat{\mathbf{h}}_{j,\ell,k} \triangleq \arg\min_{\mathbf{h}} \|\mathbf{Y}_j - \lambda_{\ell,k} \mathbf{h}\, \mathbf{p}_{\ell,k}^T\|_F^2. \tag{8}$$

---

[4]The algorithms and analysis in this paper remain the same in the presence of shadowing, provided that the users are allocated to the strongest BSs. However, the geometric interpretations that are made based on the location of the user in the cell will no longer be valid.

[5]No additional improvement in the UL performance can be gleaned by separating the pilot and data transmissions across cells [1], [9], [18].

[6]For example, using the orthogonal frequency division multiplexing (OFDM) parameters in long-term evolution (LTE) systems as in [1], i.e., $C_u = 7$ OFDM symbols, $N_{\mathrm{smooth}} = 14$ subcarriers, and assuming the pilots are reused over $L = 7$ hexagonal cells, the maximum number of supported users in the L cells is $C_u N_{\mathrm{smooth}} = 98$ users. Therefore, the number of users per cell is $C_u N_{\mathrm{smooth}}/L = 14$ users. However, note that the value of $C_u = 7$ has been chosen to allow user velocities of 350 km/h [27]. For lower user speeds and with cell sectoring, larger number users can be supported and the assumption $KL \leq C_u$ will easily be satisfied.

Solving (8) yields

$$
\widehat{\mathbf{h}}_{j,\ell,k} = \mathbf{Y}_j \left(\lambda_{\ell,k}^2 \, \mathbf{p}_{\ell,k}^H \mathbf{p}_{\ell,k} \right)^{-1} \lambda_{\ell,k} \mathbf{p}_{\ell,k}^* = \frac{1}{C_u \lambda_{\ell,k}} \mathbf{Y}_j \mathbf{p}_{\ell,k}^*
$$

$$
= \mathbf{h}_{j,\ell,k} + \frac{1}{C_u \lambda_{\ell,k}} \sum_{m=0}^{L-1} \sum_{n=0}^{K-1} \rho_{m,n} \mathbf{h}_{j,m,n} \mathbf{x}_{m,n}^T \mathbf{p}_{\ell,k}^*
$$

$$
+ \frac{1}{C_u \lambda_{\ell,k}} \mathbf{W}_j \mathbf{p}_{\ell,k}^*. \tag{9}
$$

In order to estimate the data from the received observations, it is necessary to remove the term corresponding to the transmitted superimposed pilot $\lambda_{j,m} \mathbf{h}_{j,j,m} \mathbf{p}_{j,m}^T$ from the observation vector in (7). Using $\lambda_{j,m} \widehat{\mathbf{h}}_{j,j,m} \mathbf{p}_{j,m}^T$ as an estimate for this term, the estimate of $\mathbf{x}_{j,m}$ can then be obtained from the observation $\mathbf{Y}_j$ using an MF and a decision operation as follows

$$
\widetilde{\mathbf{x}}_{j,j,m}^T = \frac{1}{M \rho_{j,m} \beta_{j,j,m}} \widehat{\mathbf{h}}_{j,j,m}^H \left(\mathbf{Y}_j - \lambda_{j,m} \widehat{\mathbf{h}}_{j,j,m} \mathbf{p}_{j,m}^T \right) \tag{10}
$$

$$
\widehat{\mathbf{x}}_{j,j,m} = \eta \left(\widetilde{\mathbf{x}}_{j,j,m} \right). \tag{11}
$$

The SINR of user m in cell j, at the output of an MF that employs the channel estimate in (9), is derived in Appendix A and is given in (12) (shown at the bottom of the page). The SINR in (12), when $M \to \infty$, can be written as

$$
\mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} = \frac{\lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,j,m}^2}{\frac{1}{C_u} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \mu_{\ell,k} \beta_{j,\ell,k}^2}. \tag{13}
$$

The corresponding per-user rate in the uplink when using Gaussian signaling is given as[7]

$$
R_{j,m}^{\mathrm{SP-ul}} = \frac{C_u}{C} \log_2 \left(1 + \mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} \right). \tag{14}
$$

B. Power Control and Choice of Parameters $\lambda_{j,m}$ and $\rho_{j,m}$

From (13), it can be seen that the SINR of a user is dependent on the product of the transmit powers and large-scale fading coefficients of the remaining $LK - 1$ users in addition to the product of its own transmit power and large-scale fading coefficient. This dependence results in a situation similar to the near-far problem in code division multiple access (CDMA) systems, wherein users that have larger values of large-scale fading coefficient β swamp users that have smaller values of β. Therefore, it becomes necessary to use power control to provide a uniform user experience.

[7]In calculating the SINR, we have neglected the correlation between the signal and interference components. Therefore, the calculated SINR and the corresponding rate are approximations.

While the parameters $\mu_{\ell,k}$, $\rho_{\ell,k}$, and $\lambda_{\ell,k}$ can be optimized by maximizing the sum-rate of all the users, i.e.,

$$
\max_{\mu_{\ell,k}, \rho_{\ell,k}, \lambda_{\ell,k}} \left\{ \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} R_{\ell,k}^{\mathrm{SP-ul}} \right\} \tag{15}
$$

the optimization problem is in general non-convex and requires coordination between the BSs. As an alternative, a suboptimal solution that does not involve coordination between the BSs is obtained here for the parameters $\mu_{\ell,k}$, $\rho_{\ell,k}$, and $\lambda_{\ell,k}$. This suboptimal solution will be shown to maximize a lower bound on the sum-rate, and it is as follows.

The received signal in (7) can be equivalently written as

$$
\mathbf{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \left(\frac{\rho_{\ell,k}}{\sqrt{\mu_{\ell,k}}} \mathbf{x}_{\ell,k}^T + \frac{\lambda_{\ell,k}}{\sqrt{\mu_{\ell,k}}} \mathbf{p}_{\ell,k}^T \right) + \mathbf{W}_j
$$

$$
= \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \bar{\mathbf{h}}_{j,\ell,k} \left(\bar{\rho}_{\ell,k} \mathbf{x}_{\ell,k} + \bar{\lambda}_{\ell,k} \mathbf{p}_{\ell,k} \right)^T + \mathbf{W}_j \tag{16}
$$

where

$$
\bar{\mathbf{h}}_{j,\ell,k} \triangleq \sqrt{\mu_{\ell,k}} \mathbf{h}_{j,\ell,k} \sim \mathcal{CN} \left(\mathbf{0}, \bar{\beta}_{j,\ell,k} \mathbf{I}_M \right) \tag{17}
$$

$$
\bar{\beta}_{j,\ell,k} \triangleq \beta_{j,\ell,k} \cdot \mu_{\ell,k} \tag{18}
$$

$$
\bar{\rho}_{\ell,k} \triangleq \sqrt{\frac{\rho_{\ell,k}^2}{\mu_{\ell,k}}} > 0 \tag{19}
$$

$$
\bar{\lambda}_{\ell,k} \triangleq \sqrt{\frac{\lambda_{\ell,k}^2}{\mu_{\ell,k}}} > 0 \tag{20}
$$

$$
\bar{\lambda}_{\ell,k}^2 + \bar{\rho}_{\ell,k}^2 = 1. \tag{21}
$$

From (16), it can be seen that a system having arbitrary values of $\beta_{j,\ell,k}$, $\mu_{\ell,k}$, $\rho_{\ell,k}$, and $\lambda_{\ell,k}$, can be reduced into an equivalent system with parameters $\bar{\beta}_{j,\ell,k}$, $\bar{\rho}_{\ell,k}$, and $\bar{\lambda}_{\ell,k}$, such that $0 \le \bar{\rho}_{\ell,k}, \bar{\lambda}_{\ell,k} \le 1$. Substituting (18) – (21) into (12), an equivalent expression for the SINR, as shown in (22) (shown at the bottom of the next page) can be obtained.

To obtain the parameter $\mu_{\ell,k}$, we propose using the statistics-aware power-control approach detailed in [18], wherein user m in cell j transmits at a power $\mu_{j,m} = \omega / \beta_{j,j,m}$ where ω is a design parameter. The parameter ω is chosen such that the transmitted power from a user satisfies a maximum power constraint, and users with severely low SINRs that would need a transmit power larger than this constraint would be denied service. This power control policy results in an identical received power of ω at the j'th BS for all the users in cell j. In addition, as mentioned in [18], the ratio $0 \le \beta_{j,\ell,k}/\beta_{\ell,\ell,k} \le 1$ is the relative strength of the interference received at BS j from a user in cell ℓ. This ratio is at most 1, when the user is at the edge of the j'th cell, and reduces to zero as its distance from BS j increases.

$$
\mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} = \left(\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\rho_{\ell,k}^2 \mu_{\ell,k} \beta_{j,\ell,k}^2}{C_u \lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,j,m}^2} + \frac{1}{M} \left(\sum_{\substack{\ell=0 \\ \{\ell \ne j, k \ne m\}}}^{L-1} \sum_{k=0}^{K-1} \frac{\beta_{j,\ell,k} \mu_{\ell,k}}{\rho_{j,m}^2 \beta_{j,j,m}} + \sum_{\substack{\ell=0 \\ \{\ell \ne j, k \ne m\}}}^{L-1} \sum_{k=0}^{K-1} \sum_{\substack{n=0 \\ \{n \ne \ell, p \ne k\}}}^{L-1} \sum_{p=0}^{K-1} \frac{\rho_{n,p}^2 \beta_{j,\ell,k} \beta_{j,n,p} \mu_{\ell,k}}{C_u \lambda_{j,m}^2 \rho_{j,m}^2 \beta_{j,j,m}^2} \right) \right)^{-1}
$$

$$
\tag{12}
$$

Therefore, setting $\mu_{\ell,k} = \omega/\beta_{\ell,\ell,k}$ and using the definitions of $\bar{\beta}_{j,\ell,k}$, $\bar{\rho}_{\ell,k}$, and $\bar{\lambda}_{\ell,k}$, and the inequality $0 \leq \beta_{j,\ell,k}/\beta_{\ell,\ell,k} \leq 1$, the following equations can be obtained

$$\bar{\beta}_{j,j,m} = \beta_{j,j,m} \cdot \mu_{j,m} = \omega \tag{23}$$

$$\bar{\beta}_{j,\ell,k} = \beta_{j,\ell,k} \cdot \mu_{\ell,k} \leq \beta_{\ell,\ell,k}\mu_{\ell,k} = \omega \quad \forall \ell \neq j \tag{24}$$

$$\rho_{j,m}^2 \beta_{j,j,m} = \bar{\rho}_{j,m}^2 \omega \tag{25}$$

$$\lambda_{j,m}^2 \beta_{j,j,m} = \bar{\lambda}_{j,m}^2 \omega \tag{26}$$

$$\rho_{\ell,k}^2 \beta_{j,\ell,k} \leq \rho_{\ell,k}^2 \beta_{\ell,\ell,k} = \bar{\rho}_{\ell,k}^2 \omega \quad \forall \ell \neq j \tag{27}$$

$$\lambda_{\ell,k}^2 \beta_{j,\ell,k} \leq \lambda_{\ell,k}^2 \beta_{\ell,\ell,k} = \bar{\lambda}_{\ell,k}^2 \omega \quad \forall \ell \neq j. \tag{28}$$

Substituting the above equations into (22), a lower bound on the SINR, as shown in (29) (shown at the bottom of the page), can be obtained.

However, the maximization of the lower bound on the SINR and hence, a lower bound on the sum rate, is still a non-convex problem in the parameters $\bar{\rho}_{\ell,k}$ and $\bar{\lambda}_{\ell,k}$ and requires coordination between the BSs. To circumvent this problem, we restrict the parameters $\bar{\rho}_{\ell,k}$ and $\bar{\lambda}_{\ell,k}$ such that $\bar{\rho}_{\ell,k} = \bar{\rho}, \forall \ell, k$ and $\bar{\lambda}_{\ell,k} = \bar{\lambda}, \forall \ell, k$. The choice of this restriction is motivated by the observation from (23) that the statistics-aware power control scheme results in the same large-scale path loss coefficient for all the desired users in the cell, irrespective of their locations. As a result, from the BS's perspective, each of its users are identical, and therefore, there is no benefit in assigning different values of $\bar{\rho}_{\ell,k}$ to different users. More importantly, such a restriction renders the choice of $\bar{\rho}_{\text{opt}}$ to depend only on L, K, C_u, and M as will be shown next. Setting $\bar{\rho}_{\ell,k} = \bar{\rho}, \forall \ell, k$ and $\bar{\lambda}_{\ell,k} = \bar{\lambda}, \forall \ell, k$ in (29), we obtain

$$\text{SINR}_{j,m}^{\text{SP-ul}} \geq \left(\frac{LK}{C_u (1 - \bar{\rho}^2)} + \frac{1}{M} \left(\frac{LK - 1}{\bar{\rho}^2} \right. \right.$$
$$\left. \left. + \frac{(LK-1)^2}{C_u (1 - \bar{\rho}^2)} \right) \right)^{-1}. \tag{30}$$

Differentiating the right hand side of (30) with respect to $\bar{\rho}^2$ and setting the resulting expression to zero, the value of $\bar{\rho}^2$ that maximizes the lower bound on $\text{SINR}_{j,m}^{\text{SP-ul}}$ and the UL sum rate can be obtained as

$$\bar{\rho}_{\text{opt}}^2 = \left(1 + \sqrt{\frac{\frac{LK}{C_u} + \frac{(LK-1)^2}{MC_u}}{\frac{LK-1}{M}}} \right)^{-1} \approx \left(1 + \sqrt{\frac{M + LK}{C_u}} \right)^{-1} \tag{31}$$

and the optimal value of $\bar{\lambda}^2$ can be obtained as

$$\bar{\lambda}_{\text{opt}}^2 = 1 - \bar{\rho}_{\text{opt}}^2 \approx \left(1 + \sqrt{\frac{C_u}{M + LK}} \right)^{-1} \tag{32}$$

where the approximations in (31) and (32) have been made assuming $LK \gg 1$ in order to obtain simpler expressions.

Based on the fact established in this subsection that systems using $\rho, \lambda, \beta, \mathbf{h}$, and $\bar{\rho}, \bar{\lambda}, \bar{\beta}, \bar{\mathbf{h}}$ are equivalent, we drop the overbar for ease of notation and adopt the former set of symbols in the rest of the paper. In addition, we set $\mu_{\ell,k} = \bar{\rho}_{\ell,k}^2 + \bar{\lambda}_{\ell,k}^2 = 1, \forall \ell, k$.

C. Impact of C_u on the Performance of Superimposed Pilots

Using (13) and a fixed set of parameters r, τ, and K, the following theorem presents an important condition that guarantees the superiority of methods based on superimposed pilots over the LS estimator that is based on time-multiplexed pilots.

Theorem 1: With fixed values of K, r, and τ and if $M \to \infty$, there exists a UL duration $\kappa_{j,m}$ beyond which a channel estimator based on superimposed pilots outperforms the LS based channel estimator that utilizes time-multiplexed pilots, in terms of the SINR performance, in any channel scenario $\{\beta_{j,\ell,m} \,|\, 0 \leq j, \ell \leq L - 1, \ 0 \leq m \leq K - 1\}$.

Proof: If $\kappa_{j,m}$ is defined as the number of symbols in the uplink such that (5) and (13) are equal, i.e.,

$$\frac{\beta_{j,j,m}^2}{\frac{1}{\kappa_{j,m} \lambda_{j,m}^2 \rho_{j,m}^2} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \beta_{j,\ell,k}^2} = \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,m}^2} \tag{33}$$

then it is evident from (13) and (33) that $C_u > \kappa_{j,m}$ is a sufficient condition for a method that is based on superimposed pilots to outperform the LS method that employs time-multiplexed pilots. In addition, $\kappa_{j,m}$ is given as

$$\kappa_{j,m} \triangleq \frac{\frac{1}{\lambda_{j,m}^2 \rho_{j,m}^2} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k}^2 \beta_{j,\ell,k}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,m}^2}. \tag{34}$$

This completes the proof. ∎

Remark 1: An important consequence of the above theorem is that in scenarios with negligible pilot contamination, the LS method based on superimposed pilots requires a large value of C_u to outperform the LS method based on time-multiplexed

$$\text{SINR}_{j,m}^{\text{SP-ul}} = \left(\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\bar{\rho}_{\ell,k}^2 \bar{\beta}_{j,\ell,k}^2}{C_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2 \bar{\beta}_{j,j,m}^2} + \frac{1}{M} \left(\sum_{\substack{\ell=0 \\ \{\ell \neq j, k \neq m\}}}^{L-1} \sum_{k=0}^{K-1} \frac{\bar{\beta}_{j,\ell,k}}{\bar{\rho}_{j,m}^2 \bar{\beta}_{j,j,m}} + \sum_{\substack{\ell=0 \\ \{\ell \neq j, k \neq m\}}}^{L-1} \sum_{k=0}^{K-1} \sum_{\substack{m=0 \\ \{m \neq \ell, n \neq k\}}}^{L-1} \sum_{n=0}^{K-1} \frac{\bar{\rho}_{m,n}^2 \bar{\beta}_{j,\ell,k} \bar{\beta}_{j,m,n}}{C_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2 \bar{\beta}_{j,j,m}^2} \right) \right)^{-1} \tag{22}$$

$$\text{SINR}_{j,m}^{\text{SP-ul}} \geq \left(\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\bar{\rho}_{\ell,k}^2}{C_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2} + \frac{LK-1}{M \bar{\rho}_{j,m}^2} + \sum_{\substack{\ell=0 \\ \{\ell \neq j, k \neq m\}}}^{L-1} \sum_{k=0}^{K-1} \sum_{\substack{m=0 \\ \{m \neq \ell, n \neq k\}}}^{L-1} \sum_{n=0}^{K-1} \frac{\bar{\rho}_{m,n}^2}{MC_u \bar{\lambda}_{j,m}^2 \bar{\rho}_{j,m}^2} \right)^{-1} \tag{29}$$

pilots. As an example, consider the case when $r = 1$, $\beta_{j,j,m} = 1$, $\forall m$, $\beta_{j,\ell,m} = \beta$, $\forall \ell \neq j, m$, and $\rho_{j,m}^2 = \lambda_{j,m}^2$, $\forall\, j, m$. For such a scenario, $\kappa_{j,m}$ is given as

$$\kappa_{j,m} = 2K \left(1 + \frac{1}{(L-1)\beta^2} \right). \tag{35}$$

Then, if the LS estimator based on superimposed pilots is required to maintain superiority over the LS estimator employing time-multiplexed pilots, C_u must scale inversely with β^2. This dependence on C_u is evident from the expression for the channel estimation error, which is given as

$$\Delta \mathbf{h}_{j,j,m} \triangleq \mathbf{h}_{j,j,m} - \widehat{\mathbf{h}}_{j,j,m} = -\frac{1}{C_u \lambda_{j,m}}$$
$$\times \left(\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \rho_{\ell,k} \mathbf{h}_{j,\ell,k} \mathbf{x}_{\ell,k}^T + \mathbf{W}_j \right) \mathbf{p}_{j,m}^*. \tag{36}$$

Remark 2: We build upon the discussion in [1] on grouping users based on their coherence times. While such a grouping does not offer any performance benefits to users when employing the approach in [1], the use of superimposed pilots offers low-mobility users an increase in throughput, by minimizing the channel estimation error resulting from transmitting the data alongside the pilots. This improvement in performance is a direct consequence of Theorem 1.

Remark 3: The type of pilot transmitted by a user can also be chosen based on the coherence time. While users with high-mobility or low pilot contamination would find it sufficient to use time-multiplexed pilots, users with low-mobility who suffer from significant pilot contamination due to their proximity to the cell-edge or due to shadowing would significantly benefit from employing superimposed pilots.

Remark 4: Superimposed pilots require coordination between BSs when assigning pilot sequences and synchronizing transmissions. In practical cellular networks, the cells are fairly large and it can be assumed that the interference is restricted to the first tier of cells and interference from the second and higher tiers of cells can be neglected. Therefore, it is reasonable to assume that practical deployments of superimposed pilots will require pilot assignment over only the first tier of cells, implying that coordination is limited to only this first tier. This overhead is not very different from that required by time-multiplexed pilots in the presence of pilot reuse. The coordination and synchronization requirements of superimposed pilot-based systems that allocate pilots over the first tier of cells are similar to that of time-multiplexed pilot-based systems that have a pilot reuse factor of $r = 3$ [18].

From (36), it can be seen that the error in the channel estimate includes interference resulting from transmitting data alongside the pilots. Hence, the quality of the channel estimate can be improved by eliminating the interference from the transmitted data through iterative data-aided schemes, thereby increasing the robustness of the proposed method with respect to C_u.

V. ITERATIVE DATA-AIDED CHANNEL ESTIMATION

In the iterative approach to channel estimation developed in this section, the estimated channel and data vectors of both the desired and interfering users are used in feedback in order to eliminate the first term in (36). In addition, to minimize error

propagation between the channel estimates of different users, the iteration is started from the user with the highest SINR and is progressed in the decreasing order of the SINRs of the users. It has to be noted that the objective of this section is to demonstrate that iterative methods for channel estimation with superimposed pilots provide a significantly better SINR performance than their non-iterative counterparts, and hence we restrict ourselves to a simple iterative algorithm. However, there is scope for developing improved iterative algorithms in the future.

A. Algorithm

For the sake of clarity and without loss of generality, we replace the two indices k, ℓ with a single index m that lies in the range $0 \leq m \leq N - 1$, where $N \triangleq KL$. The index m is used to index the users in all the L cells. In addition, we drop the index j and implicitly assume that the channel estimation is performed at the j'th BS. Then, (7) can be rewritten as

$$\mathbf{Y} = \sum_{m=0}^{N-1} \mathbf{h}_m \left(\rho_m \mathbf{x}_m + \lambda_m \mathbf{p}_m \right)^T + \mathbf{W}. \tag{37}$$

Since for large M, the SINRs of the users are proportional to the users' path-loss coefficients, the users are arranged in the decreasing order of their path-loss coefficients, i.e., $\beta_0 > \beta_1 > \ldots > \beta_{N-1}$.[8] Then, using an estimate of $\rho_m \mathbf{h}_m \mathbf{x}_m^T \mathbf{p}_m^*$ for each user as a correction factor to minimize the interference from other users, the corresponding channel estimate of user m can be written as

$$\widehat{\mathbf{h}}_m^{(i)} = \frac{1}{C_u \lambda_m} \left[\mathbf{Y} - \sum_{\substack{k=0 \\ k \in \mathcal{U}_m^{(i)}}}^{m-1} \rho_k \widehat{\mathbf{h}}_k^{(i)} \left(\widehat{\mathbf{x}}_k^{(i)} \right)^T \right.$$
$$\left. - \sum_{\substack{k=m \\ k \in \mathcal{U}_m^{(i)}}}^{N-1} \rho_k \widehat{\mathbf{h}}_k^{(i-1)} \left(\widehat{\mathbf{x}}_k^{(i-1)} \right)^T \right] \mathbf{p}_m^* \tag{38}$$

where $\widehat{\mathbf{h}}_m^{(0)} = \mathbf{0}$, $\forall m$ and $\mathcal{U}_m^{(i)}$ is the set of users whose estimated data is used in feedback in the i'th iteration to estimate the channel vector of user m. The approach to obtain $\mathcal{U}_m^{(i)}$ has been detailed in Appendix C, and involves selecting users such that the interference power, described in the next subsection, does not increase with each iteration. The channel estimate in the above equation is a modified version of the LS estimator defined in (9) with an added correction factor. Utilizing the resulting channel estimate in an MF and decision operation, similar to (10) and (11), the estimate of the data is obtained as follows

$$\left(\widetilde{\mathbf{x}}_m^{(i)} \right)^T = \frac{1}{M \rho_m \beta_m} \left(\widehat{\mathbf{h}}_m^{(i)} \right)^H \left(\mathbf{Y} - \lambda_m \widehat{\mathbf{h}}_m^{(i)} \mathbf{p}_m^T \right) \tag{39}$$

$$\widehat{\mathbf{x}}_m^{(i)} = \eta \left(\widetilde{\mathbf{x}}_m^{(i)} \right) \tag{40}$$

where $\widehat{\mathbf{x}}_m^{(0)} = \mathbf{0}$, $\forall m = 0, \ldots, N - 1$.

---

[8]It is assumed that the BSs have access to the exact values of the path-loss coefficients β_m and that there is no false-ordering. This assumption is reasonable since for large M, the path-loss coefficients can be computed at the BS with negligible error by averaging the power of the channel coefficients over the entire array.

Remark 5: If the matrix $\mathbf{P}$, whose columns are the superimposed pilots, is chosen as $\mathbf{P} = \text{blkdiag}\{\mathbf{P}_0, \ldots, \mathbf{P}_{L-1}\}$, where the ℓ'th block $\mathbf{P}_\ell \in \mathbb{C}^{K \times K}$ is comprised of the orthogonal pilot sequences used by the K users in cell ℓ, then the latency introduced when the non-iterative method is employed is the same as that for time-multiplexed pilots. However, when the iterative method is employed, the channel and the data vectors of the users are required and therefore, the uplink data in the entire slot will have to be aggregated before estimating the channel, which introduces a latency of C_u symbols.

Remark 6: From (9), the non-iterative method for channel estimation requires MC_u operations per user, whereas the MF and decision operations in (10) and (11) require M and C_u operations per user, respectively.

For the iterative method with ν iterations, the channel estimator, matched filter, and decision operations have a combined complexity of $O(\nu MC_u) + O(\nu M) + O(\nu C_u)$.

B. Interference Power at the BS

Let $\mathbf{e}_m^{(i)} \triangleq \mathbf{x}_m - \widetilde{\mathbf{x}}_m^{(i)}$ be the error in the estimate of the data symbols of user m obtained from the MF in the i'th iteration. Let $\Delta\mathbf{x}_m^{(i)} \triangleq \mathbf{x}_m - \widehat{\mathbf{x}}_m^{(i)}$ be the corresponding error vector after the decision operation and let $\Delta\mathbf{h}_m^{(i)} \triangleq \mathbf{h}_m - \widehat{\mathbf{h}}_m^{(i)}$ be the associated error in the channel estimate. If $\alpha_n^{(i)}$ is the variance of the elements of $\Delta\mathbf{x}_n^{(i)}$ and assuming that the elements of $\mathbf{e}_m^{(i)}$ are i.i.d. circular complex-Gaussian random variables with zero mean and variance $I_m^{(i)}$, an approximate expression for the interference power $I_m^{(i)}$ can be written as

$$I_m^{(i)} \approx \frac{1}{\beta_m^2}\left(\frac{1}{M\rho_m^2}\sum_{\substack{k=0\\k\neq m}}^{N-1}\beta_k\beta_m + \frac{\sigma^2\beta_m}{M\rho_m^2} + \frac{1}{M^2\rho_m^2}\psi_m^{(i)}\right) \tag{41}$$

where the expression for $\psi_m^{(i)}$ is given in (42) at the bottom of this page and $\psi_m^{(0)} = 0$, $\forall m$. The detailed derivation of $I_m^{(i)}$ can be found in Appendix B.

In deriving (41), the following simplifying assumptions have been made in order to obtain a closed form expression:

S1) $\mathbf{e}_m^{(i)}$ is independent of $\mathbf{x}_k$ and $\mathbf{W}$, $\forall k, i$.

S2) $\Delta\mathbf{x}_m^{(i)}$ is independent of $\mathbf{x}_k$, $\mathbf{W}$, and $\mathbf{h}_k$, $\forall k, i$.

S3) $\Delta\mathbf{x}_m^{(i)}$ is independent of $\Delta\mathbf{x}_k^{(p)}$, $\forall p \neq i, m \neq k$ and the elements of $\Delta\mathbf{x}_m^{(i)}$ are i.i.d.

S4) $\Delta\mathbf{h}_m^{(i)}$ is independent of $\mathbf{x}_k$, $\mathbf{W}$, and $\Delta\mathbf{x}_k^{(p)}$, $\forall k, p$.

In scenarios with low interference and with large M, only a few of the received symbols will be erroneous. As a result, the elements of $\Delta\mathbf{x}_m^{(i)}$ are sparse with the few non-zero elements restricted to locations that correspond to the erroneous symbols. Moreover, the vector $\mathbf{e}_m^{(i)}$ represents the error in the estimated data and in such low-interference scenarios, the elements of $\mathbf{e}_m^{(i)}$ take small values. Therefore, the simplifications (S1), (S2), and (S3) are reasonably accurate for these scenarios. Although the expression for $\Delta\mathbf{h}_m^{(i)}$, (given in (82) in Appendix B) is explicitly dependent on $\mathbf{x}_k$ and $\Delta\mathbf{x}_m^{(i)}$, we neglect the correlation between these terms since $\Delta\mathbf{h}_m^{(i)}$ is inversely proportional to C_u, and the simplification (S4) is fairly accurate when C_u is large with respect to N and when scenarios with low interference are considered. Since $\mathbf{e}_m^{(i)}$ is assumed to be a zero-mean random variable, $\Delta\mathbf{x}_k^{(i)}$ is also a zero-mean random variable, provided the constellation points in χ and their probability density functions are symmetric about the origin. This is true since by definition, $\Delta\mathbf{x}_k^{(i)}$ and $\mathbf{e}_m^{(i)}$ are related to each other through the following equation

$$\Delta\mathbf{x}_k^{(i)} = \mathbf{x}_k - \eta\left(\mathbf{x}_k - \mathbf{e}_m^{(i)}\right). \tag{43}$$

From (43), an expression for the variance of the elements of $\Delta\mathbf{x}_k^{(i)}$, i.e., $\alpha_k^{(i)}$ can be found as

$$\begin{aligned}\alpha_k^{(i)} &\triangleq \mathbb{E}\left\{\left|\left[\Delta\mathbf{x}_k^{(i)}\right]_n\right|^2\right\} = \int|\Delta x|^2\, p_{\Delta\mathbf{x}_k^{(i)}}(\Delta x)d\Delta x\\ &= \int_{x\in\chi}\int|x - \eta(x - e)|^2\, p_{\mathbf{e}_k^{(i)},\mathbf{x}_k}(e, x)\, de\, dx\\ &= \int_{x\in\chi}\int|x - \eta(x - e)|^2\, p_{\mathbf{e}_k^{(i)}}(e)\, p_{\mathbf{x}_k}(x)\, de\, dx \end{aligned} \tag{44}$$

where $p_{\mathbf{e}_k^{(i)}}(\cdot)$, $p_{\Delta\mathbf{x}_k^{(i)}}(\cdot)$, and $p_{\mathbf{x}_k}(\cdot)$ are the probability density functions of the elements of $\mathbf{e}_k^{(i)}$, $\Delta\mathbf{x}_k^{(i)}$, and $\mathbf{x}_k$, respectively, and $p_{\mathbf{e}_k^{(i)},\mathbf{x}_k}(\cdot)$ is the joint density function of the random variables $\mathbf{e}_k^{(i)}$ and $\mathbf{x}_k$. The latter has been written as the product of their individual distributions in the final expression of (44), thanks to (S1).

Important example of $\alpha_k^{(i)}$: When the elements of $\mathbf{x}_m$ are uniformly distributed and take values from a unit-power P-quaternary amplitude modulation (QAM) constellation, then under the assumption that the symbol errors in $\Delta\mathbf{x}_k^{(i)}$ are dominated by the closest neighboring symbols, the expression for

$$\psi_m^{(i)}\Big|_{i\geq 1} = \frac{M^2}{C_u\lambda_m^2}\left[\sum_{k\in\mathcal{U}_m^{(i)},k<m}\rho_k^2\left\{\beta_k^2\alpha_k^{(i)} + \frac{1}{M}\sum_{n=0}^{N-1}\beta_n\beta_k\alpha_k^{(i)} + \frac{\left(1+\alpha_k^{(i)}\right)}{M^2}\psi_k^{(i)}\right\} + \sum_{k\notin\mathcal{U}_m^{(i)}}\rho_k^2\left\{\beta_k^2 + \frac{1}{M}\sum_{n=0}^{N-1}\beta_n\beta_k\right\}\right.$$

$$\left. + \sum_{k\in\mathcal{U}_m^{(i)},m\leq k\leq N}\rho_k^2\left\{\beta_k^2\alpha_k^{(i-1)} + \sum_{n=0}^{N-1}\frac{1}{M}\beta_n\beta_k\alpha_k^{(i-1)} + \frac{\left(1+\alpha_k^{(i-1)}\right)}{M^2}\psi_k^{(i-1)}\right\} + \frac{\sigma^2}{M}\left(\sum_{n=0}^{N-1}\beta_n\right)\right]. \tag{42}$$
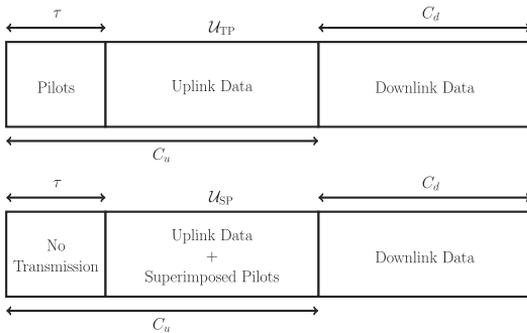
Fig. 1. Frame structure of a hybrid system with users employing time-multiplexed and superimposed pilots.

$\alpha_m^{(i)}$ can be written as

$$\alpha_m^{(i)} = \begin{cases} \dfrac{24}{\sqrt{P}\left(\sqrt{P}+1\right)} Q\left(\sqrt{\dfrac{\frac{3}{(P-1)}}{I_m^{(i)}}}\right), & i \geq 1 \\ 1, & i = 0 \end{cases} \quad (45)$$

where $Q\left(\cdot\right)$ is the Q-function. The detailed derivation of the above expression can be found in Appendix D.

VI. HYBRID SYSTEM

One of the main advantages of superimposed pilots over time-multiplexed pilots is that it does not require a separate set of symbols for pilot transmission. This property can be used to construct a hybrid system that contains two disjoint sets of users, with the users in one of the sets employing time-multiplexed pilots, and the users in the other set employing superimposed pilots. The following theorem shows that this hybrid system has a higher throughput and supports a larger number of users than a system that employs only time-multiplexed pilots.

Theorem 2: In a system that employs time-multiplexed pilots and is designed to maximize the UL and DL sum-rate,[9] 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 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 time-multiplexed and superimposed pilots, which is capable of supporting $C_u - \tau$ additional users and offers a higher sum-rate in the UL than the optimal system that employs only time-multiplexed pilots.

Proof: Consider the frame structure in Fig. 1 , wherein there are two sets of users $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$. The users in the set $\mathcal{U}_{\mathrm{TP}}$ employ time-multiplexed pilots, with parameters selected using approaches such as in [18]. The users in the set $\mathcal{U}_{\mathrm{SP}}$ maintain radio silence during the pilot training phase of the users in $\mathcal{U}_{\mathrm{TP}}$, i.e., for τ symbols in the frame, and transmit orthogonal pilots superimposed with data during the uplink data phase of $C_u - \tau$ symbols. Since these users maintain radio silence during the pilot training phase of τ symbols, they do not affect the quality of the channel estimates of the users in $\mathcal{U}_{\mathrm{TP}}$. As a result, under the assumption of asymptotic orthogonality of the channels,

there is no interference from the users in $\mathcal{U}_{\mathrm{SP}}$ to those in $\mathcal{U}_{\mathrm{TP}}$. Therefore, the per-cell sum-rate in the UL for the users in $\mathcal{U}_{\mathrm{TP}}$ remains unchanged and can be found from (6) to be

$$\mathrm{R}_j^{\mathrm{ul}}\left(\mathcal{U}_{\mathrm{TP}}\right) = \frac{\left(C_u - \tau\right)}{C} \sum_{\substack{k=0 \\ k \in \bar{\mathcal{U}}_{\mathrm{TP}}}}^{K-1} \log_2\left(1 + \frac{\beta_{j,j,k}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r)}} \beta_{j,\ell,k}^2}\right). \quad (46)$$

Assuming, for the sake of simplicity, that all the users in $\mathcal{U}_{\mathrm{SP}}$ are located in the j'th cell, the sum-rate of the users in $\mathcal{U}_{\mathrm{SP}}$ can be found using (13) and (14) as

$$\mathrm{R}_j^{\mathrm{ul}}\left(\mathcal{U}_{\mathrm{SP}}\right) = \frac{\left(C_u - \tau\right)}{C} \sum_{m \in \mathcal{U}_{\mathrm{SP}}} \log_2\left(1 + \mathrm{SINR}_m\left(\mathcal{U}_{\mathrm{SP}}\right)\right) \quad (47)$$

$$\mathrm{SINR}_m\left(\mathcal{U}_{\mathrm{SP}}\right) \triangleq \frac{\beta_{j,j,m}^2}{\sum_{k \in \mathcal{U}_{\mathrm{SP}}} \frac{\rho_{j,k}^2 \beta_{j,j,k}^2}{\left(C_u - \tau\right)\rho_{j,m}^2 \lambda_{j,m}^2}}. \quad (48)$$

In obtaining the above expression, it has been assumed that the transmit power p_u of the users in $\mathcal{U}_{\mathrm{TP}}$ is small enough such that the interference to the users in $\mathcal{U}_{\mathrm{SP}}$ can be neglected.[10] Therefore, from (46) and (48), the combined rate $\mathrm{R}_j^{\mathrm{ul}}\left(\mathcal{U}_{\mathrm{SP}}\right) + \mathrm{R}_j^{\mathrm{ul}}\left(\mathcal{U}_{\mathrm{TP}}\right)$ is strictly greater than $\mathrm{R}_j^{\mathrm{ul}}\left(\mathcal{U}_{\mathrm{TP}}\right)$. In addition, since the data slot is made up of $C_u - \tau$ symbols, it is possible to allocate $C_u - \tau$ orthogonal pilots and therefore, the set $\mathcal{U}_{\mathrm{SP}}$ can contain a maximum of $C_u - \tau$ users. This concludes the proof. ∎

In the above theorem, given a system with users employing time-multiplexed pilots, we have shown that additional users employing superimposed pilots can always be added to the system, resulting in a hybrid system that offers a higher throughput.

In the following section, we utilize the concept of the above theorem to partition a given set of users employing time-multiplexed pilots into two disjoint subsets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ that contain users transmitting time-multiplexed pilots and superimposed pilots, respectively. There are two main benefits of performing such a partition: (i) there is an overall improvement in the throughput as a result of the reduced inter-cell interference; and (ii) there is a reduction in the number of users that use time-multiplexed pilots, thereby allowing for more aggressive pilot reuse since r is a function of the number of users employing time-multiplexed pilots [18].

VII. A SIMPLE IMPLEMENTATION OF THE HYBRID SYSTEM

Given a set of K users per cell in L cells with channel gains $\beta_{j,\ell,k}, \forall j, \ell = 1, \ldots, L$, and $k = 1, \ldots, K$, the problem of partitioning users into disjoint sets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ can be accomplished by minimizing the overall UL inter-cell and intra-cell interference. This choice of objective function is motivated by Theorem 1, wherein it is observed that users at the cell edge cause significant pilot contamination and benefit from being assigned superimposed pilots, whereas users that are close to the BS cause negligible interference and could be assigned time-multiplexed pilots that are potentially shared with users in neighboring cells.

---

[9]Such as the scheme described in [18].

[10]This assumption is valid since the SINR and the rate of the users in $\mathcal{U}_{\mathrm{TP}}$ are independent of the transmit power p_u when $M \to \infty$. It has to be noted that this assumption has been made for the sake of simplicity and the theorem is valid even if this assumption does not hold.

A. Framework

If the users in $\mathcal{U}_{\mathrm{TP}}$ transmit pilots with unit power and data at a power p_u, then the received signal from the hybrid system in the UL phase at BS j can be written as

$$\mathbf{Y}_j = \mathbf{Y}_j^{\mathrm{TP}} + \mathbf{Y}_j^{\mathrm{SP}} + \mathbf{W}_j \tag{49}$$

where $\mathbf{Y}_j^{\mathrm{TP}}$ and $\mathbf{Y}_j^{\mathrm{SP}}$ are the received signals from the users in $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$, respectively. From Fig. 1, $\mathbf{Y}_j^{\mathrm{TP}}$ and $\mathbf{Y}_j^{\mathrm{SP}}$ can be written as

$$\mathbf{Y}_j^{\mathrm{TP}} \triangleq \sum_{\substack{\ell=0 \\ (\ell,k) \in \mathcal{U}_{\mathrm{TP}}}}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \left[\boldsymbol{\phi}_{\ell,k}^T, \sqrt{p_u} \mathbf{x}_{\ell,k}^T \right] \tag{50}$$

$$\mathbf{Y}_j^{\mathrm{SP}} \triangleq \sum_{\substack{\ell=0 \\ (\ell,k) \in \mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \left[\mathbf{0}_{1 \times \tau}, \rho \mathbf{x}_{\ell,k}^T + \lambda \mathbf{p}_{\ell,k}^T \right] \tag{51}$$

where the tuple (ℓ, k) is used to denote user k in cell l.

If user (j, m) is a member of $\mathcal{U}_{\mathrm{TP}}$, then the LS estimate of its channel can be written as [1]

$$\widehat{\mathbf{h}}_{j,j,m} = \frac{1}{\tau} \mathbf{Y}_j \mathbf{b}_{j,m}^{\mathrm{TP}} \mathbf{h}_{j,j,m} + \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r) \\ (\ell,m) \in \mathcal{U}_{\mathrm{TP}}}} \mathbf{h}_{j,\ell,m} + \frac{1}{\tau} \mathbf{W}_j \mathbf{b}_{j,m}^{\mathrm{TP}} \tag{52}$$

where $\mathbf{b}_{j,m}^{\mathrm{TP}} \triangleq \left[\boldsymbol{\phi}_{j,m}^H, \mathbf{0}_{(1 \times (C_u - \tau))} \right]^T$. If $M \gg K$, the SINR in the UL when using the channel estimate in (52) can be obtained similar to (5) as

$$\mathrm{SINR}_{j,m}^{\mathrm{TP-ul}} \approx \frac{\beta_{j,j,m}^2}{\sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r) \\ (\ell,m) \in \mathcal{U}_{\mathrm{TP}}}} \beta_{j,\ell,m}^2} \tag{53}$$

where the approximation in (53) is made for the sake of simplicity and is valid when M is sufficiently large.

If user (j, m) is a member of $\mathcal{U}_{\mathrm{SP}}$, then the LS estimate of its channel can be written as

$$\widehat{\mathbf{h}}_{j,j,m} = \frac{1}{(C_u - \tau)\lambda} \mathbf{Y}_j \mathbf{b}_{j,m}^{\mathrm{SP}}$$

$$= \mathbf{h}_{j,j,m} + \frac{\rho}{(C_u - \tau)\lambda} \sum_{\substack{\ell=0 \\ (\ell,k) \in \mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \mathbf{h}_{j,\ell,k} \mathbf{x}_{\ell,k}^T \mathbf{p}_{j,m}^*$$

$$+ \sqrt{p_u} \sum_{\substack{\ell=0 \\ (\ell,k) \in \mathcal{U}_{\mathrm{TP}}}}^{L-1} \sum_{k=0}^{K-1} \frac{\mathbf{h}_{j,\ell,k} \mathbf{x}_{\ell,k}^T \mathbf{p}_{j,m}^*}{(C_u - \tau)\lambda} + \mathbf{W}_j \mathbf{b}_{j,m}^{\mathrm{SP}} \tag{54}$$

where $\mathbf{b}_{j,m}^{\mathrm{SP}} \triangleq \left[\mathbf{0}_{(1 \times \tau)}, \mathbf{p}_{j,m}^H \right]^T$. Since it can be seen from (53) that the UL SINR of the users in $\mathcal{U}_{\mathrm{TP}}$ is independent of the UL transmit power p_u, we assume that p_u is small enough with respect to the transmit powers of the users in $\mathcal{U}_{\mathrm{SP}}$. As a result, the users in $\mathcal{U}_{\mathrm{SP}}$ do not experience significant interference during the data transmission phase of the users in $\mathcal{U}_{\mathrm{TP}}$ and result in the transmissions of $\mathcal{U}_{\mathrm{SP}}$ and $\mathcal{U}_{\mathrm{TP}}$ becoming independent of each

other.[11] Then (54) simplifies as

$$\widehat{\mathbf{h}}_{j,j,m} \approx \mathbf{h}_{j,j,m} + \frac{\rho}{\lambda} \sum_{\substack{\ell=0 \\ (\ell,k) \in \mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \frac{\mathbf{h}_{j,\ell,k} \mathbf{x}_{\ell,k}^T \mathbf{p}_{j,m}^*}{(C_u - \tau)} + \mathbf{W}_j \mathbf{b}_{j,m}^{\mathrm{SP}}. \tag{55}$$

Then the SINR in the UL for the users in $\mathcal{U}_{\mathrm{SP}}$ can be obtained from (13) as

$$\mathrm{SINR}_{j,m}^{\mathrm{SP-ul}} \approx \frac{\beta_{j,j,m}^2}{\frac{1}{(C_u - \tau)\lambda^2} \sum_{\substack{\ell=0 \\ (\ell,k) \in \mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \beta_{j,\ell,k}^2} \tag{56}$$

where, similar to (53), the approximation in (56) is made for the sake of simplicity and is valid when M is sufficiently large.

B. Algorithm to Obtain $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$.

The goal in this subsection is to obtain an algorithm for partitioning users into the sets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ by minimizing the total UL inter-cell and intra-cell interference. In order to accomplish this, we quantify the amount of interference caused by a user that is assigned to either of the sets $\mathcal{U}_{\mathrm{TP}}$ or $\mathcal{U}_{\mathrm{SP}}$.

Let $I_{j,m}^{\mathrm{TP-ul}}$ or $I_{j,m}^{\mathrm{SP-ul}}$ be the contributions of user (j, m) to the total UL inter/intra-cell interference power when assigned to $\mathcal{U}_{\mathrm{TP}}$ or $\mathcal{U}_{\mathrm{SP}}$, respectively. If users (j, m) and (ℓ, k) are members of $\mathcal{U}_{\mathrm{TP}}$, then from the denominator of (53), the amount of interference that user (j, m) causes to user (ℓ, k) in the UL is $\beta_{\ell,j,k}^2 \delta_{m,k}$. Likewise, from (56), if both users are members of $\mathcal{U}_{\mathrm{SP}}$, then the amount of interference that user (j, m) causes to user (ℓ, k) in the UL is $\beta_{\ell,j,m}^2 / \left((C_u - \tau)\lambda^2 \right)$. Therefore, $I_{j,m}^{\mathrm{TP-ul}}$ and $I_{j,m}^{\mathrm{SP-ul}}$ can be obtained as

$$I_{j,m}^{\mathrm{TP-ul}} = \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r) \\ (\ell,k) \in \mathcal{U}_{\mathrm{TP}}}}^{} \sum_{k=0}^{K-1} \beta_{\ell,j,k}^2 \delta_{m,k} = \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r) \\ (\ell,m) \in \mathcal{U}_{\mathrm{TP}}}} \beta_{\ell,j,m}^2 \tag{57}$$

$$I_{j,m}^{\mathrm{SP-ul}} = \frac{1}{(C_u - \tau)\lambda^2} \sum_{\substack{\ell=0 \\ (\ell,k) \in \mathcal{U}_{\mathrm{SP}}}}^{L-1} \sum_{k=0}^{K-1} \beta_{\ell,j,m}^2. \tag{58}$$

From the above equations, the total cost due to UL inter/intra-cell interference can be expressed as

$$I(\mathcal{U}_{\mathrm{TP}}, \mathcal{U}_{\mathrm{SP}}) = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \left(I_{\ell,k}^{\mathrm{TP-ul}} \mathbf{1}_{\{(\ell,k) \in \mathcal{U}_{\mathrm{TP}}\}} \right.$$

$$\left. + I_{\ell,k}^{\mathrm{SP-ul}} \mathbf{1}_{\{(\ell,k) \in \mathcal{U}_{\mathrm{SP}}\}} \right) \tag{59}$$

Using (59) as the objective function, the sets $\mathcal{U}_{\mathrm{TP}}$ and $\mathcal{U}_{\mathrm{SP}}$ can be obtained as the solution of the following optimization problem

$$(\mathcal{U}_{\mathrm{TP}}, \mathcal{U}_{\mathrm{SP}}) = \arg \min_{\substack{\mathcal{U}_{\mathrm{TP}} \subseteq \mathcal{U} \\ \mathcal{U}_{\mathrm{SP}} \subseteq \mathcal{U}}} I(\mathcal{U}_{\mathrm{TP}}, \mathcal{U}_{\mathrm{SP}})$$

$$\text{subject to} \quad \mathcal{U}_{\mathrm{TP}} \cup \mathcal{U}_{\mathrm{SP}} = \mathcal{U}$$

$$\mathcal{U}_{\mathrm{TP}} \cap \mathcal{U}_{\mathrm{SP}} = \varnothing \tag{60}$$

---

[11]This assumption is made for the sake of clarity and simplicity. In the absence of this assumption, the BS will have to estimate and remove $\mathbf{Y}_j^{\mathrm{TP}}$ from $\mathbf{Y}_j$ before estimating the channels of the users in $\mathcal{U}_{\mathrm{SP}}$.

Algorithm 1: Greedy algorithm to select $\mathcal{U}_{\text{TP}}$ and $\mathcal{U}_{\text{SP}}$.

Data: $\beta_{j,\ell,k}, \ \forall j, \ell = 0, \ldots, L-1, \ k = 0, \ldots, K-1$
Initialize: $\mathcal{U}_{\text{TP}} \leftarrow \mathcal{U}, \mathcal{U}_{\text{SP}} \leftarrow \varnothing$

1: Compute $\left(\tilde{\ell}, \tilde{k}\right)$ as in (61)
2: Set $\mathcal{U}'_{\text{TP}} \leftarrow \mathcal{U}_{\text{TP}} \setminus \left(\tilde{\ell}, \tilde{k}\right)$ and $\mathcal{U}'_{\text{SP}} \leftarrow \mathcal{U}_{\text{SP}} \cup \left(\tilde{\ell}, \tilde{k}\right)$
3: **If** $\mathcal{U}_{\text{TP}} \neq \varnothing$ **and if** $I\left(\mathcal{U}'_{\text{TP}}, \mathcal{U}'_{\text{SP}}\right) \leq I\left(\mathcal{U}_{\text{TP}}, \mathcal{U}_{\text{SP}}\right)$ **then**
4: $\quad \mathcal{U}_{\text{TP}} := \mathcal{U}'_{\text{TP}}, \mathcal{U}_{\text{SP}} := \mathcal{U}'_{\text{SP}}$
5: $\quad$ Return to Step (1).
6: **else**
7: $\quad$ STOP
8: **end if**

---

where $\mathcal{U}$ is the set of all users in the L cells. However, the optimization problem in (60) is combinatorial in nature with $2^{\text{Card}(\mathcal{U})}$ possible choices for $\mathcal{U}_{\text{TP}}$ and $\mathcal{U}_{\text{SP}}$, making it computationally hard to obtain the optimal solution. A workaround is to employ a greedy approach to partition $\mathcal{U}$ into $\mathcal{U}_{\text{TP}}$ and $\mathcal{U}_{\text{SP}}$. At each step of this algorithm, given $\mathcal{U}_{\text{TP}}$ and $\mathcal{U}_{\text{SP}}$, a user $\left(\tilde{\ell}, \tilde{k}\right)$ in $\mathcal{U}_{\text{TP}}$ is chosen as

$$\left(\tilde{\ell}, \tilde{k}\right) = \arg \max_{(\ell,k) \in \mathcal{U}_{\text{TP}}} I_{\ell,k}^{\text{TP}-\text{ul}}. \tag{61}$$

Setting $\mathcal{U}'_{\text{TP}} = \mathcal{U}_{\text{TP}} \setminus \left(\tilde{\ell}, \tilde{k}\right)$ and $\mathcal{U}'_{\text{SP}} = \mathcal{U}_{\text{SP}} \cup \left(\tilde{\ell}, \tilde{k}\right)$, user $\left(\tilde{\ell}, \tilde{k}\right)$ is added to $\mathcal{U}_{\text{SP}}$ if

$$I\left(\mathcal{U}'_{\text{TP}}, \mathcal{U}'_{\text{SP}}\right) \leq I\left(\mathcal{U}_{\text{TP}}, \mathcal{U}_{\text{SP}}\right). \tag{62}$$

The algorithm is initialized with $\mathcal{U}_{\text{TP}} = \mathcal{U}$ and is terminated when (62) is no longer satisfied or when $\mathcal{U}_{\text{TP}}$ is empty. The approach described above is summarized in Algorithm 1.

The complexity of the greedy algorithm used for designing the hybrid system can be obtained as follows. The terms $I^{\text{TP}-\text{ul}}$ and $I^{\text{SP}-\text{ul}}$ require a maximum of $\text{Card}(\mathcal{U})$ operations to compute, and therefore, computing $I(\mathcal{U}_{\text{TP}}, \mathcal{U}_{\text{SP}})$ requires $\text{Card}(\mathcal{U})^2$ operations. Assuming that the greedy algorithm runs till the condition $\mathcal{U}_{\text{TP}} = \varnothing$ is satisfied, then an upper bound on the computational complexity of the greedy algorithm is $\text{Card}(\mathcal{U})^3$ operations. Moreover, an overhead of $2\text{Card}(\mathcal{U})$ data transmissions is required for sending the large-scale path-loss coefficients to a central node and receiving the sets $\mathcal{U}_{\text{TP}}$ and $\mathcal{U}_{\text{SP}}$.

It has to be noted that Algorithm 1 is sub-optimal, but it is useful for illustrating the concept of the hybrid system. Partitioning algorithms that offer superior performance compared to Algorithm 1 with lower coordination overhead are left as topics for future research.

VIII. SIMULATION RESULTS

We compare the UL SINR and UL bit-error rate (BER) performance of the LS-based and (in some examples) eigenvalue decomposition (EVD)-based methods that use time-multiplexed pilots to the performance of the channel estimator that uses superimposed pilots, at the output of a MF that employs these channel estimates. Two scenarios are considered for this comparison.
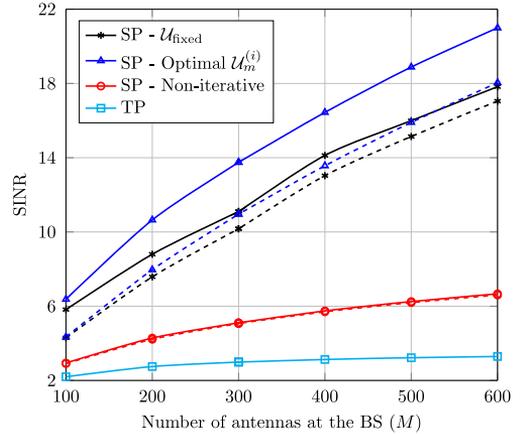
Fig. 2. The UL SINR of a user in the reference BS vs. M in Scenario 2. The values of ρ and λ are computed from (31) and (32), respectively, and since they are approximations, they result in a non-smooth SINR behavior for the iterative methods. The solid and dashed lines represent simulated and theoretical curves, respectively.

Scenario 1: The users are uniformly distributed in hexagonal cells of radius 1 km with the BS at the center. In addition, users are located at a distance of at least 100 m from the BS. *Scenario 2:* Users in both the reference and interfering cells are in a fixed configuration and are equally spaced on a circle of a given radius with the BS in the center. The size of the hexagonal cell is 1 km and unless otherwise specified, the users are on a circle of radius 800 m.

Unless otherwise specified, the following parameters are used in both scenarios. The channel estimation methods are tested with $L = 7$ cells and $K = 5$ users per cell. A P-QAM constellation is employed and the path-loss coefficient is assumed to be 3. The simulations for the superimposed pilots-based iterative channel estimation scheme have been performed for 4 iterations. The number of symbols in the uplink time slot C_u is set to 100, and for computing the rate, C is set to 200 symbols. The values of ρ and λ are computed from (31) and (32), respectively, and ω is set to 1, where ω is the design parameter in the statistics-aware power control scheme. The signal-to-noise ratio (SNR), i.e., ω/σ^2 is set to 10 dB. The methods based on time-multiplexed pilots have been simulated with $r = 1$ and $p_u = 1$. In addition, the chosen channel estimation methods have been observed to perform better with the statistics aware-power control scheme, and therefore, this power control scheme has been employed for both time-multiplexed and superimposed pilots. The plots in Scenario 1 are generated by averaging over 10^4 realizations of user locations across the cell. For each realization of user location, the channel vectors are generated and 200 bits are transmitted per user. The BER is computed by counting the bit errors for all the users in the reference cell. Similarly, the plots in Scenario 2 are generated for a fixed user location by averaging over 10^4 channel realizations with 200 bits transmitted per user for each realization.

Fig. 2 shows the variation of the UL SINR of an arbitrary user with respect to M in Scenario 2, whereas in Fig. 3, the approximate rate of an arbitrary user, calculated using 16-QAM constellation, is plotted for the same scenario. We compute the
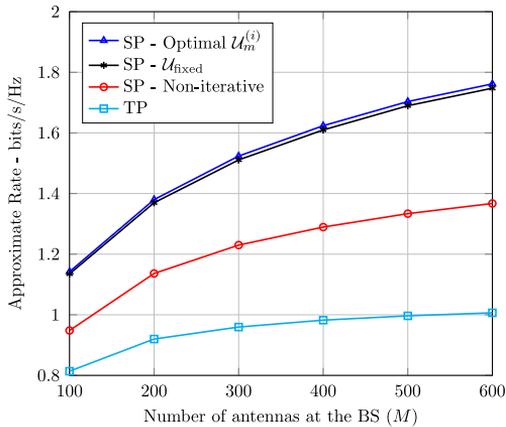
Fig. 3. Approximate per-user UL rate obtained using a 16-QAM constellation vs. M in the reference BS in Scenario 2. The maximum UL rate that can be achieved with the 16-QAM constellation, with half the symbols in a coherence block used for UL transmission, is 2 bps/Hz.
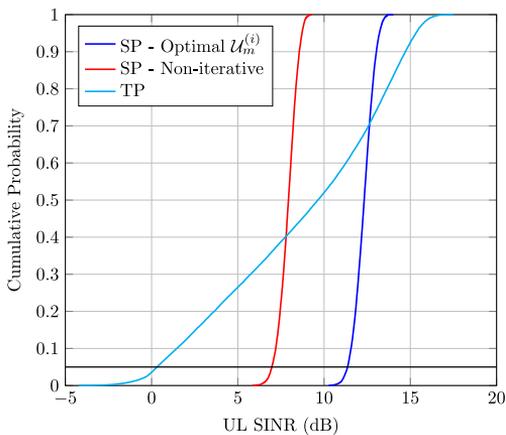
Fig. 4. Cumulative distribution of the UL SINR in dBs for users in Scenario 1 with $M = 300$ antennas. The black line indicates SINRs with probability ≥ 0.95.

achievable rate for 16-QAM signaling, modeling a practical scenario where highly mobile users are requesting moderate-to-high data rates. The SINR when the proposed method is employed, is seen to linearly increase in the number of antennas, whereas the SINR performance is observed to saturate for the LS-based method that uses time-multiplexed pilots. This trajectory of the proposed method could be potentially maintained using techniques such as adaptive modulation and coding, thereby implying that the effects of pilot contamination can be eliminated.

In Fig. 4, the cumulative distribution of the UL SINR in Scenario 1 is plotted. The interference power is averaged over 100 channel and data realizations for each realization of user location. While the LS-based method employing time-multiplexed pilots offers a higher SINR than the LS method employing superimposed pilots with a probability of approximately 0.6, the latter method can be seen to offer a significantly higher minimum SINR compared to the former method. Moreover, the
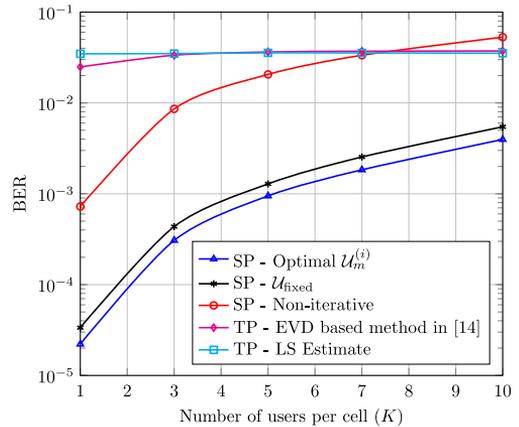
Fig. 5. BER in the UL vs. K in Scenario 1 with $M/K = 50$ and $C_u = 70$ symbols.

users employing superimposed pilots have a smaller variation in their SINR than those employing time-multiplexed pilots. This is because the SINR of a user when superimposed pilots are employed is limited by the interference from the other users in the same cell, and the statistics-aware power control scheme renders the intra-cell interference power independent of the user location within the cell. The iterative method based on superimposed pilots is observed to offer a remarkably higher SINR performance with respect to its non-iterative counterpart and the LS-based method employing time-multiplexed pilots.

In Fig. 5, the BER is plotted against the number of users per cell in Scenario 1, with K ranging from 1 to 10 and $C_u = 70$ symbols. Since $L = 7$ cells, $K = 10$ implies that the superimposed pilot-based system cannot support any new users without sharing pilots across cells. The ratio M/K is set to 50. While the non-iterative channel estimator based on superimposed pilots performs better in the UL at lower values of K than the estimators based on time-multiplexed pilots, the non-iterative estimator performs poorly at higher values of K. This is because the data transmitted alongside the pilots causes self-interference and this interference power increases with the number of users in the system. Therefore, it is necessary to resort to iterative techniques to mitigate this additional interference and it can be seen that the iterative methods offer a better performance than methods based on time-multiplexed pilots when LK is close to C_u.

In Fig. 6, the users are distributed as in Scenario 2 and the distance of the users from the BS is varied between 0.2 and 0.9 km. For the chosen range of user distance, the total rate in the UL is plotted against the corresponding received signal-to-interference ratio (SIR). The received SIR of an arbitrary user m in cell j is defined as

$$\mathrm{SIR}_j^{\mathrm{Rx}} \triangleq \frac{\omega}{\sum_{\ell \neq j} \sum_k \beta_{j,\ell,k}^2}. \qquad (63)$$

We assume $L = 19$ hexagonal cells, i.e., a central cell with two tiers of interfering cells. Each cell has $M = 1000$ antennas, $K = 5$ users, and the value of C_u is chosen as 40 symbols. Although L is set to 19, the optimization described in Algorithm 1 and the computation of the performance metrics is performed over
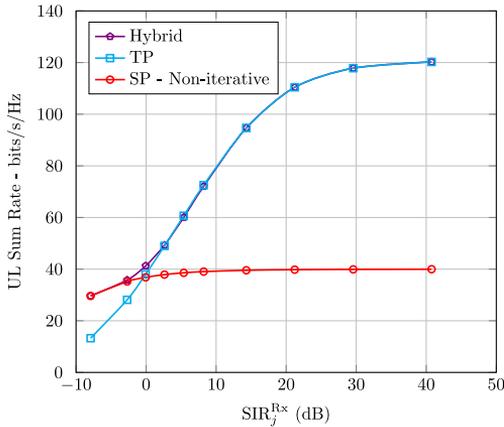
Fig. 6.   UL sum rate vs. $\mathrm{SIR}_j^{\mathrm{Rx}}$ in Scenario 2 with $M = 1000$ antennas.

7 cells which consist of the central and the first tier of cells. The value of ω for users in $\mathcal{U}_{\mathrm{SP}}$ is set to 10 and p_u for the users in $\mathcal{U}_{\mathrm{TP}}$ is set to 1. The data symbols are Gaussian distributed and the sum rate in Fig. 6 is obtained by averaging over 10^3 realizations of the channel and data symbols.

In Fig. 6, high and low values of SIR correspond to users located close to the BS and at the cell-edge, respectively. It can be observed that channel estimation methods based only on superimposed pilots (even the non-iterative formulation) are better in high interference scenarios, i.e., when the interfering users are at the cell-edge, whereas time-multiplexed pilots are better in low-interference scenarios. This behavior is a direct consequence of Theorem 1 since higher interference scenarios have smaller values of κ, resulting in superimposed pilots outperforming time-multiplexed pilots. However, at smaller values of user radius, the impact of pilot contamination is low but the self-interference in superimposed pilots resulting from transmitting the data alongside the pilots leads to a poorer performance compared to methods based on time-multiplexed pilots. In addition, it can be seen that the hybrid system adapts to the level of inter and intra-cell interference and offers a performance that is resilient to the location of the user within the cell.

IX. CONCLUSION AND DISCUSSION

We have proposed superimposed pilots as a superior alternative to time-multiplexed data and pilots for uplink channel estimation in massive MIMO. In the limit of an infinite number of antennas, a hybrid system using both superimposed pilots and time-multiplexed data and pilots offers a higher UL rate and supports larger number of users than the optimal system that utilizes only time-multiplexed data and pilots. The resilience to pilot contamination can be significantly improved with superimposed pilots through the use of an iterative data-aided channel estimation scheme that utilizes the data symbols of both the desired and interfering users in the feedback loop. Computer simulations in both a realistic scenario, in which users are distributed uniformly over the entire cell, and a high-interference scenario, in which users are concentrated at the cell edge, show that channel estimation methods using superimposed pilots offer a significant performance improvement over those that use time-multiplexed pilots.

The objective of this paper is to advocate superimposed pilots for practical use in massive MIMO systems by showing their superiority through theoretical and simulation based investigations. In standard MIMO communications, superimposed pilots are typically argued to be useful only for the scenario with high user mobility, and therefore, have not found practical application. On the contrary, in massive MIMO, superimposed pilots in a hybrid system provide superior performance in general. Therefore, there is a strong reason for superimposed pilots to make their way to practical use.

The proposed iterative data-aided channel estimation scheme and the greedy algorithm for partitioning users are suboptimal algorithms for corresponding non-convex problems. Algorithms that offer performance close to the optimal at low computational complexities and overheads are of interest for future research. Moreover, the downlink performance of superimposed pilots is another topic of practical importance, which we have partially addressed in [28].

APPENDIX A

Uplink SINR of the Non-Iterative Channel Estimation Method

Using the notation described in Section V-A, (7) can be rewritten as

$$\mathbf{Y} = \sum_{m=0}^{N-1} \mathbf{h}_m \left(\rho_m \mathbf{x}_m + \lambda_m \mathbf{p}_m \right)^T + \mathbf{W} \qquad (64)$$

From (9), the estimation error of the channel estimate can be obtained as

$$\Delta \mathbf{h}_m \triangleq \mathbf{h}_m - \widehat{\mathbf{h}}_m = -\frac{1}{C_u \lambda_m} \left(\sum_{k=0}^{N-1} \rho_k \mathbf{h}_k \mathbf{x}_k^T + \mathbf{W} \right) \mathbf{p}_m^*. \qquad (65)$$

From (10) and (65), the estimate of the received data after MF with the estimated channel can be written as

$$
\begin{aligned}
\widetilde{\mathbf{x}}_m^T &= \frac{1}{M \rho_m \beta_m} \widehat{\mathbf{h}}_m^H \left(\mathbf{Y} - \lambda_m \widehat{\mathbf{h}}_m \mathbf{p}_m^T \right) \\
&= \frac{1}{M \rho_m \beta_m} \left(\mathbf{h}_m^H - \Delta \mathbf{h}_m^H \right) \\
&\quad \times \left(\sum_{k=0}^{N-1} \mathbf{h}_k \left(\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k \right)^T + \mathbf{W} \right. \\
&\quad \left. - \lambda_m \left(\mathbf{h}_m - \Delta \mathbf{h}_m \right) \mathbf{p}_m^T \right) = \mathbf{g}^T + \mathbf{i}^T \qquad (66)
\end{aligned}
$$

where $\mathbf{g}$ and $\mathbf{i}$ are the signal and interference components of the matched filtered signal, respectively, which can be written as

$$\mathbf{g} \triangleq \frac{\|\mathbf{h}_m\|^2}{M \beta_m} \mathbf{x}_m \qquad (67)$$

$$\mathbf{i} \triangleq \sum_{n=1}^{5} \mathbf{i}_n \qquad (68)$$

$$\mathbf{i}_1 \triangleq \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \frac{\mathbf{h}_m^H \mathbf{h}_n}{M \rho_m \beta_m} \left(\lambda_n \mathbf{p}_n + \rho_n \mathbf{x}_n \right) + \frac{\left(\mathbf{h}_m^H \mathbf{W} \right)^T}{M \rho_m \beta_m} \qquad (69)$$

$$\mathbf{i}_2 \triangleq \frac{\lambda_m}{M \rho_m \beta_m} \mathbf{h}_m^H \Delta \mathbf{h}_m \mathbf{p}_m \qquad (70)$$

$$\mathbf{i}_3 \triangleq -\frac{1}{M\beta_m} \Delta\mathbf{h}_m^H \mathbf{h}_m \mathbf{x}_m \tag{71}$$

$$\mathbf{i}_4 \triangleq -\frac{1}{M\rho_m\beta_m} \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \Delta\mathbf{h}_m^H \mathbf{h}_n \left(\lambda_n \mathbf{p}_n + \rho_n \mathbf{x}_n\right) - \frac{\left(\Delta\mathbf{h}_m^H \mathbf{W}\right)^T}{M\rho_m\beta_m} \tag{72}$$

$$\mathbf{i}_5 \triangleq -\frac{\lambda_m}{M\rho_m\beta_m} \|\Delta\mathbf{h}_m\|^2 \mathbf{p}_m. \tag{73}$$

The average interference power can be found as

$$\mathbb{E}\left\{\|\mathbf{i}\|^2\right\} = \mathbb{E}\left\{\left\|\sum_{n=1}^{5} \mathbf{i}_n\right\|^2\right\}. \tag{74}$$

Then, using the definitions of $\mathbf{i}_n, \forall n$ in $(69) - (73)$ and the definition of $\Delta\mathbf{h}_m$ in (65), the following expressions can be easily obtained

$$\mathbb{E}\left\{\|\mathbf{i}_1\|^2\right\} \approx \frac{C_u}{M\rho_m^2\beta_m} \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \beta_n \mu_n \tag{75}$$

$$\mathbb{E}\left\{\|\mathbf{i}_2\|^2 + \|\mathbf{i}_3\|^2 + \|\mathbf{i}_4\|^2\right\} \approx \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \sum_{\substack{k=0 \\ k\neq n}}^{N-1} \frac{\rho_k^2 \beta_n \beta_k \mu_n}{M\lambda_m^2 \rho_m^2 \beta_m^2}$$
$$+ \sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{\lambda_m^2 \rho_m^2 \beta_m^2} + \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \sum_{\substack{p=0 \\ p\neq m \\ n\neq p}}^{N-1} \frac{\rho_n^2 \rho_p^2 \beta_p \beta_n}{C_u \rho_m^2 \lambda_m^2 \beta_m^2} \tag{76}$$

$$\mathbb{E}\left\{\|\mathbf{i}_5\|^2\right\} \approx \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \sum_{\substack{p=0 \\ p\neq m \\ n\neq p}}^{N-1} \frac{\rho_n^2 \rho_p^2 \beta_n \beta_p}{C_u \lambda_m^2 \rho_m^2 \beta_m^2} \tag{77}$$

$$\mathbb{E}\left\{\left(\mathbf{i}_3^H + \mathbf{i}_4^H\right)\mathbf{i}_5\right\} \approx -\sum_{\substack{p=0 \\ p\neq m \\ n\neq p}}^{N-1} \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \frac{\rho_n^2 \rho_p^2 \beta_n \beta_p}{C_u \lambda_m^2 \rho_m^2 \beta_m^2} \tag{78}$$

where the approximation errors in $(75) - (78)$ are proportional to either N/M, N/C_u, or C_u/M. In addition, the remaining terms of the form $\mathbf{i}_n^H \mathbf{i}_p, \forall\, n \neq p$ in the expansion of (74) are proportional to N/M or N/C_u. If M is large with respect to N and C_u, then the approximation errors and terms that are proportional to N/M and N/C_u can be neglected. Similarly, error terms that are proportional to N/C_u can also be dropped, and if $\sigma^2 \ll C_u$, then the effect of noise can also be neglected. Then, substituting $(75) - (78)$ into the expansion of (74), the interference power is obtained as

$$\mathbb{E}\left\{\|\mathbf{i}\|^2\right\} \approx \sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{\lambda_m^2 \rho_m^2 \beta_m^2} + \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \frac{C_u \beta_n \mu_n}{M\rho_m^2 \beta_m}$$
$$+ \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \sum_{\substack{k=0 \\ k\neq n}}^{N-1} \frac{\rho_k^2 \beta_n \beta_k \mu_n}{M\lambda_m^2 \rho_m^2 \beta_m^2}. \tag{79}$$

Using (79), the SINR can be obtained as

$$\mathrm{SINR}_m^{\mathrm{SP-ul}} \triangleq \frac{\mathbb{E}\left\{\|\mathbf{g}\|^2\right\}}{\mathbb{E}\left\{\|\mathbf{i}\|^2\right\}}$$

$$= \frac{C_u}{\displaystyle\sum_{n=0}^{N-1} \frac{\rho_n^2 \mu_n \beta_n^2}{\lambda_m^2 \rho_m^2 \beta_m^2} + \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \frac{C_u \beta_n \mu_n}{M\rho_m^2 \beta_m} + \sum_{\substack{n=0 \\ n\neq m}}^{N-1} \sum_{\substack{k=0 \\ k\neq n}}^{N-1} \frac{\rho_k^2 \beta_n \beta_k \mu_n}{M\lambda_m^2 \rho_m^2 \beta_m^2}}. \tag{80}$$

It completes the derivation of (12).

APPENDIX B
Interference Power of the Iterative Method

To derive the SINR, using the definition of $\Delta\mathbf{x}_m^{(i)} \triangleq \mathbf{x}_m - \widehat{\mathbf{x}}_m^{(i)}$, the channel estimate in (38) can be simplified as

$$\widehat{\mathbf{h}}_m^{(i)} = \mathbf{h}_m + \frac{1}{C_u\lambda_m}\left(\sum_k \rho_k \mathbf{h}_k \mathbf{x}_k^T - \sum_{k\in\mathcal{U}_j, k<m} \rho_k \widehat{\mathbf{h}}_k^{(i)}\left(\widehat{\mathbf{x}}_k^{(i)}\right)^T\right.$$
$$\left. - \sum_{k\in\mathcal{U}_j, m\leq k\leq N} \rho_k \widehat{\mathbf{h}}_k^{(i-1)}\left(\widehat{\mathbf{x}}_k^{(i-1)}\right)^T + \mathbf{W}_j\right)\mathbf{p}_m^* \tag{81}$$

where

$$\Delta\mathbf{h}_m^{(i)} = -\frac{1}{C_u\lambda_m}\left(\sum_{k\in\mathcal{U}_j, k<m} \rho_k\left\{\mathbf{h}_k\left(\Delta\mathbf{x}_k^{(i)}\right)^T + \Delta\mathbf{h}_k^{(i)}\mathbf{x}_k^T\right.\right.$$
$$\left. - \Delta\mathbf{h}_k^{(i)}\left(\Delta\mathbf{x}_k^{(i)}\right)^T\right\} + \sum_{k\in\mathcal{U}_j, m\leq k\leq N} \rho_k\left\{\mathbf{h}_k\left(\Delta\mathbf{x}_k^{(i-1)}\right)^T\right.$$
$$\left. + \Delta\mathbf{h}_k^{(i-1)}\mathbf{x}_k^T - \Delta\mathbf{h}_k^{(i-1)}\left(\Delta\mathbf{x}_k^{(i-1)}\right)^T\right\}$$
$$\left. + \sum_{k\notin\mathcal{U}_j} \rho_k \mathbf{h}_k \mathbf{x}_k^T + \mathbf{W}\right)\mathbf{p}_m^*. \tag{82}$$

The received symbols after MF in (39) are then given as

$$\widehat{\mathbf{x}}_m^T = \frac{1}{M\rho_m}\left(\mathbf{h}_m^H - \left(\Delta\mathbf{h}_m^{(i)}\right)^H\right)\left(\sum_{k=0}^{N-1} \mathbf{h}_k\left(\rho_k\mathbf{x}_k + \lambda_k\mathbf{p}_k\right)^T\right.$$
$$\left. + \mathbf{W} - \lambda_m\left(\mathbf{h}_m - \Delta\mathbf{h}_m^{(i)}\right)\mathbf{p}_m^T\right)$$
$$= \frac{1}{M}\|\mathbf{h}_m\|^2 \mathbf{x}_m^T + \sum_{k=1}^{7} \mathbf{a}_k^T \tag{83}$$

where

$$\mathbf{a}_1 \triangleq \frac{1}{M\rho_m} \sum_{\substack{k=0 \\ k\neq m}}^{N-1} \mathbf{h}_m^H \mathbf{h}_k \left(\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k\right) \tag{84}$$

$$\mathbf{a}_2 \triangleq \frac{1}{M\rho_m} \left(\mathbf{h}_m^H \mathbf{W}\right)^T \tag{85}$$

$$\mathbf{a}_3 \triangleq \frac{\lambda_m}{M\rho_m} \mathbf{h}_m^H \Delta\mathbf{h}_m^{(i)} \mathbf{p}_m \tag{86}$$

$$\mathbf{a}_4 \triangleq -\frac{1}{M} \left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_m \mathbf{x}_m \tag{87}$$

$$\mathbf{a}_5 \triangleq -\frac{1}{M\rho_m} \sum_{\substack{k=0 \\ k\neq m}}^{N-1} \left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_k \left(\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k\right) \tag{88}$$

$$\mathbf{a}_6 \triangleq -\frac{1}{M\rho_m} \left(\left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{W}\right)^T \tag{89}$$

$$\mathbf{a}_7 \triangleq -\frac{\lambda_m}{M\rho_m} \left\|\Delta\mathbf{h}_m^{(i)}\right\|^2 \mathbf{p}_m. \tag{90}$$

Under the assumption that the interference power at each of the received symbols is the same, the average interference power of the m'th user at the j'th cell is given as

$$I_m^{(i)} = \frac{1}{C_u} \mathbb{E}\left\{\left\|\sum_{k=1}^{7} \mathbf{a}_k\right\|^2\right\} \approx \frac{1}{C_u} \left[\mathbb{E}\left\{\sum_{k=1}^{5} \|\mathbf{a}_k\|^2\right\}\right] \tag{91}$$

where the terms $\mathbf{a}_6$, $\mathbf{a}_7$, and $\mathbf{a}_p^H \mathbf{a}_q$, $\forall p, q$ have been dropped. Further, it can be shown straightforwardly that

$$\mathbb{E}\left\{\|\mathbf{a}_1\|^2\right\} = \frac{C_u}{M\rho_m^2} \sum_{\substack{k=0 \\ k\neq m}}^{N-1} \beta_k \beta_m \tag{92}$$

$$\mathbb{E}\left\{\|\mathbf{a}_2\|^2\right\} = \frac{C_u \sigma^2 \beta_m}{M\rho_m^2}. \tag{93}$$

Moreover, $\mathbb{E}\left\{\|\mathbf{a}_3\|^2\right\}$, $\mathbb{E}\left\{\|\mathbf{a}_4\|^2\right\}$, and $\mathbb{E}\left\{\|\mathbf{a}_5\|^2\right\}$ can be written as

$$\mathbb{E}\left\{\|\mathbf{a}_3\|^2\right\} = \frac{\lambda_m^2}{M^2\rho_m^2} \mathbb{E}\left\{\mathbf{h}_m^H \Delta\mathbf{h}_m^{(i)} \mathbf{p}_m^T \mathbf{p}_m^* \left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_m\right\}$$

$$= \frac{C_u \lambda_m^2}{M^2\rho_m^2} \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_m \mathbf{h}_m^H \Delta\mathbf{h}_m^{(i)}\right\} \tag{94}$$

$$\mathbb{E}\left\{\|\mathbf{a}_4\|^2\right\} = \frac{1}{M^2} \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_m \mathbf{x}_m^T \mathbf{x}_m^* \mathbf{h}_m^H \Delta\mathbf{h}_m^{(i)}\right\}$$

$$= \frac{1}{M^2} \mathbb{E}\left\{\mathbf{x}_m^T \mathbf{x}_m^*\right\} \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_m \mathbf{h}_m^H \Delta\mathbf{h}_m^{(i)}\right\}$$

$$= \frac{C_u}{M^2} \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_m \mathbf{h}_m^H \Delta\mathbf{h}_m^{(i)}\right\} \tag{95}$$

and

$$\mathbb{E}\left\{\|\mathbf{a}_5\|^2\right\} = \frac{1}{M^2\rho_m^2} \sum_{\substack{\ell=0 \\ \ell\neq m}}^{N-1} \sum_{\substack{k=0 \\ k\neq m}}^{N-1} \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \mathbf{h}_\ell \mathbf{h}_k^H \Delta\mathbf{h}_m^{(i)}\right\}$$

$$\times \mathbb{E}\left\{\left(\rho_\ell \mathbf{x}_\ell + \lambda_\ell \mathbf{p}_\ell\right)^H \left(\rho_k \mathbf{x}_k + \lambda_k \mathbf{p}_k\right)\right\}$$

$$= \frac{C_u}{M^2\rho_m^2} \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \left(\sum_{\substack{k=0 \\ k\neq m}}^{N-1} \mathbf{h}_k \mathbf{h}_k^H\right) \Delta\mathbf{h}_m^{(i)}\right\}. \tag{96}$$

Summing up (94), (95), and (96), we obtain

$$\mathbb{E}\left\{\sum_{k=3}^{5} \|\mathbf{a}_k\|^2\right\} = \frac{C_u}{M^2\rho_m^2}$$

$$\times \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \left(\sum_{k=0}^{N-1} \mathbf{h}_k \mathbf{h}_k^H\right) \Delta\mathbf{h}_m^{(i)}\right\}. \tag{97}$$

Now, let $\psi_m^{(i)}$ be defined as the second term in (97), i.e.,

$$\psi_m^{(i)}\big|_{i\geq 1} \triangleq \mathbb{E}\left\{\left(\Delta\mathbf{h}_m^{(i)}\right)^H \left(\sum_{n=0}^{N-1} \mathbf{h}_n \mathbf{h}_n^H\right) \Delta\mathbf{h}_m^{(i)}\right\}. \tag{98}$$

Using (82) and the simplifications (S1) to (S4), (98) can be simplified to obtain (42). Substituting (92), (93), (97), and (42) into (91), $I_m^{(i)}$ can be obtained as

$$I_m^{(i)} \approx \frac{1}{M\rho_m^2} \sum_{\substack{k=0 \\ k\neq m}}^{N-1} \beta_k \beta_m + \frac{\sigma^2 \beta_m}{M\rho_m^2} + \frac{1}{M^2\rho_m^2} \psi_m^{(i)}. \tag{99}$$

It completes the derivation of (41).

APPENDIX C
Choice of the Set of Users $\mathcal{U}_m^{(i)}$

Let $\mathcal{S}$ be a set of the KL users in the system and let $\mathcal{P}(\mathcal{S})$ be its power set. In addition, for the sake of clarity, let the additional argument $\mathcal{U}_m^{(i)}$ be added to the functions $I_m^{(i)}$ and $\psi_m^{(i)}$ in this section. Now, the optimal set $\mathcal{U}_m^{(i)}$ can be obtained by solving the following optimization problem

$$\mathcal{U}_m^{(i)} = \arg\min_{\mathcal{U}\in\mathcal{P}(\mathcal{S})} \left\{I_m^{(i)}(\mathcal{U})\right\}. \tag{100}$$

Substituting (41) into (100) yields

$$\mathcal{U}_m^{(i)} = \arg\min_{\mathcal{U}\in\mathcal{P}(\mathcal{S})} \left\{\psi_m^{(i)}(\mathcal{U})\right\}. \tag{101}$$

Now, $\psi_m^{(i)}(\mathcal{U})$ can be rewritten as

$$\psi_m^{(i)}(\mathcal{U}) = c + \sum_{n=0}^{N-1} \left\{\xi_n \mathbf{1}_{\{n\notin\mathcal{U}\}} + \epsilon_n^{(i)}(\mathcal{U})\mathbf{1}_{\{n\in\mathcal{U},n<m\}}\right.$$

$$\left. + \epsilon_n^{(i-1)}(\mathcal{U})\mathbf{1}_{\{n\in\mathcal{U},n\geq m\}}\right\} \tag{102}$$

where c, ξ_n, and $\epsilon_n^{(i)}(\mathcal{U})$ are defined as

$$c \triangleq \frac{M\sigma^2}{C_u\lambda_m^2}\left(\sum_{k=0}^{N-1}\beta_k\right) \tag{103}$$

$$\xi_n \triangleq \frac{M^2\rho_n^2}{C_u\lambda_m^2}\left\{\beta_n^2 + \frac{1}{M}\sum_{k=0}^{N-1}\beta_k\beta_n\right\} \tag{104}$$

$$\epsilon_n^{(i)}(\mathcal{U}) \triangleq \frac{M^2\rho_n^2}{C_u\lambda_m^2}\left\{\beta_n^2\alpha_n^{(i)} + \sum_{k=0}^{N-1}\frac{1}{M}\beta_k\beta_n\alpha_n^{(i)}\right. $$
$$\left. + \frac{\left(1+\alpha_n^{(i)}\right)}{M^2}\psi_n^{(i)}(\mathcal{U})\right\}. \tag{105}$$

It can be seen from (102) that the optimization problem (101) is separable over the user indices, implying that the decision to include user n in $\mathcal{U}_m^{(i)}$ is independent of the other $N-1$ users. Therefore, the channel and data estimates of user n are used in the i'th iteration if the following condition is satisfied

$$n \in \mathcal{U}_m^{(i)} \quad \text{iff} \quad \psi_m^{(i)}\big|_{n\in\mathcal{U}_m^{(i)}} < \psi_m^{(i)}\big|_{n\notin\mathcal{U}_m^{(i)}}. \tag{106}$$

From (102) and (106), the set $\mathcal{U}_m^{(i)}$ is obtained as

$$\mathcal{U}_m^{(i)} = \left\{n \in \mathbb{N}\,\middle|\,\epsilon_n^{(i)}(\mathcal{U}) < \xi_n \text{ when } n < m\right.$$
$$\left. \text{and } \epsilon_n^{(i-1)}(\mathcal{U}) < \xi_n \text{ when } n \geq m\right\}. \tag{107}$$

Equivalently, using (104) and (105), the above expression simplifies to

$$\mathcal{U}_m^{(i)} = \left\{n \in \mathbb{N}\,\middle|\,\alpha_n^{(i)} < \gamma_n^{(i)} \text{ when } n < m\right.$$
$$\left. \text{and } \alpha_n^{(i-1)} < \gamma_n^{(i-1)} \text{ when } n \geq m\right\} \tag{108}$$

where

$$\gamma_n^{(i)} \triangleq \frac{\left\{\beta_n^2 + \frac{1}{M}\sum_{k=0}^{N-1}\beta_n\beta_k - \frac{\psi_n^{(i)}\big|_{n\in\mathcal{U}_m^{(i)}}}{M^2}\right\}}{\left\{\beta_n^2 + \frac{1}{M}\sum_{k=0}^{N-1}\beta_n\beta_k + \frac{\psi_n^{(i)}\big|_{n\in\mathcal{U}_m^{(i)}}}{M^2}\right\}}. \tag{109}$$

If $\mathbf{x}_m$ takes values from the P-QAM constellation, then substituting (45) into (108), the set $\mathcal{U}_m^{(i)}$ can be obtained as

$$\mathcal{U}_m^{(i)} = \left\{n \in \mathbb{N}\,\middle|\,I_n^{(i)} < f_n^{(i)} \text{ when } n < m\right.$$
$$\left. \text{and } I_n^{(i)} < f_n^{(i-1)} \text{ when } n \geq m\right\} \tag{110}$$

where $f_n^{(i)}$ is defined as

$$f_n^{(i)} \triangleq \frac{3}{P-1}Q^2\left(\frac{\sqrt{P}\left(\sqrt{P}+1\right)\gamma_n^{(i)}}{24}\right). \tag{111}$$

However, since the decision rules are based on approximate SINR expressions, it is worth commenting that the reliability of the decision rule in (108) decreases with increasing user and iteration indices. Alternatively, a fixed and conservative decision rule can be used to obtain $\mathcal{U}$ as follows

$$\mathcal{U}_{\text{fixed}} = \left\{m \in \mathbb{N}\,\middle|\,I_m^{(2)}(\{m\}) < I_m^{(2)}(\varnothing) = I_m^{(1)}(\varnothing)\right\}. \tag{112}$$

The decision rule in (112) results in a set $\mathcal{U}_{\text{fixed}}$ that is computed at the beginning of the first iteration and is left unchanged for the subsequent iterations.

APPENDIX D

Derivation of $\alpha_m^{(i)}$ for a P-QAM constellation

For P-QAM constellation and $i \geq 1$, the integral over $\mathbf{x}_m$ in (44) reduces to a summation, which can be written as

$$\alpha_m^{(i)} = \sum_{x\in\chi}\int |x - \eta(x-e)|^2\,p_{\mathbf{e}_m^{(i)}}(e)\,p_{\mathbf{x}_m}(x)\,de. \tag{113}$$

Since the P symbols are equally likely, $p_{\mathbf{x}_m}(x) = 1/P$, $\forall\,x$ and under the assumption that the errors $x - \eta(x-e)$ are dominated by the closest neighboring symbols, the above equation reduces to

$$\alpha_m^{(i)} = \frac{1}{P}\sum_{x\in\chi}d_x^2 k_x Q\left(\frac{\frac{d_x}{2}}{\sqrt{\frac{I_m^{(i)}}{2}}}\right) \tag{114}$$

where d_x is the distance between the symbol x and its closest neighbor and k_x is the number of symbols at a distance of d_x from x. The Q-function in the above equation results from the assumption on the statistics of $\mathbf{e}_m^{(i)}$. For a unit-power P-QAM constellation, $d_x = \sqrt{6/P-1}$, $\forall x$ [29]. In addition, it can be easily verified that $k_x = 2$ for the 4 corner symbols, $k_x = 3$ for the $(\sqrt{P}-2)4$ symbols on the outer edges, and $k_x = 4$ for the remaining $P - 4\sqrt{P} + 4$ symbols. Substituting these values into (114) yields

$$\alpha_m^{(i)}\big|_{i\geq 1} = \frac{24}{\sqrt{P}\left(\sqrt{P}+1\right)}Q\left(\sqrt{\frac{\frac{3}{(P-1)}}{I_m^{(i)}}}\right). \tag{115}$$

Moreover, since $\Delta\mathbf{x}_{\ell,m}^{(0)} = \mathbf{x}_{\ell,m}$, the value of $\alpha_m^{(0)}$ is 1. It completes the derivation of (45).

REFERENCES

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] J. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[6] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[7] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[8] H. Yang and T. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[9] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[10] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[11] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[12] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[13] H. Q. Ngo and E. Larsson, "EVD-based channel estimation in multi-cell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 3249–3252.

[14] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[15] J. Vinogradova, E. Björnson, and E. G. Larsson, "On the separability of signal and interference-plus-noise subspaces in blind pilot decontamination," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 3421–3425.

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

[17] K. Upadhya and S. A. Vorobyov, "An array processing approach to pilot decontamination for massive MIMO," in *Proc. IEEE 6th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, Cancun, Mexico, Dec. 2015, pp. 453–456.

[18] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[19] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large Rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.

[20] S. He, J. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3007–3021, Jun. 2007.

[21] M. Coldrey and P. Bohlin, "Training-based MIMO systems – Part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[22] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.

[23] T. Cui and C. Tellambura, "Pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE Global Telecommun. Conf.*, St. Louis, MO, USA, Dec. 2005, vol. 4, pp. 2229–2233.

[24] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, Mar. 2005.

[25] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Shanghai, China, Mar. 2016, pp. 3366–3370.

[26] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, Mar. 2016.

[27] C. Cox, *An Introduction to LTE: LTE, LTE-Advanced, SAE and 4G Mobile Communications*, 2nd ed. Chichester, U.K.: Wiley, 2012.

[28] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Downlink performance of superimposed pilots in massive MIMO systems in the presence of pilot contamination," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Washington, DC, USA, Dec. 2016, pp. 665–669.

[29] J. Proakis and M. Salehi, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2008.

Karthik Upadhya (S'16) received the B.E. degree in electronics and communication engineering from Visvesvaraya Technological University, Belgaum, India, in 2007, and the M.Tech. degree in communication systems from the Indian Institute of Technology Madras, Chennai, India, in 2011. He is currently working toward the Ph.D. degree at Aalto University, Espoo, Finland. His research interests include wireless communications and array signal processing. Before joining the Ph.D. program, he was a Member of the Technical Staff in the Saankhya Labs, Bengaluru, India, from 2011 to 2013, and a Research Assistant in the Indian Institute of Science, Bengaluru, India, from 2013 to 2014.

Sergiy A. Vorobyov (M'02–SM'05) received the M.Sc. and Ph.D. degrees in systems and control from Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, in 1994 and 1997, respectively. He is a Professor in the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland. He was previously with the University of Alberta, Edmonton, AB, Canada as an Assistant Professor from 2006 to 2010, Associate Professor from 2010 to 2012, and became Full Professor there in 2012. Since his graduation, he also held various research and faculty positions at Kharkiv National University of Radio Electronics, the Institute of Physical and Chemical Research, Tokyo, Japan, McMaster University, Hamilton, ON, Canada; Duisburg-Essen University, Duisburg, Germany, Darmstadt University of Technology, Darmstadt, Germany, and the Joint Research Institute between Heriot-Watt University and Edinburgh University, Edinburgh, U.K. His research interests include optimization and liner algebra methods in signal processing and communications; statistical and array signal processing; sparse signal processing; estimation and detection theory; sampling theory; and multi-antenna, very large, cooperative, and cognitive systems.

Dr. Vorobyov received the 2004 IEEE Signal Processing Society Best Paper Award, the 2007 Alberta Ingenuity New Faculty Award, the 2011 Carl Zeiss Award (Germany), the 2012 NSERC Discovery Accelerator Award, and other awards. Since 2016, he has been serving as an Area Editor of the IEEE SIGNAL PROCESSING LETTERS. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010 and the IEEE SIGNAL PROCESSING LETTERS from 2007 to 2009. He was a member of the Sensor Array and Multichannel Signal Processing and Signal Processing for Communications and Networking Technical Committees of the IEEE Signal Processing Society in 2007–2012 and 2010–2016, respectively. He was the Track Chair of Asilomar 2011, Pacific Grove, CA, USA; the Technical Co-Chair of IEEE CAMSAP 2011, San Juan, Puerto Rico; the Tutorial Chair of ISWCS 2013, Ilmenau, Germany; and the Technical Co-Chair of IEEE SAM 2018, Sheffield, U.K.

Mikko Vehkapera (M'10) received the Ph.D. degree from Norwegian University of Science and Technology, Trondheim, Norway, in 2010. From 2010 to 2013, he was a Postdoctoral Researcher in the School of Electrical Engineering, and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden, and from 2013 to 2015, an Academy of Finland Postdoctoral Researcher in the School of Electrical Engineering, Aalto University, Espoo, Finland. He is currently an Assistant Professor (Lecturer) in the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, U.K. He held visiting appointments in the Massachusetts Institute of Technology, Cambridge, MA, USA, Kyoto University, Kyoto, Japan, Tokyo Institute of Technology, Tokyo, Japan, and University of Erlangen-Nuremberg, Erlangen, Germany. His research interests include wireless communications, information theory, and signal processing. He was a corecipient of the Best Student Paper Award at IEEE International Conference on Networks and IEEE Sweden Joint VT-COM-IT Chapter Best Student Conference Paper Award 2015.

Publication V

Downlink Performance of Superimposed Pilots in Massive MIMO Systems

Karthik Upadhya, *Student Member, IEEE*,

Sergiy A. Vorobyov, *Fellow, IEEE*,

Mikko Vehkaperä, *Member, IEEE*

Abstract

In this paper, we investigate the downlink throughput performance of a massive multiple-input multiple-output (MIMO) system that employs superimposed pilots for channel estimation. The component of downlink (DL) interference that results from transmitting data alongside pilots in the uplink (UL) is shown to decrease at a rate proportional to the square root of the number of antennas at the BS. The normalized mean-squared error (NMSE) of the channel estimate is compared with the Bayesian Cramér-Rao lower bound that is derived for the system, and the former is also shown to diminish with increasing number of antennas at the base station (BS). Furthermore, we show that staggered pilots are a particular case of superimposed pilots and offer the downlink throughput of superimposed pilots while retaining the UL spectral and energy efficiency of regular pilots. We also extend the framework for designing a hybrid system, consisting of users that transmit either regular or superimposed pilots, to minimize both the UL and DL interference. The improved NMSE and DL rates of the channel estimator based on superimposed pilots are demonstrated by means of simulations.

Index Terms

Massive MIMO, pilot contamination, superimposed pilots, staggered pilots, downlink performance.

K. Upadhya, S. A. Vorobyov, and M. Vehkaperä are with the Department of Signal Processing and Acoustics, Aalto University, FI-00076 Aalto, Finland (E-mails: karthik.upadhya@aalto.fi, svor@ieee.org, mikko.vehkapera@aalto.fi). Parts of this paper were presented at IEEE GlobalSIP, Greater Washington, D.C., 2016 and IEEE ICASSP, New Orleans, Louisiana, 2017.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) communication systems employ base stations (BS) with a large number of antennas and have garnered significant interest in recent years as a candidate for future fifth generation (5G) cellular systems [1]–[5]. These systems promise a logarithmic increase in the uplink (UL) spectral and energy efficiency, with respect to the number of antennas at the BS, when the exact channel state information is assumed to be available for designing the detector [6]. In practice, the channel state information (CSI) has to be obtained at the BS using orthogonal pilot sequences that are transmitted by the users. These orthogonal sequences typically occupy a portion of the time-frequency resource dedicated for pilot transmission (henceforth referred to as regular pilots (RP)). The overhead for obtaining the CSI increases linearly in the number of orthogonal pilot sequences transmitted, and therefore, in order to limit this overhead, pilot sequences are shared/reused across cells in multi-cell systems. This sharing results in inter-cell interference in both the UL and downlink (DL), which is known as *pilot contamination* [7]. Pilot contamination diminishes the promised gains of massive MIMO systems and hence is considered a major impediment [8]. Approaches for pilot decontamination have garnered significant interest in recent years and they primarily rely on separating the users based on properties such as asymptotic orthogonality between user channels, non-overlapping angle of arrivals of the signal at the BS, and pilot reuse [9]–[14].

Superimposed pilots (SP) have been extensively studied for channel estimation in MIMO systems [15]–[19], especially in the context of rapidly changing channels in which reserving a set of symbols for pilot transmission would be impractical. Recently, in [20], SPs have been studied as an alternative pilot structure to mitigate/avoid pilot contamination in massive MIMO. SPs have also been investigated for use in massive MIMO systems in [21], [22]. In [21], the authors derive expressions for the UL spectral and energy efficiency of SP and compare them with those for RP. In [22], the authors have considered the case when the number of symbols in the UL time-slot is larger than the number of users in the system.

In [20], approximate expressions have been derived for the UL signal-to-interference-plus-noise ratio (SINR) and rate at the output of a matched filter (MF) that employs the least-squares (LS)-based channel estimate in an iterative and non-iterative way. The expressions have been derived under the condition that the total number of users in the system is smaller than the number of symbols in the UL time-slot. The importance of power control for a system employing SP

has been highlighted and the fractions of power that should be assigned to pilots and data, respectively, in order to maximize an approximation on the UL per cell rate, have been derived. It has been found that with increasing number of antennas at the BS, the optimal fraction of the power assigned to the data would decrease proportional to the square root of the number of antennas at the BS. In addition, a hybrid system that employs both RP and SP has been introduced to minimize the total UL interference, and shown to be superior to a system that is optimized for maximal spectral efficiency [11] but employs only RP.

In this paper, we provide additional important theoretical results with regard to SP for massive MIMO systems through performance metrics such as the normalized mean-squared error (NMSE) of the channel estimate and especially the DL rate. In particular, the following are the contributions of this paper.

- Closed-form expressions for the DL achievable rate are derived when the channel estimates obtained from SP are employed in a MF precoder at the BS.
- We discuss the relationship between staggered pilots and SP and derive the DL rate for the former scheme.
- We derive expressions for the NMSE and compare it against the Bayesian Cramér-Rao lower bound (CRLB) that we also derive for the system.
- The hybrid system described in [20], which consists of users that transmit both RP and SP, is extended to the DL and is designed by minimizing both the UL and DL interference.
- Simulations are carried out to validate the MSE and DL performance of SP and the hybrid system.

Some initial results, for the CRLB and approximate DL rate, have been reported in [23] without detailed derivations. In addition, some results for the hybrid system, with approximate UL and DL rates, have been reported in [24].

The paper is organized as follows. In Section II, we briefly review the system model for the UL and introduce the system model for the DL. In Section III, the DL rate is derived when the channel estimates are employed in an MF precoder. In addition, the expressions for the MSE and the corresponding CRLB of the channel estimate are derived for a system that employs SP. These metrics are then compared with the corresponding metrics that are obtained for a system employing RP. In Section IV, staggered pilots are shown to be a particular case of SP and the DL rate for this scheme is derived. In Section V, the framework for the hybrid system proposed in [20] is extended to include the downlink. Using simulations, Section VI discusses

the performance of the hybrid system and compares the MSE and DL performance of RP and SP. Section VII concludes the paper. Some of the lengthy proofs and derivations are detailed in the appendix.

Notation : Lower case and upper case boldface letters denote column vectors and matrices, respectively. The notations $(\cdot)^*, (\cdot)^T, (\cdot)^H$, and $(\cdot)^{-1}$ represent the conjugate, transpose, Hermitian transpose, and inverse, respectively. The Kronecker product is denoted by $\otimes$. The notation $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the complex normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $\mathbb{E}\{\cdot\}$ denotes the expectation operator. The notation $\mathbf{I}_N$ is used to denote an $N \times N$ identity matrix and $\|\cdot\|$ denotes the Euclidean norm of a vector. Upper case calligraphic letters denote sets and $\mathbf{1}_{\{\mathcal{S}\}}$ represents the indicator function over the set $\mathcal{S}$, while $\delta_{n,m}$ denotes the Kronecker delta function. The empty set is denoted by $\varnothing$, whereas the symbols $\cup$ and $\setminus$ stands for the union and the relative complement operations, respectively. The operator $\lfloor x \rceil$ returns the largest integer smaller than x. The trace of matrix $\boldsymbol{A}$ is represented as $\mathrm{trace}\{\boldsymbol{A}\}$.

II. SYSTEM MODEL

We consider a time-division duplexing (TDD) massive MIMO system with L cells and K single-antenna users per cell. Each cell has a BS with M antennas. In the UL phase, the users transmit C_u symbols, which include both data and pilots. Using the tuple (ℓ, k) to denote user k in cell ℓ, the matrix of received symbols $\boldsymbol{Y}_j \in \mathbb{C}^{M \times C_u}$ at BS j can be written as

$$\boldsymbol{Y}_j = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu_{\ell k}} \boldsymbol{h}_{j\ell k} \boldsymbol{s}_{\ell k}^T + \boldsymbol{W}_j \tag{1}$$

where $\boldsymbol{h}_{j\ell k} \in \mathbb{C}^M$ is the channel response between BS j and user (ℓ, k), $\boldsymbol{s}_{\ell k} \in \mathbb{C}^{C_u}$ is the vector of symbols transmitted by user (ℓ, k) with power $\mu_{\ell k}$, and $\boldsymbol{W}_j \in \mathbb{C}^{M \times C_u}$ is the matrix of additive white Gaussian noise at BS j with each column distributed as $\mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ and being mutually independent of the other columns. The channel vectors $\boldsymbol{h}_{j\ell k}$ are assumed to be distributed as $\mathcal{CN}(\boldsymbol{0}, \beta_{j\ell k} \boldsymbol{I})$ where $\beta_{j\ell k}$ denotes the large-scale path-loss coefficient. In addition, the channel is assumed to be constant during the coherence time, i.e., C symbols whereas $\beta_{j\ell k}$ is constant for a significantly longer duration than C symbols.

The symbol $\boldsymbol{s}_{\ell k}$ is dependent on the nature of the pilot transmitted in the UL. For example, when RP is employed, a part of $\boldsymbol{s}_{\ell k}$ is reserved for pilots and the remaining part is used for data transmission. Whereas, when SP is employed, the whole of $\boldsymbol{s}_{\ell k}$ contains both pilots and data.

Assuming channel reciprocity, if the BS ℓ uses the precoder $\boldsymbol{g}_{\ell k}$ and if $d_{\ell k} \in \mathbb{C}$ is the data symbol transmitted to user (ℓ, k) by BS ℓ, then the received symbol at user (j, m) can be written as

$$\widehat{d}_{jm} = \sqrt{\gamma} \sum_{\ell=0}^{L-1} \boldsymbol{h}_{\ell j m}^H \sum_{k=0}^{K-1} \sqrt{\nu_{\ell k}} \boldsymbol{g}_{\ell k} d_{\ell k} + w_{jm} \tag{2}$$

where γ is the DL SNR of user (j, m) and is assumed to be same in all the cells, and $w_{j,m}$ is zero-mean unit-variance additive Gaussian noise at the user terminal. The symbols $d_{\ell k}, \forall \ (\ell, k)$ are assumed to be distributed as $d_{\ell k} \sim \mathcal{CN}(0, 1)$ and are statistically independent of the channel vectors $\boldsymbol{h}$ and the UL symbols $\boldsymbol{s}$. The parameter $\nu_{\ell k} = q_{\ell k}/\mathbb{E}\left\{\|\boldsymbol{g}_{\ell k}\|^2\right\}$ normalizes the average transmit power to user (ℓ, k) to be $q_{\ell k}$ [11], [25].

We assume that all pilot transmissions are synchronized. This assumption is common in massive MIMO literature [6], [13], [26] since such a system is easy to analyze mathematically. However, in practice, network-wide synchronization may be infeasible. Local synchronization may be achieved using the cyclic prefix of the orthogonal frequency-division multiplexing (OFDM) symbols and the methods developed herein will suppress coherent interference from only the synchronized users [26]. For the purposes of this work, we assume that asynchronous cells are distant enough such that the effects of the interference resulting from them is negligible.

III. EFFECT OF PILOT CONTAMINATION ON THE DOWNLINK

In TDD massive MIMO, under the assumption of channel reciprocity, the precoder for data transmission in the DL is designed using the channel estimate that is obtained from UL training. Therefore, the throughput in the DL depends on the quality of the channel that has been estimated in the UL. In this section, the quality of the channel estimate obtained from both RP and SP transmission schemes are quantified through the normalized MSE and the latter is compared with the CRLB. In addition, the closed form expressions for the DL achievable rate at the user terminal are derived and compared when the channel estimates are used in an MF precoder.

A. Regular Pilots

With RP, each user transmits a $\tau \geq K$ length pilot sequence for channel estimation followed by UL data. Let the length-τ pilot sequences be taken from the columns of a scaled unitary matrix $\boldsymbol{\Phi} \in \mathbb{C}^{\tau \times \tau}$ such that $\boldsymbol{\Phi}^H \boldsymbol{\Phi} = \tau \boldsymbol{I}_\tau$. These orthogonal pilot sequences are distributed across $r^{\text{RP}} \triangleq \lfloor \tau/K \rfloor$ cells, where r^{RP} is assumed to be a positive integer. In other words, the
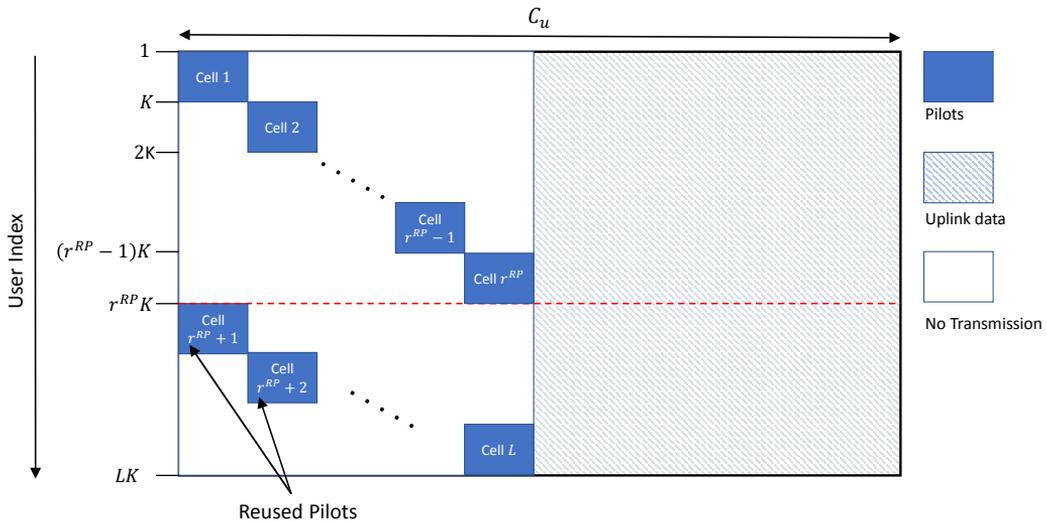
Fig. 1. Pilot structure of RP when $\mathbf{\Phi}$ is block diagonal with each block containing the orthogonal pilots in a cell. Each of the shaded blocks in the figure correspond to a block in the block diagonal matrix. The horizontal axis represents the C_u symbols in the UL slot. The vertical axis represents user indexes. Therefore, from the figure, we see that users in $\mathcal{L}_1\left(r^{\mathrm{RP}}\right)$ which includes cell 1, cell $r^{\mathrm{RP}} + 1$ etc., transmit pilots in the first τ symbols whereas users in the other cells remain silent.

pilot sequence $\boldsymbol{\phi}_{\ell k}$ that is transmitted by user (ℓ, k) is reused at every r^{RP}th cell. The pilot structure as well as reuse is depicted in Fig 1. Assuming that all the pilot transmissions are synchronized, the LS estimate of the channel can be easily found as [7], [20]

$$\widehat{\boldsymbol{h}}_{jjm}^{\mathrm{RP}} = \boldsymbol{h}_{jjm} + \sum_{\ell \in \mathcal{L}_j(r^{\mathrm{RP}})\backslash j} \sqrt{\frac{\mu_{\ell m}}{\mu_{jm}}} \boldsymbol{h}_{j\ell m} + \boldsymbol{w}_{jm} \tag{3}$$

where $\boldsymbol{w}_{jm} = \boldsymbol{W}_j \boldsymbol{\phi}_{jm}^* / \left(\tau\sqrt{\mu_{jm}}\right)$ and $\mathcal{L}_j(r^{\mathrm{RP}})$ is the subset of the L cells that use the same pilot sequences as cell j. The normalized MSE of the channel estimate $\widehat{\boldsymbol{h}}_{jjm}^{\mathrm{RP}}$ is defined as

$$\mathrm{NMSE}_{jm}^{\mathrm{RP}} \triangleq \frac{\mathbb{E}\left\{\|\widehat{\boldsymbol{h}}_{jjm}^{\mathrm{RP}} - \boldsymbol{h}_{jjm}\|^2\right\}}{\mathbb{E}\left\{\|\boldsymbol{h}_{jjm}\|^2\right\}} = \frac{1}{\beta_{jjm}} \left(\sum_{\ell \in \mathcal{L}_j(r^{\mathrm{RP}})\backslash j} \frac{\mu_{\ell m}}{\mu_{jm}} \beta_{j\ell m} + \frac{\sigma^2}{\tau\mu_{jm}} \right) . \tag{4}$$

The first term in (4) is the estimation error due to pilot contamination from users in the neighboring cells which employ the same pilots as user (j, m).

If C_d symbols are transmitted from the BS to the user terminals in the DL phase, then the rate in the downlink for user (j, m) can be expressed as [7]

$$R_{jm}^{\mathrm{RP-dl}} = \frac{C_d}{C} \log_2 \left(1 + \mathrm{SINR}_{jm}^{\mathrm{RP-dl}} \right) \tag{5}$$

where $C = C_u + C_d$ is the smallest channel coherence time of all the users in the system, and $\mathrm{SINR}_{jm}^{\mathrm{RP-dl}}$ is the DL SINR at user (j, m). If the channel estimate in (3) is used in an MF precoder, then [27]

$$\mathrm{SINR}_{jm}^{\mathrm{RP-dl}} = \frac{\nu_{jm} \beta_{jjm}^2}{\displaystyle\sum_{\ell \in \mathcal{L}_j(r^{\mathrm{RP}}) \backslash j} \frac{\mu_{jm}}{\mu_{\ell m}} \nu_{\ell m} \beta_{\ell jm}^2 + \frac{1}{M} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\nu_{\ell k}}{\mu_{\ell k}} \left(\displaystyle\sum_{n \in \mathcal{L}_\ell(r^{\mathrm{RP}})} \mu_{nk} \beta_{\ell jm} \beta_{\ell nk} + \frac{\sigma^2 \beta_{\ell jm}}{\tau} \right) + \frac{1}{M^2 \gamma}} \tag{6}$$

where

$$\nu_{\ell k} = \frac{q_{\ell k}}{\mathbb{E}\left\{ \|\boldsymbol{g}_{\ell k}\|^2 \right\}} = \frac{q_{\ell k}}{M} \left(\sum_{n \in \mathcal{L}_\ell(r^{\mathrm{RP}})} \frac{\mu_{nk}}{\mu_{\ell k}} \beta_{\ell nk} + \frac{\sigma^2}{\tau \mu_{\ell k}} \right)^{-1}. \tag{7}$$

The estimation error due to pilot contamination limits the asymptotic $(M \to \infty)$ DL SINR of user (j, m) to [7]

$$\mathrm{SINR}_{jm}^{\mathrm{RP-dl}} = \frac{\widetilde{\nu}_{jm} \beta_{jjm}^2}{\displaystyle\sum_{\ell \in \mathcal{L}_j(r^{\mathrm{RP}}) \backslash j} \widetilde{\nu}_{\ell m} \beta_{\ell jm}^2}. \tag{8}$$

where $\widetilde{\nu}_{\ell k} = M \nu_{\ell k}$. Note that $\widetilde{\nu}_{\ell k}$ is independent of M.

B. Superimposed Pilots

When employing SP, the estimate of the channel is obtained from pilots that are transmitted at a reduced power alongside the data. The LS estimate of the channel can be written as [20]

$$\widehat{\boldsymbol{h}}_{j\ell k}^{\mathrm{SP}} = \sum_{n \in \mathcal{L}_j(r^{\mathrm{SP}})} \sqrt{\frac{\mu_{nk}}{\mu_{\ell k}}} \boldsymbol{h}_{jnk} + \frac{\rho}{C_u \lambda} \sum_{n=0}^{L-1} \sum_{p=0}^{K-1} \sqrt{\frac{\mu_{np}}{\mu_{\ell k}}} \boldsymbol{h}_{jnp} \boldsymbol{x}_{np}^T \boldsymbol{p}_{\ell k}^* + \frac{\boldsymbol{W}_j \boldsymbol{p}_{\ell k}^*}{C_u \lambda \sqrt{\mu_{\ell k}}} \tag{9}$$

where $\boldsymbol{p}_{jm} \in \mathbb{C}^{C_u}$ and $\boldsymbol{x}_{jm} \in \mathbb{C}^{C_u}$ are, respectively, the pilot and data vectors transmitted by user (j, m), $r^{\mathrm{SP}} \triangleq \lfloor C_u/K \rfloor$ is a positive integer representing the number of cells over which SP are reused, $\mathcal{L}_j(r^{\mathrm{SP}})$ is the subset of the L cells that use the same pilot sequences as cell j. In addition, the pilots are taken from the columns of a scaled unitary matrix $\boldsymbol{P} \in \mathbb{C}^{C_u \times C_u}$ such that $\boldsymbol{P}^H \boldsymbol{P} = C_u \boldsymbol{I}_{C_u}$, and therefore, $\boldsymbol{p}_{\ell k}^H \boldsymbol{p}_{np} = C_u \delta_{\ell n} \delta_{kp}$. The parameters $\lambda^2 > 0$ and $\rho^2 > 0$ are the fractions of the UL transmit power reserved for pilots and data, respectively, such that $\lambda^2 + \rho^2 = 1$. Moreover, in (9), it is assumed that every user in the system uses the same value of λ and ρ.
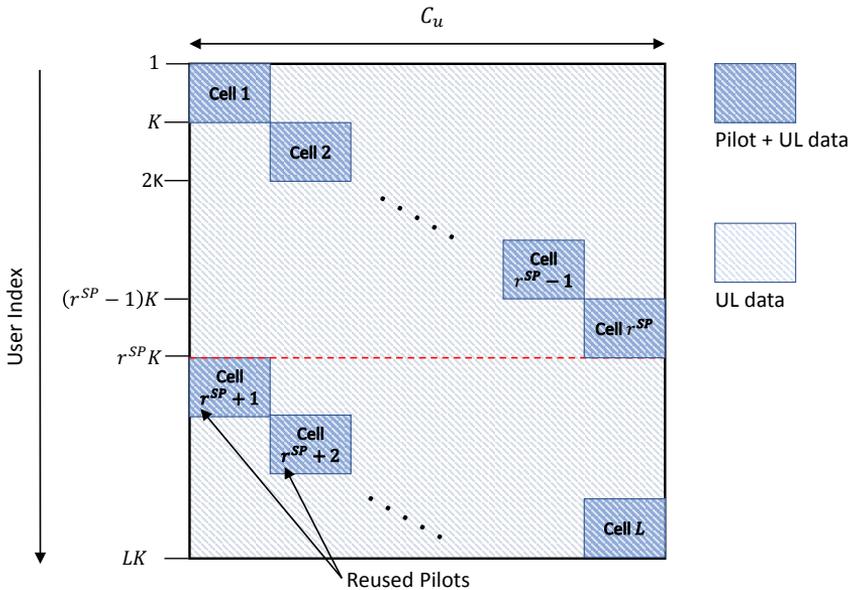
Fig. 2. Pilot structure of SP when $\boldsymbol{P}$ is block diagonal. Users transmit orthogonal pilots and data on top of each other and the number of pilots covers the entire UL slot.

Similar to (4), the normalized MSE for the channel estimate obtained from SP is defined as

$$
\begin{aligned}
\text{NMSE}_{jm}^{\text{SP}} &\triangleq \frac{\mathbb{E}\left\{\|\widehat{\boldsymbol{h}}_{jjm}^{\text{SP}} - \boldsymbol{h}_{jjm}\|^2\right\}}{\mathbb{E}\left\{\|\boldsymbol{h}_{jjm}\|^2\right\}} \\
&= \frac{1}{\beta_{jjm}} \left(\sum_{\ell \in \mathcal{L}_j(r^{\text{SP}})\backslash j} \frac{\mu_{\ell m}}{\mu_{jm}} \beta_{j\ell m} + \frac{\rho^2}{C_u \lambda^2} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\mu_{\ell k}}{\mu_{jm}} \beta_{j\ell k} + \frac{\sigma^2}{\lambda^2 C_u \mu_{jm}} \right) .
\end{aligned}
\tag{10}
$$

The first error term in (10) results from reusing pilots every r^{SP} cells, whereas the second error term results from transmitting pilots alongside data. As in the case of RP, both the errors lead to interference in the DL phase. Under the assumption that the interference from outside the r^{SP} contiguous cells which contain the reference BS can be neglected, the CRLB for the channel estimate can be derived as (the derivation is in Appendix C)

$$
\text{CRLB}\left(\boldsymbol{h}_{jjm}\right) = \frac{M}{\frac{C_u}{\sigma^2} + \frac{1}{\mu_{jm}\beta_{jjm}}} \approx \frac{M\sigma^2}{C_u}
\tag{11}
$$

where the approximation is valid when $\sigma^2/C_u \ll \mu_{jm}\beta_{jjm}$. Therefore, we have the relation,

$$
\text{NMSE}_{jm}^{\text{SP}} \geq \frac{1}{M\beta_{jjm}} \text{CRLB}\left(\boldsymbol{h}_{jjm}\right) \approx \frac{\sigma^2}{\beta_{jjm}C_u} .
\tag{12}
$$

The Bayesian CRLB is a lower bound on the MSE of a minimum mean-squared error (MMSE) channel estimator, and its value in (11) is the MSE of an MMSE estimator when $\rho = 0$ or all the power is allocated to the pilots. In addition, the approximation in (11) is the MSE of the LS estimator when $\rho = 0$.

The NMSE in (10) is parameterized by both ρ^2 and λ^2. However, the CRLB is loose for non-zero values of ρ^2 and an estimator will attain this bound only when $\rho^2 = 0$, i.e., when all the power is allocated to the pilots. Nevertheless, the CRLB is a standard and useful benchmark to evaluate the performance of the proposed method. For example, in the same context of massive MIMO, the performance of a semi-blind channel estimation method is also compared against the CRLB in [28]. We will also see in Section VI that the CRLB in (11) is achieved[1] by the estimator in (9) when $M \to \infty$.

A lower bound on the DL ergodic capacity can be obtained for superimposed pilots using a similar approach as in [27]. Rewriting (2) as

$$
\begin{aligned}
\widehat{d}_{jm} = {} & \sqrt{\gamma \nu_{jm}} \mathbb{E} \left\{ \boldsymbol{h}_{jjm}^H \boldsymbol{g}_{jm} \right\} d_{jm} + \sqrt{\gamma \nu_{jm}} \left(\boldsymbol{h}_{jjm}^H \boldsymbol{g}_{jm} - \mathbb{E} \left\{ \boldsymbol{h}_{jjm}^H \boldsymbol{g}_{jm} \right\} \right) d_{jm} \\
& + \sqrt{\gamma} \sum_{(\ell,k) \neq (j,m)} \sum \sqrt{\nu_{\ell k}} \boldsymbol{h}_{\ell jm}^H \boldsymbol{g}_{\ell k} d_{\ell k} + w_{jm} ,
\end{aligned}
\tag{13}
$$

and noting that the first term is uncorrelated with the subsequent terms, a lower bound on the ergodic capacity can be computed as [29]

$$
R_{jm}^{\text{SP}-\text{dl}} = \frac{C_d}{C} \mathbb{E} \left\{ \log_2 \left(1 + \text{SINR}_{jm}^{\text{SP}-\text{dl}} \right) \right\}
\tag{14}
$$

where

$$
\text{SINR}_{jm}^{\text{SP}-\text{dl}} = \frac{\nu_{jm} | \mathbb{E} \left\{ \boldsymbol{h}_{jjm}^H \boldsymbol{g}_{jm} \right\} |^2}{\sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \nu_{\ell k} \mathbb{E} \left\{ |\boldsymbol{h}_{\ell jm}^H \boldsymbol{g}_{\ell k}|^2 \right\} - \nu_{jm} | \mathbb{E} \left\{ \boldsymbol{h}_{jjm}^H \boldsymbol{g}_{jm} \right\} |^2 + \frac{1}{\gamma}} .
\tag{15}
$$

In (14) and (15), the expectation outside the logarithm is with respect to the user locations whereas the inner expectation is with respect to the channel and noise vectors. In addition, these expressions are valid for any combining scheme. However, to obtain closed-form expressions for precoders such as zero-forcing (ZF) or MMSE, we require the channel estimation error to be independent of the estimate. But, as outlined in [21], the LS channel estimate and the estimation error when SP is employed are not Gaussian. As a result, even if a linear minimum mean-squared error (LMMSE) channel estimate were to be employed, it would only result in

---

[1]In fact, it is the approximation in (11) that is achieved by the estimator.

the estimation error being uncorrelated with the estimate but not independent of it. This renders it difficult/impossible in general to obtain closed-form expressions for precoders such as ZF and MMSE. We will therefore obtain a closed form expression for the SINR for MF precoding and numerically evaluate (14) and (15) in Section VI for methods such as ZF. Setting $\boldsymbol{g}_{\ell k} = \widehat{\boldsymbol{h}}_{\ell\ell k}$, the DL SINR for the MF precoder has been obtained in Appendix A as

$$\mathrm{SINR}_{jm}^{\mathrm{SP-dl}} = \nu_{jm}\beta_{jjm}^2 \left(\sum_{\ell\in\mathcal{L}_j(r^{\mathrm{SP}})\backslash j} \frac{\mu_{jm}}{\mu_{\ell m}}\nu_{\ell m}\beta_{\ell jm}^2 + \frac{1}{M}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\frac{\nu_{\ell k}\beta_{\ell jm}}{\mu_{\ell k}} \left(\sum_{n\in\mathcal{L}_\ell(r^{\mathrm{SP}})}\mu_{nk}\beta_{\ell nk} + \frac{\sigma^2}{C_u\lambda^2} \right) \right.$$
$$\left. + \frac{\rho^2}{C_u\lambda^2}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\frac{\nu_{\ell k}}{\mu_{\ell k}} \left(\mu_{jm}\beta_{\ell jm}^2 + \frac{1}{M}\sum_{n=0}^{L-1}\sum_{p=0}^{K-1}\mu_{np}\beta_{\ell np}\beta_{\ell jm} \right) + \frac{1}{M^2\gamma} \right)^{-1} \tag{16}$$

where

$$\nu_{\ell k} = \frac{q_{\ell k}}{\mathbb{E}\{\|\boldsymbol{g}_{\ell k}\|^2\}} = \frac{q_{\ell k}\mu_{\ell k}}{M} \left(\sum_{n\in\mathcal{L}_\ell(r^{\mathrm{SP}})}\mu_{nk}\beta_{\ell nk} + \sum_{n=0}^{L-1}\sum_{p=0}^{K-1}\mu_{np}\frac{\rho^2}{C_u\lambda^2}\beta_{\ell np} + \frac{\sigma^2}{C_u\lambda^2} \right)^{-1}. \tag{17}$$

In [20], the optimal values of ρ^2 and λ^2 were computed by maximizing an approximation of the UL sum-rate which was obtained assuming that $C_u \geq LK$. However, since in practice $C_u \leq LK$, the results obtained in [20] do not necessarily hold. In this paper, we obtain the optimal values of ρ^2 and λ^2 by maximizing an achievable lower bound on the channel capacity when $C_u \leq LK$.

Proposition 1. *The values of ρ^2 and λ^2 that maximize the achievable rate in the UL are*

$$\rho_{\mathrm{opt}}^2 = \left(1 + \sqrt{M}\kappa \right)^{-1} \tag{18}$$

$$\lambda_{\mathrm{opt}}^2 = \left(1 + \frac{1}{\kappa\sqrt{M}} \right)^{-1} \tag{19}$$

where

$$\kappa \triangleq \sqrt{\frac{\alpha_1 + \frac{\alpha_2}{M} + \frac{\alpha_4}{M} + \frac{\alpha_6}{M}}{\alpha_3 + \alpha_4 + \alpha_5}} \tag{20}$$

$$\alpha_1 \triangleq \sum_{k=0}^{N-1}\frac{\mu_k^2\beta_k^2}{\mu_0(C_u-1)} \,,\; \alpha_2 \triangleq \sum_{k=0}^{N-1}\sum_{n=0}^{N-1}\frac{\mu_k\mu_n\beta_k\beta_n}{\mu_0(C_u-1)} \,;\; \alpha_3 \triangleq \sum_{k=0}^{N-1}\frac{\mu_k d_k C_u}{(C_u-1)} \tag{21}$$

$$\alpha_4 \triangleq \sum_{k=0}^{N-1}\frac{\mu_k e_k C_u}{C_u-1} + \frac{\sigma^4}{\mu_0 C_u} \,;\; \alpha_5 \triangleq \sum_{k\in\mathcal{P}_0}\sigma^2\frac{\mu_k}{\mu_0}\beta_k \,;\; \alpha_6 \triangleq \sum_{k=0}^{N-1}\frac{\sigma^2}{C_u}\frac{\mu_k}{\mu_0}\beta_k \,. \tag{22}$$

Proof. The achievable rate in the UL as well as expressions for ρ_{opt}^2 and λ_{opt}^2 are derived in Appendix B. $\qquad\square$

Remark 1: Note that the exact expressions for achievable rate for SP when $LK \geq C_u$ have been derived earlier in [21] (c.f. Theorem 1 and Corollary 1). However, Theorem 1 in [21] underestimates the UL rate since it treats the pilot that is transmitted alongside data in each UL symbol as interference, whereas Corollary 1 over-estimates the rate since it assumes that the pilots are perfectly removed. On the other hand, in the expression for the UL rate derived in Appendix B, we side-step this issue by multiplying the received observations with a unitary matrix that relegates all the interference resulting from transmitting pilots alongside data to a single symbol. This symbol can then be discarded since we are anyway interested in only a lower bound on the ergodic capacity. The remaining $C_u - 1$ symbols of the reference user are free of interference from the UL pilot of that user, and therefore, standard methods can be used to calculate the UL throughput in these symbols.

Substituting (18) and (19) into (10), the expression for the NMSE becomes

$$\text{NMSE}_{j,m}^{\text{SP}}\bigg|_{\rho_{\text{opt}},\lambda_{\text{opt}}} = \frac{1}{\beta_{jjm}} \left(\frac{1}{\sqrt{M}\kappa C_u} \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\mu_{\ell k}}{\mu_{jm}} \beta_{j\ell k} + \sum_{\ell \in \mathcal{L}_j(r^{\text{SP}})\backslash j} \frac{\mu_{\ell m}}{\mu_{jm}} \beta_{j\ell m} \right.$$
$$\left. + \left(1 + \frac{1}{\sqrt{M}\kappa}\right) \frac{\sigma^2}{C_u \mu_{jm}} \right) . \tag{23}$$

Thus, with optimized values of ρ^2 and λ^2, the component of the $\text{NMSE}_{j,m}^{\text{SP}}$ resulting from transmitting data alongside pilots reduces proportional to the square root of the number of antenna elements. This behavior is in contrast to (4), wherein $\text{NMSE}_{j,m}^{\text{RP}}$ is independent of M. Consequently, the reduction in the NMSE also leads to a higher DL throughput, as shown below.

Substituting (18) and (19) into (16), the expression for the DL SINR becomes

$$\text{SINR}_{jm}^{\text{SP}-\text{dl}} = \frac{\nu_{jm}\beta_{jjm}^2}{\sum\limits_{\ell \in \mathcal{L}_j(r^{\text{SP}})\backslash j} \frac{\mu_{jm}}{\mu_{\ell m}}\nu_{\ell m}\beta_{\ell jm}^2 + \frac{t_1}{\sqrt{M}\kappa C_u} + \frac{t_2}{M}} \tag{24}$$

where

$$t_1 = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \nu_{\ell k} \frac{\mu_{jm}}{\mu_{\ell k}} \beta_{\ell jm}^2 \tag{25}$$

$$t_2 = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \frac{\nu_{\ell k}\beta_{\ell jm}}{\mu_{\ell k}} \left(\sum_{n \in \mathcal{L}_\ell(r^{\text{SP}})} \mu_{nk}\beta_{\ell nk} + \left(1 + \frac{1}{\sqrt{M}\kappa}\right) \frac{\sigma^2}{C_u} \right)$$
$$+ \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sum_{n=0}^{L-1} \sum_{p=0}^{K-1} \frac{\mu_{np}\nu_{\ell k}\beta_{\ell np}\beta_{\ell jm}}{\sqrt{M}\mu_{\ell k}\kappa C_u} + \frac{1}{M\gamma} . \tag{26}$$

From the above expression, it can be observed that $t_1/\sqrt{M}$, which is the component of interference from users that do not share a pilot with user (j, m), decreases proportional to the square root of M. Since $C_u \gg K$, more orthogonal pilot sequences are available when SPs are employed in comparison with RP. As a result, SP can be reused over a larger number of cells, i.e., $r^{\mathrm{SP}} > r^{\mathrm{RP}}$. Therefore, we have the following result.

Proposition 2. *If $r^{\mathrm{SP}} > r^{\mathrm{RP}}$, the ceiling of $\mathrm{SINR}_{jm}^{\mathrm{SP-dl}}$ when $M \to \infty$ is higher than that of $\mathrm{SINR}_{jm}^{\mathrm{RP-dl}}$.*

Proof. When $M \to \infty$ (16) becomes

$$\mathrm{SINR}_{jm}^{\mathrm{SP-dl}} = \frac{\widetilde{\nu}_{jm}\beta_{jjm}^2}{\sum\limits_{\ell \in \mathcal{L}_j(r^{\mathrm{SP}})\backslash j} \frac{\mu_{jm}}{\mu_{\ell m}}\widetilde{\nu}_{\ell m}\beta_{\ell jm}^2} \ . \tag{27}$$

Therefore, comparing (8) and (27), the denominator of (27) is smaller than that of (8) when $r^{\mathrm{SP}} > r^{\mathrm{RP}}$. Here $\widetilde{\nu}_{jm} = \lim_{M \to \infty} M\nu_{jm}$ and is independent of M. $\qquad\square$

The values of ρ^2 and λ^2 can be computed in practice in a designated BS in the network. The remaining BSs transmit the large-scale path-loss coefficients to this BS. While the transmission of these coefficients requires overhead, the large-scale statistics of the channel is valid for one to two orders of magnitude longer than the channel coherence time [30], [31]. In addition, since these coefficients are only scalars, we expect that this overhead is negligible.

IV. STAGGERED PILOTS AS A PARTICULAR CASE OF SUPERIMPOSED PILOTS

When transmitting staggered pilots [32], [33], the users in each cell stagger[2] their pilot transmissions so that the users of no two cells within the r^{SP} cells, which share the C_u UL pilots, are transmitting UL pilots simultaneously, i.e., if the users in a particular cell are transmitting UL pilots, the users in the remaining $r^{\mathrm{SP}} - 1$ cells transmit data. This pilot structure is depicted in Fig. 3 Let $\boldsymbol{Y}_n$ be the observation at BS j when the users in the nth cell (where $0 \le n \le r^{\mathrm{SP}} - 1$) transmit UL pilots. Note that the index j has been dropped from $\boldsymbol{Y}_n$ for the sake of simplicity of notation. Then, $\boldsymbol{Y}_n \in \mathbb{C}^{M \times \tau}$ can be written as

$$\boldsymbol{Y}_n \triangleq \sum_{\ell \in \mathcal{L}_n(r^{\mathrm{SP}})} \sum_k \sqrt{\mu_{\ell k} p_p}\boldsymbol{h}_{j\ell k}\boldsymbol{\phi}_{nk}^T + \sum_{\ell \notin \mathcal{L}_n(r^{\mathrm{SP}})} \sum_k \sqrt{\mu_{\ell k} p_d}\boldsymbol{h}_{j\ell k}\left(\boldsymbol{x}_{\ell k}^n\right)^T + \boldsymbol{W}_n \tag{28}$$

---

[2]To achieve the staggered pilot frame structure, we assume that the BSs coordinate the UL pilot transmissions of their users.
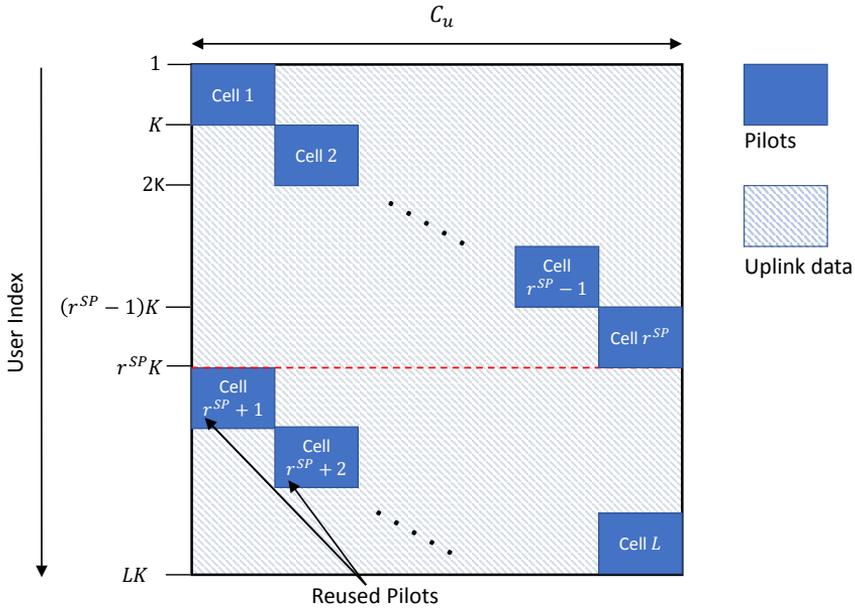
Fig. 3. Pilot structure of staggered pilots. In contrast with RP, the pilot transmissions cover the entire UL slot, as with SP. However, in contrast with SP, a particular user does not transmit pilots and data simultaneously.

where $\boldsymbol{\phi}_{n,k} \, \forall n, k$ are the orthogonal pilot sequences described in Subsection III-A, p_p and p_d are the powers at which the pilots and the data, respectively, are transmitted, and $\boldsymbol{x}_{\ell,k}^n \in \mathbb{C}^\tau$ is the vector of data symbols transmitted by user (ℓ, k) in the nth block. We then have the following proposition.

Proposition 3. *The UL in a system that employs staggered pilots in* (28) *is a particular case of superimposed pilots if $p_p = \mu\lambda^2 C_u/\tau$, $p_d = \mu\rho^2$ and $\boldsymbol{P} = \sqrt{\frac{C_u}{\tau}}\mathrm{blkdiag}\,\{\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_{L-1}\}$.*

An important conclusion of Proposition 3 is that staggered pilots are capable of achieving the downlink throughput of SP while maintaining the UL spectral efficiency of RP. Indeed, utilizing the same approach used to derive (16), a lower bound on the DL ergodic capacity when the channel estimate obtained from staggered pilots is employed in a MF precoder can be obtained as

$$R_{jm}^{\mathrm{ST-dl}} = \frac{C_d}{C} \log_2\left(1 + \mathrm{SINR}_{jm}^{\mathrm{ST-dl}}\right) \tag{29}$$

where

$$\text{SINR}_{jm}^{\text{ST-dl}} = \nu_{jm}\beta_{jjm}^2\left(\sum_{\ell\in\mathcal{L}_j(r^{\text{SP}})\backslash j}\nu_{\ell m}\frac{\mu_{jm}}{\mu_{\ell m}}\beta_{j\ell m}^2 + \frac{1}{M}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\frac{\nu_{\ell k}\beta_{\ell jm}}{\mu_{\ell k}}\left(\sum_{n\in\mathcal{L}_\ell(r^{\text{SP}})}\mu_{nk}\beta_{\ell nk} + \frac{\sigma^2}{p_p\tau}\right)\right.$$

$$\left.+\frac{p_d}{\tau p_p}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\frac{\nu_{\ell k}}{\mu_{\ell k}}\left(\mu_{jm}\beta_{\ell jm}^2\mathbf{1}_{\{j\notin\mathcal{L}_\ell(r^{\text{SP}})\}} + \frac{1}{M}\sum_{n\notin\mathcal{L}_\ell(r^{\text{SP}})}\sum_{p=0}^{K-1}\mu_{np}\beta_{\ell jm}\beta_{\ell np}\right) + \frac{1}{M^2\gamma}\right)^{-1} \qquad (30)$$

and

$$\nu_{\ell k} = \frac{q_{\ell k}}{\mathbb{E}\left\{\|\boldsymbol{g}_{\ell k}\|^2\right\}} = \frac{q_{\ell k}}{M}\left(\sum_{n\in\mathcal{L}_\ell(r^{\text{SP}})}\frac{\mu_{nk}}{\mu_{\ell k}}\beta_{\ell nk} + \frac{p_d}{\tau p_p}\sum_{n\notin\mathcal{L}_\ell(r^{\text{SP}})}\sum_{p=0}^{K-1}\frac{\mu_{np}}{\mu_{\ell k}}\beta_{\ell np} + \frac{\sigma^2}{\tau p_p\mu_{\ell k}}\right)^{-1} . \qquad (31)$$

Therefore, staggered pilots can achieve the DL performance of RP with a reuse factor r^{SP} with an overhead equivalent to that of RP with pilot-reuse factor r^{RP}. As a result, similar to Proposition 2, we have

Proposition 4. *If $r^{\text{SP}} > r^{\text{RP}}$, the ceiling of $\text{SINR}_{jm}^{\text{ST-dl}}$ when $M \to \infty$ is higher than that of* $\text{SINR}_{jm}^{\text{RP-dl}}$.

The concept described in this section can be further demonstrated by a simple example. Consider a system with two users A and B. Without loss of generality, it is assumed that the large-scale path-loss between the BS and users A and B are unity. In the UL phase, let user B transmit data with power $\rho^2 \geq 0$ when user A transmits its pilot at unit power. In the DL phase, user B receives interference at a power ρ^2 from the DL transmission to user A. Thus, increasing the number of antennas M at the BS increases the array gain at the BS, allowing for user B to transmit with a smaller power ρ^2, thereby reducing the interference it sees in the DL.

V. EXTENSION OF HYBRID SYSTEM TO DL

Using the DL SINR analysis in Section III, we extend the hybrid system in [20] to include the DL as well. The hybrid system consists of two sets of users $\mathcal{U}_{\text{RP}}$ and $\mathcal{U}_{\text{SP}}$ that transmit RP and SP, respectively. As shown in Fig. 4, users in $\mathcal{U}_{\text{RP}}$ obtain channel estimates in the UL using RPs, transmitted over τ symbols, and use these estimates to detect data using a spatial filter. However, users in $\mathcal{U}_{\text{SP}}$ maintain radio silence during the pilot training phase of the users in $\mathcal{U}_{\text{RP}}$, i.e., for τ symbols in the frame, and transmit orthogonal pilots superimposed with data during the UL data phase of $C_u - \tau$ symbols. By this construction, users in $\mathcal{U}_{\text{SP}}$ effectively transmit the zero vector for the τ training symbols and since the zero vector is orthogonal to all other vectors,
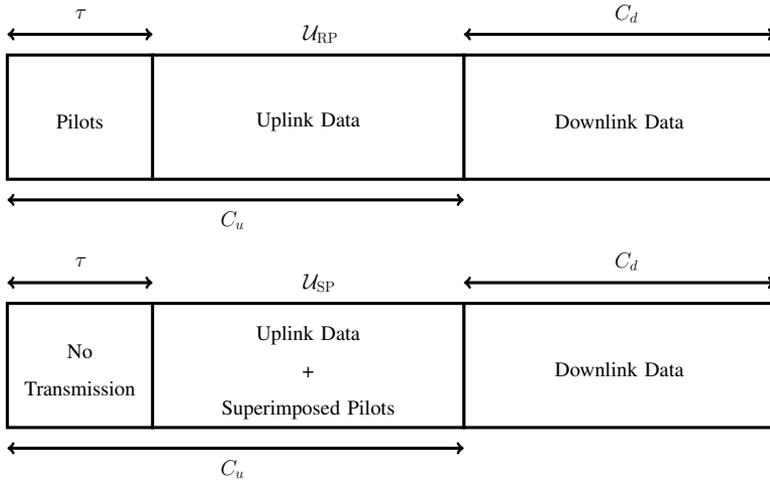
Fig. 4. Frame structure of a hybrid system with users employing RP and SP pilots.

the users in $\mathcal{U}_{\mathrm{SP}}$ can be viewed as having orthogonal pilots, thus not affecting the performance of any user in $\mathcal{U}_{\mathrm{RP}}$.

Employing the channel estimates obtained from $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$ in a MF precoder and combiner, the SINR in the UL (see [20]) and DL from users in $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$ when $M \gg K$ [3] can be obtained as

$$\mathrm{SINR}_{jm}^{\mathrm{RP-ul}} = \frac{\beta_{jjm}^2}{\displaystyle\sum_{(\ell,m)\in\mathcal{U}_{\mathrm{RP}}\backslash j} \beta_{j\ell m}^2} \tag{32}$$

$$\mathrm{SINR}_{jm}^{\mathrm{RP-dl}} = \frac{\widetilde{\nu}_{jm}\beta_{jjm}^2}{\displaystyle\sum_{(\ell,m)\in\mathcal{U}_{\mathrm{RP}}\backslash j} \widetilde{\nu}_{\ell m}\beta_{\ell jm}^2} \tag{33}$$

$$\mathrm{SINR}_{jm}^{\mathrm{SP-ul}} \approx \frac{\beta_{jjm}^2}{\frac{1}{(C_u-\tau)\lambda^2} \displaystyle\sum_{\ell}\sum_{\substack{k \\ (\ell,k)\in\mathcal{U}_{\mathrm{SP}}}} \beta_{j\ell k}^2} \tag{34}$$

$$\mathrm{SINR}_{jm}^{\mathrm{SP-dl}} \approx \frac{\widetilde{\nu}_{jm}\beta_{jjm}^2}{\frac{\rho^2}{(C_u-\tau)\lambda^2} \displaystyle\sum_{\ell}\sum_{\substack{k \\ (\ell,k)\in\mathcal{U}_{\mathrm{SP}}}} \widetilde{\nu}_{\ell m}\beta_{\ell jm}^2} \tag{35}$$

[3] Since this section is devoted to designing a suboptimal algorithm to partition users into those that use RP and SP, we rely on approximate asymptotic expressions (33) – (35), for the sake of simplicity, since the problem is anyway solved sub-optimally.

where the approximations in (34), and (35) reflect on the assumption that the users in $\mathcal{U}_{\text{RP}}$ and $\mathcal{U}_{\text{SP}}$ do not interfere with each other. This assumption is valid if the UL transmission power of the users in $\mathcal{U}_{\text{RP}}$ is significantly smaller than those in $\mathcal{U}_{\text{SP}}$. This assumption is made for the sake of simplicity and clarity only. In the absence of this assumption, the BS will have to estimate and remove the interference from the users in $\mathcal{U}_{\text{RP}}$ before estimating the channel vectors of the users in $\mathcal{U}_{\text{SP}}$. In addition, for the sake of simplicity, we assume that $r^{\text{RP}} = 1$ and that the interference from the cells other than the ones adjacent to the reference cell are negligible. Furthermore, the UL transmit powers $\mu_{\ell k}$ are subsumed into the coefficients $\beta_{j\ell k}$.

In [20], the objective of the hybrid system design has been defined as to partition the users into disjoint sets $\mathcal{U}_{\text{RP}}$ and $\mathcal{U}_{\text{SP}}$ by minimizing the overall UL interference. Using (33) and (34), we extend here the objective to jointly minimize the UL and DL interference.

Let $I_{jm}^{\text{RP}-\text{ul}}$ or $I_{jm}^{\text{RP}-\text{dl}}$, respectively, be the contributions of user (j, m) to the total UL and DL inter/intra-cell interference power when assigned to $\mathcal{U}_{\text{RP}}$. Similarly, let $I_{jm}^{\text{SP}-\text{ul}}$ or $I_{jm}^{\text{SP}-\text{dl}}$, respectively, be the contributions of user (j, m) to the total DL inter/intra-cell interference power when assigned to $\mathcal{U}_{\text{SP}}$. If users (j, m) and (ℓ, k) are members of $\mathcal{U}_{\text{RP}}$, then from the denominator of (32), the amount of interference that user (j, m) causes to user (ℓ, k) in the UL is $\beta_{\ell j k}^2 \delta_{m,k}$. Similarly, from (33), the amount of interference that user (j, m) causes to user (ℓ, k) in the DL is $\beta_{n\ell k}^2 \delta_{j,\ell} \delta_{m,k}$, $\forall n \neq \ell$, $n \in \mathcal{L}_j(r)$, $(n, k) \in \mathcal{U}_{\text{RP}}$. Likewise, from (34) and (35), if both users are members of $\mathcal{U}_{\text{SP}}$ then the amount of interference that user (j, m) causes to user (ℓ, k) in the UL and DL is $\beta_{\ell j m}^2 / ((C_u - \tau)\lambda^2)$ and $\rho^2 \beta_{njm}^2 / ((C_u - \tau)\lambda^2)$, $\forall n \neq j$, $n = 0, \ldots, L - 1$, respectively. Therefore, $I_{jm}^{\text{RP}-\text{ul}}$, $I_{jm}^{\text{RP}-\text{dl}}$, $I_{jm}^{\text{SP}-\text{ul}}$, and $I_{jm}^{\text{SP}-\text{dl}}$ can be obtained as

$$I_{jm}^{\text{RP}-\text{ul}} = \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r) \\ (\ell,k) \in \mathcal{U}_{\text{RP}}}} \sum_k \beta_{\ell j k}^2 \delta_{m,k} = \sum_{\substack{\ell \neq j \\ \ell \in \mathcal{L}_j(r) \\ (\ell,m) \in \mathcal{U}_{\text{RP}}}} \beta_{\ell j m}^2 \tag{36}$$

$$I_{j,m}^{\text{RP}-\text{dl}} = \sum_{\substack{n \neq j \\ n,\ell \in \mathcal{L}_j(r) \\ (n,k) \in \mathcal{U}_{\text{RP}}}} \sum_\ell \sum_k \beta_{n\ell k}^2 \delta_{j,\ell} \delta_{m,k} = \sum_{\substack{n \neq j \\ n \in \mathcal{L}_j(r) \\ (n,m) \in \mathcal{U}_{\text{RP}}}} \beta_{njm}^2 \tag{37}$$

$$I_{jm}^{\text{SP}-\text{ul}} = \frac{1}{(C_u - \tau)\lambda^2} \sum_\ell \sum_{\substack{k \\ (\ell,k) \in \mathcal{U}_{\text{SP}}}} \beta_{\ell j m}^2 \tag{38}$$

$$I_{jm}^{\text{SP}-\text{dl}} = \frac{\rho^2}{(C_u - \tau)\lambda^2} \sum_\ell \sum_{\substack{k \\ (\ell,k) \in \mathcal{U}_{\text{SP}}}} \beta_{\ell j m}^2 = \rho^2 I_{jm}^{\text{SP}-\text{ul}}. \tag{39}$$

Let $\xi^{\mathrm{ul}} > 0$ and $\xi^{\mathrm{dl}} > 0$ be the weights for the interference powers in the UL and DL, respectively, such that $\xi^{\mathrm{ul}} + \xi^{\mathrm{dl}} = 1$. Then, the total cost due to inter/intra-cell interference can be expressed as

$$I\left(\mathcal{U}_{\mathrm{RP}}, \mathcal{U}_{\mathrm{SP}}\right) = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \left(T_{\ell k}^{\mathrm{RP}} \mathbf{1}_{\{(\ell,k)\in\mathcal{U}_{\mathrm{RP}}\}} + T_{\ell k}^{\mathrm{SP}} \mathbf{1}_{\{(\ell,k)\in\mathcal{U}_{\mathrm{SP}}\}}\right) \tag{40}$$

where $T_{\ell k}^{\mathrm{RP}}$ and $T_{\ell k}^{\mathrm{SP}}$ are the costs incurred when user (ℓ, k) is assigned to $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$, respectively, and are defined as

$$T_{\ell k}^{\mathrm{RP}} \triangleq \xi^{\mathrm{ul}} I_{\ell k}^{\mathrm{RP-ul}} + \xi^{\mathrm{dl}} I_{\ell k}^{\mathrm{RP-dl}} \tag{41}$$

$$T_{\ell k}^{\mathrm{SP}} \triangleq \xi^{\mathrm{ul}} I_{\ell k}^{\mathrm{SP-ul}} + \xi^{\mathrm{dl}} I_{\ell k}^{\mathrm{SP-dl}} . \tag{42}$$

Minimizing (40) over the possible choices of $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$, the optimal sets $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$ can be obtained as the solution of the following optimization problem

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

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

$$\mathcal{U}_{\mathrm{RP}} \cap \mathcal{U}_{\mathrm{SP}} = \varnothing \tag{43}$$

where $\mathcal{U}$ is the set of all users in the L cells. However, obtaining the solution to the optimization problem in (43) is combinatorial in nature with $2^{\mathrm{card}\{\mathcal{U}\}}$ possible choices for $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$. A simple greedy algorithm to partition the users by minimizing only the overall UL interference has been devised in [20], and it can be straightforwardly extended to jointly minimize both the UL and DL interference powers.

VI. SIMULATION RESULTS

A. Downlink and Channel Estimation Performance

We compare the DL throughput and MSE performance of systems that employ the LS-based channel estimates obtained from RP to the performance of the massive MIMO systems that obtain channel estimates from SP and staggered pilots.

Unless otherwise specified, the simulation parameters are as follows. The users are uniformly distributed in hexagonal cells and are at a distance of at least 100m from the BS. The inter-BS separation is 2km. The channel estimation methods are tested with $L = 91$ cells and $K = 5$ users per cell. Both the SP and staggered pilots have reuse factors $r^{\mathrm{SP}} = 7$. The path loss coefficient
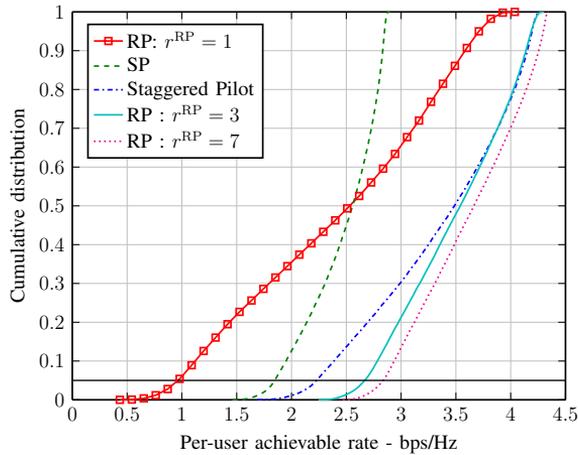
Fig. 5. Cumulative distribution of DL throughput for $M = 100$ antennas. The black line indicates rates with probability ≥ 0.95. The DL rates for SP and staggered pilots that have a probability ≤ 0.05 are significantly higher than those of RP with $r^{\mathrm{RP}} = 1$.

is assumed to be 3. The number of symbols in the UL and DL, i.e., C_u and C_d, respectively, are both chosen as 35 symbols. The values of ρ and λ are computed from (18) and (19), respectively, and are used for both SP and staggered pilots. The UL transmit power $\mu_{\ell k}$ is chosen based on the statistical channel-inversion power-control scheme [11], i.e., $\mu_{\ell k} = \omega/\beta_{\ell \ell k}$ where ω is a design parameter. The signal-to-noise ratio (SNR) in the UL and DL, i.e., ω/σ^2 and γ, respectively, is set to 10dB. The plots are generated by averaging over 10^4 realizations of user locations across the cell. For the sake of simplicity, the effects of shadowing are not taken into account in this paper, but the conclusions are valid provided the users associate themselves with the strongest BS.

In Fig. 5, the cumulative distribution of the DL rate of an arbitrary user in the reference BS is plotted for SP, RP, and staggered pilots for $M = 100$ antennas. The values of pilot reuse ratio for RP are $r^{\mathrm{RP}} = 1$, $r^{\mathrm{RP}} = 3$, and $r^{\mathrm{RP}} = 7$ in the plot. It can be observed that rates obtained from the channel estimate based on SP and staggered pilots are significantly higher than those obtained from RP with $r^{\mathrm{RP}} = 1$. Furthermore, it has to be noted that no additional UL overhead is required by staggered pilots and SP to achieve this DL throughput. However, while RP with $r^{\mathrm{RP}} = 3$ offers comparable performance to staggered pilots with $r^{\mathrm{SP}} = 7$, it has to be noted that, in addition to the increased UL overhead, pilot reuse with RP comes with the additional requirement that all users transmit their UL pilots simultaneously. This requirement will result in
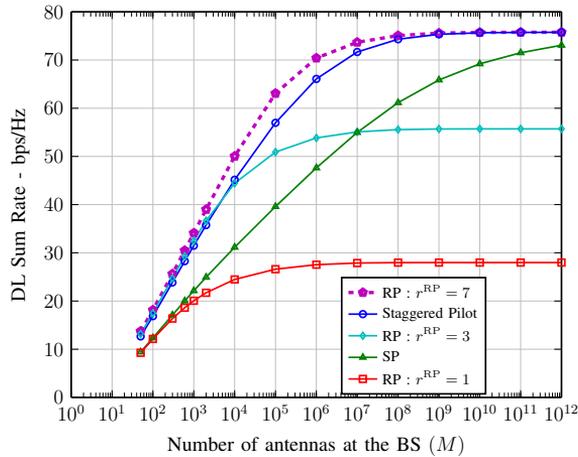
Fig. 6. DL Sum Rate vs M with MF precoder. SP and Staggered pilots offer an asymptotic DL throughput equivalent to that of RP with $r^{\mathrm{RP}} = 7$, even though the UL overhead is as much as that of RP with $r^{\mathrm{RP}} = 1$.
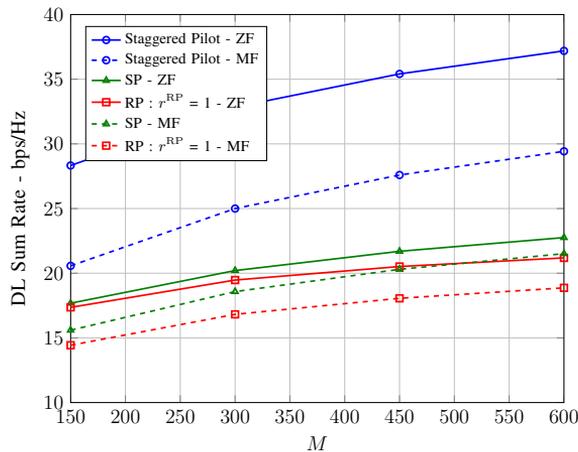
Fig. 7. DL Sum Rate vs M with ZF precoder.

pilot reuse capable of being implemented across only a few cells in the network, and therefore, the remaining cells can benefit from using either SP or staggered pilots.

In Fig. 6, the DL sum rate of staggered, RP, and SP is plotted against the number of antennas M when the channel estimates are used in an MF precoder. The DL rate ceiling of SP and staggered pilots is significantly higher than that of RP with $r^{\mathrm{RP}} = 1$. Moreover, the DL rate of staggered pilots is close to that of RP with $r^{\mathrm{RP}} = 7$ and SP achieves this rate asymptotically. The
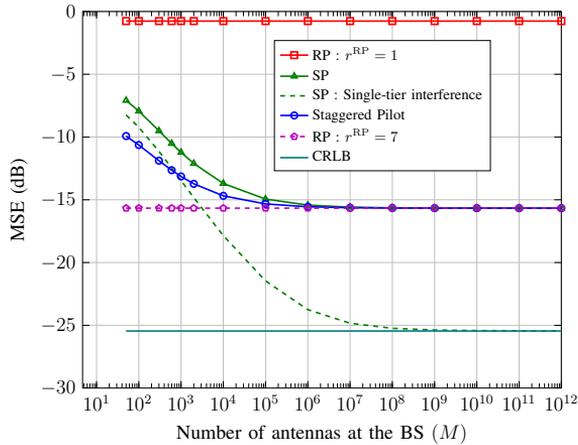
Fig. 8. NMSE vs M. Similar to Fig. 6, the asymptotic MSEs of SP and staggered pilots are equivalent to that of RP with $r^{\mathrm{RP}} = 7$. In addition, the component of interference from users that transmit a pilot orthogonal to that of the reference user reduces asymptotically to zero.

key difference between SP and staggered pilots is that in the former, the strongest interference seen by a particular user in the reference cell is due to the remaining users of that cell, which results from transmitting data alongside pilots. However, in the latter case, this component of interference is absent. Since the strongest component of the interference has been removed in the case of staggered pilots, they are capable of achieving a rate close to that of RP with $r^{\mathrm{RP}} = 7$. In addition, SP and staggered pilots have the same ceiling on the achievable rate, which is evident from the values of the achievable rate for large M.

In Fig. 7, the DL sum rates are plotted when the channel estimates are used in a ZF precoder which is given as

$$\boldsymbol{g}_{jm} = \widehat{\boldsymbol{H}}_j \left(\widehat{\boldsymbol{H}}_j^H \widehat{\boldsymbol{H}}_j \right)^{-1} \boldsymbol{e}_m \tag{44}$$

where $\widehat{\boldsymbol{H}}_j = \left[\widehat{\boldsymbol{h}}_{jj0}, \ldots, \widehat{\boldsymbol{h}}_{jj,K-1} \right]$ and $\boldsymbol{e}_m$ is the mth column of $\boldsymbol{I}_K$. The parameter $\nu_{\ell k}$ is chosen to constrain the instantaneous transmit power to 1. The sum rates are obtained numerically with $L = 7$ and $K = 5$. It can be observed that the behavior of SP and RP is similar to that in Fig. 6, whereas the gap between staggered pilots and SP/RP is significantly higher than that in Fig. 6.

In Fig. 8, the MSE of the channel estimate is plotted against M. Similar to the behavior in Fig. 6, the MSE of the channel estimate obtained from SP and staggered pilots asymptotically approaches the MSE of the estimate from RP with $r^{\mathrm{RP}} = 7$. In addition, since we have assumed

TABLE I

UL AND DL PERFORMANCE OF RP, SP, AND HYBRID SYSTEMS

	UL Sum Rate (bps/Hz)	DL Sum Rate(bps/Hz)	Total Rate(bps/Hz)
Hybrid System	48.12	84.94	133.07
RPs ($r^{\mathrm{RP}} = 1$)	50.42	65.85	116.27
SPs	35.30	75.02	110.32

that the interference from second and subsequent tiers of cells are negligible when deriving the CRLB, the interference from these cells results in a gap between the MSE of SP and the CRLB. In the absence of this component of the interference, it can be seen that the MSE of the channel estimate attains the CRLB asymptotically as $M \to \infty$.

B. Hybrid System

The hybrid system is simulated with $L = 19$ hexagonal cells, i.e., a central cell with two tiers of interfering cells. Each cell has $K = 5$ users and the values of C_u and C_d are both chosen as 40 symbols. Although L is set to 19, the partitioning of users and the computation of the performance metrics is performed over 7 cells which consist of the central and the first tier of cells. The weights ξ^{ul} and ξ^{dl} are both set to 0.5. The value of ω for the users in $\mathcal{U}_{\mathrm{SP}}$ is set to 10 and μ for the users in $\mathcal{U}_{\mathrm{RP}}$ is set to 1.

For obtaining Figs. 9 and 10, the users are assumed to be distributed uniformly on a circle around the BS. Then, the sum rates in the UL and DL are plotted in the figures against the radius of the circle around the BS. As can be observed in Figs. 9 and 10, when the user radius is smaller than 0.6, RP are superior both in the UL and DL. However, SP are superior both in the UL and DL, when the user radius is larger than 0.8. Therefore, for user radius in the ranges $[0, 0.6]$ the greedy algorithm chooses RP and it chooses SP in the range $[0.8, 1]$. However, in the range $[0.6, 0.8]$, RP offer a better performance in the UL but a poorer performance in the DL, with respect to SP. Therefore, the choices of $\mathcal{U}_{\mathrm{RP}}$ and $\mathcal{U}_{\mathrm{SP}}$ are determined by ξ^{ul} and ξ^{dl}. Since for these simulations ξ^{ul} and ξ^{dl} are both chosen as 0.5, the greedy algorithm attempts to strike a balance between the UL and DL throughputs and offers a total throughput that is in between that of the systems that employ only RP or SP. In addition, since the algorithm is greedy, the
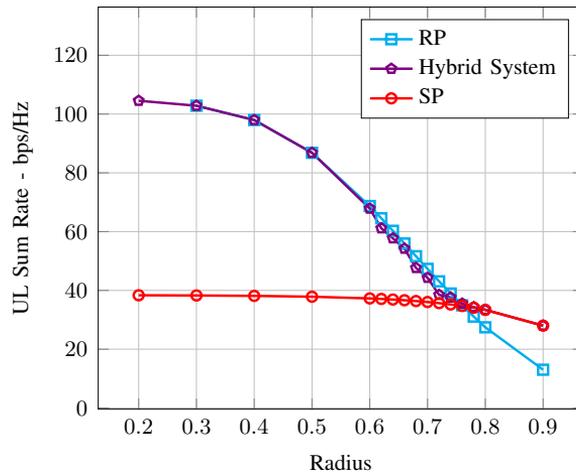
Fig. 9. Sum rate in the UL over users in the first tier of cells vs. user radius. The algorithm assigns users RP and SP in scenarios with low and high interference, respectively.
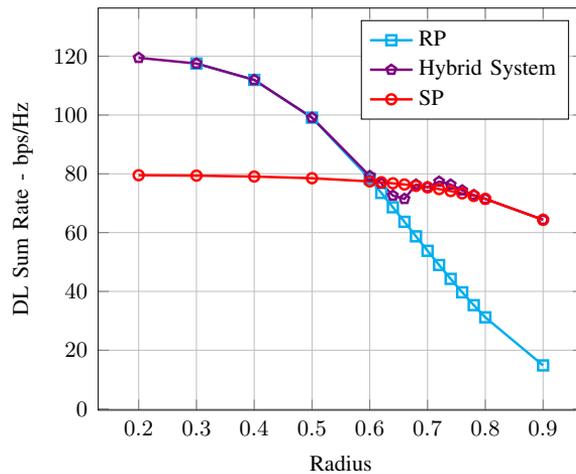
Fig. 10. Sum rate in the DL over users in the first tier of cells vs. user radius. The non-smooth nature of the DL rate of the hybrid system in the range of user radius $[0.6, 0.8]$ is due to the suboptimal nature of the greedy algorithm.

variation of this throughput with respect to the user radius is non-smooth in nature, as can be seen in the figures.

Table I details the UL and DL performance of a system with users transmitting RP, SP, and a hybrid of both, when the users are uniformly distributed across the cells. The hybrid system

offers roughly 14.44% higher total rate than the system that employs only RP. Moreover, both SP and the hybrid system offer a significantly higher throughput in the DL, albeit at the cost of a lower UL throughput than when compared with RP. However, the hybrid system enables controlling the trade-off between the UL and DL throughputs using the weights ξ^{ul} and ξ^{dl}.

It has to be noted that there is an important difference between the results in Section VI-B and those in [20], [24]. In the latter, the computed rates are approximate for finite M, since the correlation between the signal and interference components have been ignored and approximated to be zero. However, using the approach in Appendix A, the signal and interference terms are uncorrelated and both the UL and DL rates shown in Figs. 9 and 10, and Table I are lower bounds on the achievable rates.

VII. Conclusion

We have shown that SPs offer a significantly better asymptotic MSE and DL performance than RPs. This improvement is attributed to utilizing the array gain of the antenna for reducing the fraction of UL power allocated to data in favor of allocating a larger fraction of power for pilot transmission. We have also shown that staggered pilots are a particular case of SPs and therefore, offer a DL performance similar to that of SPs, while offering the same UL spectral and energy efficiency as RPs. Furthermore, we have shown that higher asymptotic DL throughput offered by SPs and staggered pilots are at the same or lower UL transmission overhead than RPs. We have also extended the hybrid system to partition the users into two disjoint sets of users that use RPs and SPs by minimizing both the UL and DL interference. We show, by means of simulation, that the hybrid system offers a higher throughput than when only RPs or SPs are employed.

APPENDIX A

Downlink SINR for Channel Estimate Based on SP Pilots

For MF precoding, $\boldsymbol{g}_{\ell k} = \widehat{\boldsymbol{h}}_{\ell\ell k}$. Then,

$$\mathbb{E}\left\{\boldsymbol{h}_{jjm}^{H}\boldsymbol{g}_{jm}\right\} = M\beta_{jjm} \tag{45}$$

$$\mathbb{E}\left\{|\boldsymbol{h}_{\ell jm}^{H}\boldsymbol{g}_{\ell k}|^{2}\right\} = M^{2}\beta_{\ell jm}^{2}\frac{\mu_{jm}}{\mu_{\ell m}}\mathbf{1}_{\{k=m,j\in\mathcal{L}_{\ell}(r^{\mathrm{SP}})\}} + M\sum_{n\in\mathcal{L}_{\ell}(r^{\mathrm{SP}})}\frac{\mu_{nk}}{\mu_{\ell k}}\beta_{\ell jm}\beta_{\ell nk}$$

$$+ M\sum_{n=0}^{L-1}\sum_{p=0}^{K-1}\frac{\rho^{2}}{C_{u}\lambda^{2}}\frac{\mu_{np}}{\mu_{\ell k}}\beta_{\ell np}\beta_{\ell jm} + \frac{M^{2}\rho^{2}}{C_{u}\lambda^{2}}\frac{\mu_{jm}}{\mu_{\ell k}}\beta_{\ell jm}^{2} + \frac{M\sigma^{2}}{C_{u}\lambda^{2}\mu_{\ell k}} \tag{46}$$

$$\frac{1}{\nu_{\ell k}} = \frac{\mathbb{E}\left\{\|\boldsymbol{g}_{\ell k}\|^{2}\right\}}{q_{\ell k}} = \frac{M}{q_{\ell k}}\left(\sum_{n\in\mathcal{L}_{\ell}(r^{\mathrm{SP}})}\frac{\mu_{nk}}{\mu_{\ell k}}\beta_{\ell nk} + \sum_{n=0}^{L-1}\sum_{p=0}^{K-1}\frac{\rho^{2}}{C_{u}\lambda^{2}}\frac{\mu_{np}}{\mu_{\ell k}}\beta_{\ell np} + \frac{\sigma^{2}}{C_{u}\lambda^{2}\mu_{\ell k}}\right) \tag{47}$$

Substituting (45) to (47) into (15), the SINR can be obtained as

$$\mathrm{SINR}_{jm}^{\mathrm{SP-dl}} = \nu_{jm}\beta_{jjm}^{2}\left(\sum_{\ell\in\mathcal{L}_{j}(r^{\mathrm{SP}})\backslash j}\frac{\mu_{jm}}{\mu_{\ell m}}\nu_{jm}\beta_{\ell jm}^{2} + \frac{1}{M}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\frac{\nu_{\ell k}\beta_{\ell jm}}{\mu_{\ell k}}\left(\sum_{n\in\mathcal{L}_{\ell}(r^{\mathrm{SP}})}\mu_{nk}\beta_{\ell nk} + \frac{\sigma^{2}}{C_{u}\lambda^{2}}\right)\right.$$

$$\left. + \frac{\rho^{2}}{C_{u}\lambda^{2}}\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\frac{\nu_{\ell k}}{\mu_{\ell k}}\left(\mu_{jm}\beta_{\ell jm}^{2} + \frac{1}{M}\sum_{n=0}^{L-1}\sum_{p=0}^{K-1}\mu_{np}\beta_{\ell np}\beta_{\ell jm}\right) + \frac{1}{M^{2}\gamma}\right)^{-1} \tag{48}$$

This completes the derivation of (16).

APPENDIX B

Calculation of ρ_{opt}^{2} and $\lambda_{\mathrm{opt}}^{2}$

Equation (1) can be written as

$$\widetilde{\boldsymbol{Y}}_{j} = \sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\sqrt{\mu_{\ell k}}\boldsymbol{h}_{j\ell k}\left(\lambda\boldsymbol{p}_{\ell k} + \rho\widetilde{\boldsymbol{x}}_{\ell k}\right)^{T} + \widetilde{\boldsymbol{W}}_{j} \tag{49}$$

where $\widetilde{\boldsymbol{x}}_{\ell k} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I})$. The receiver applies the following linear invertible transformation to the received observation

$$\boldsymbol{Y}_{j} \triangleq \frac{1}{\sqrt{C_{u}}}\widetilde{\boldsymbol{Y}}_{j}\boldsymbol{P}^{*} = \sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}\sqrt{\mu_{\ell k}}\boldsymbol{h}_{j\ell k}\left(\lambda\sqrt{C_{u}}\boldsymbol{e}_{\ell k} + \rho\boldsymbol{x}_{\ell k}\right)^{T} + \boldsymbol{W}_{j} \tag{50}$$

where $\boldsymbol{e}_{\ell k} = \boldsymbol{P}^{H}\boldsymbol{p}_{\ell k}/C_{u}$ has ones in locations corresponding to the column index of $\boldsymbol{p}_{\ell k}$ in $\boldsymbol{P}$ and zeros elsewhere, $\boldsymbol{x}_{\ell k} = \boldsymbol{P}^{H}\widetilde{\boldsymbol{x}}_{\ell k}/\sqrt{C_{u}}$, and $\boldsymbol{W}_{j} = \widetilde{\boldsymbol{W}}\boldsymbol{P}^{H}/\sqrt{C_{u}}$. Note that the distributions of $\boldsymbol{x}_{\ell k}$ and $\boldsymbol{W}_{j}$ are unchanged since $\boldsymbol{P}/\sqrt{C_{u}}$ is a unitary matrix. Therefore, from the perspective of calculating the achievable rate, both (49) and (50) are equivalent. We simplify notation by

dropping the subscript j and replacing the tuple (ℓ, k) with a single index k such that $0 \leq k \leq N - 1$, where $N \triangleq LK - 1$. Defining $\boldsymbol{y}_t$ as the tth vector of received observations and x_{nt} as the tth element of $\boldsymbol{x}_n$, $\boldsymbol{y}_t$ for $0 \leq t \leq C_u - 1$ can be written as

$$\boldsymbol{y}_t = \sum_{k \in \mathcal{P}_t} \lambda \sqrt{\mu_k C_u} \boldsymbol{h}_k + \sum_{k=0}^{N-1} \rho \sqrt{\mu_k} \boldsymbol{h}_k x_{kt} + \boldsymbol{w}_t \tag{51}$$

where $\mathcal{P}_t = \{k \in \{0, \ldots, LK - 1\} \mid \boldsymbol{p}_k = \boldsymbol{p}_t\}$ is the set of all users that transmit their pilot in symbol t, and $\mathcal{P}_t = \varnothing$ for $t \geq LK$.

Without loss of generality, let user 0 be the reference user, and let this user transmit its pilot in symbol 0 in (51). The LS estimate of the channel of user 0 can then be written as

$$\widehat{\boldsymbol{h}}_0 = \frac{1}{\sqrt{\mu_0} \sqrt{C_u} \lambda} \boldsymbol{y}_0 = \sum_{k \in \mathcal{P}_0} \sqrt{\frac{\mu_k}{\mu_0}} \boldsymbol{h}_k + \sum_{k=0}^{N-1} \frac{\rho}{\sqrt{C_u} \lambda} \sqrt{\frac{\mu_k}{\mu_0}} \boldsymbol{h}_k x_{k0} + \frac{\boldsymbol{w}_t}{\sqrt{C_u \mu_0} \lambda} . \tag{52}$$

Then, the output of the MRC when $t \geq 1$ is

$$\widehat{x}_{0t} = \widehat{\boldsymbol{h}}_0^H \boldsymbol{y}_t = \rho \sqrt{\mu_0} \mathbb{E} \left\{ \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0 \right\} x_{0t} + \rho \sqrt{\mu_0} \left(\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0 - \mathbb{E} \left\{ \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0 \right\} \right) x_{0t} + \sum_{k \in \mathcal{P}_t} \lambda \sqrt{\mu_k C_u} \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k$$

$$+ \sum_{k=1}^{N-1} \rho \sqrt{\mu_k} \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k x_{kt} + \widehat{\boldsymbol{h}}_0^H \boldsymbol{w}_t . \tag{53}$$

Since x_{0t} when $t \geq 1$ is independent of x_{n0} for all $0 \leq n \leq LK - 1$, x_{0t} is independent of $\widehat{\boldsymbol{h}}_0$. As a result, in (53), the first term is uncorrelated with the remaining terms. Then, (53) can be written as $\widehat{x}_{0t} = s_t + i_t$ where

$$s_t \triangleq \rho \sqrt{\mu_0} \mathbb{E} \left\{ \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0 \right\} x_{0t} \tag{54}$$

$$i_t \triangleq \rho \sqrt{\mu_0} \left(\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0 - \mathbb{E} \left\{ \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0 \right\} \right) x_{0t} + \sum_{k \in \mathcal{P}_t} \lambda \sqrt{\mu_k C_u} \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k + \sum_{k=1}^{N-1} \rho \sqrt{\mu_k} \widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k x_{kt} + \widehat{\boldsymbol{h}}_0^H \boldsymbol{w}_t . \tag{55}$$

Defining $\Psi(x) = \log_2(1 + x)$, a lower bound on the ergodic capacity for user 0 in symbol t can be obtained as [29]

$$R_{0t} = \Psi \left(\frac{\mathbb{E}\{|s_t|^2\}}{\mathbb{E}\{|i_t|^2\}} \right) . \tag{56}$$

Since $s \triangleq \mathbb{E}\{|s_t|^2\}$ is independent of t for $t \geq 1$ and since $\log_2(1 + 1/x)$ is convex in x, an achievable lower bound on the capacity can be obtained using Jensen's inequality as

$$R_0 = \frac{1}{C_u} \sum_{t=0}^{C_u-1} \Psi \left(\frac{s}{\mathbb{E}\{|i_t|^2\}} \right) \geq \frac{1}{C_u} \sum_{t=1}^{C_u-1} \Psi \left(\frac{s}{\mathbb{E}\{|i_t|^2\}} \right) \geq \frac{C_u - 1}{C_u} \Psi \left(\frac{s}{\frac{1}{C_u-1} \sum_{t=1}^{C_u-1} \mathbb{E}\{|i_t|^2\}} \right) \tag{57}$$

where, in the first inequality, the throughput in symbol 0 in which user 0 transmits both pilot and data is ignored. Jensen's inequality is applied in the second inequality to render the right-hand side independent of t. Now,

$$i \triangleq \frac{1}{C_u - 1} \sum_{t=1}^{C_u - 1} \mathbb{E}\left\{|i_t|^2\right\} = \sum_{k=0}^{N-1} \rho^2 \mu_k \mathbb{E}\left\{|\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k|^2\right\} - \rho^2 \mu_0 \left|\mathbb{E}\left\{\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0\right\}\right|^2$$

$$+ \frac{1}{C_u - 1} \sum_{t=1}^{C_u - 1} \sum_{k \in \mathcal{P}_t} \mu_k \lambda^2 C_u \mathbb{E}\left\{|\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k|^2\right\} + \mathbb{E}\left\{|\widehat{\boldsymbol{h}}_0^H \boldsymbol{w}_t|^2\right\}$$

$$\leq \frac{C_u}{C_u - 1} \sum_{k=0}^{N-1} \mu_k \mathbb{E}\left\{|\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k|^2\right\} - \rho^2 \mu_0 \left|\mathbb{E}\left\{\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_0\right\}\right|^2 + \mathbb{E}\left\{|\widehat{\boldsymbol{h}}_0^H \boldsymbol{w}_t|^2\right\} \quad (58)$$

where, to obtain the inequality, we have used the property that $\bigcup_{t=1}^{C_u - 1} \mathcal{P}_t = \Omega \backslash \mathcal{P}_0 \subset \Omega$ with $\Omega \triangleq \{0, \ldots, N-1\}$ being the set of all users, and that $C_u/(C_u - 1) > 1$. Using (52), $\mathbb{E}\left\{|\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k|^2\right\}$ in (58) can be obtained as

$$\mathbb{E}\left\{|\widehat{\boldsymbol{h}}_0^H \boldsymbol{h}_k|^2\right\} = \begin{cases} v_k & k \notin \mathcal{P}_0 \\ M^2 \beta_k^2 \frac{\mu_k}{\mu_0} + v_k & k \in \mathcal{P}_0 \end{cases} \quad (59)$$

where

$$v_k \triangleq M^2 \rho^2 \left(\frac{c_k}{C_u \lambda^2} + \frac{1}{M \rho^2} d_k + \frac{1}{M \rho^2 \lambda^2} e_k\right) \quad (60)$$

$$c_k \triangleq \frac{\mu_k}{\mu_0} \beta_k^2 + \frac{1}{M} \sum_{n=0}^{N-1} \frac{\mu_n}{\mu_0} \beta_n \beta_k \ , \ d_k \triangleq \sum_{n \in \mathcal{P}_0} \frac{\mu_n}{\mu_0} \beta_n \beta_k \ , \ e_k \triangleq \frac{\beta_k \sigma^2}{\mu_0 C_u} \quad (61)$$

Substituting (60) into (58) and noting that $\boldsymbol{w}_t$ is independent of $\widehat{\boldsymbol{h}}_0$, we obtain

$$i \leq M^2 \rho^2 \left[\frac{\rho^2}{\lambda^2} \alpha_1 + \frac{\rho^2}{M \lambda^2} \alpha_2 + \frac{\lambda^2}{\rho^2} \frac{\alpha_3}{M} + \frac{1}{\rho^2 \lambda^2} \frac{\alpha_4}{M} + \frac{1}{\rho^2} \frac{\alpha_5}{M} + \frac{1}{\lambda^2} \frac{\alpha_6}{M} + \alpha_7\right] \quad (62)$$

where

$$\alpha_1 = \sum_{k=0}^{N-1} \frac{\mu_k^2 \beta_k^2}{\mu_0 (C_u - 1)} \ , \ \alpha_2 = \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} \frac{\mu_k \mu_n \beta_k \beta_n}{\mu_0 (C_u - 1)} \ ; \ \alpha_3 = \sum_{k=0}^{N-1} \frac{\mu_k d_k C_u}{(C_u - 1)} \quad (63)$$

$$\alpha_4 = \sum_{k=0}^{N-1} \frac{\mu_k e_k C_u}{C_u - 1} + \frac{\sigma^4}{\mu_0 C_u} \ ; \ \alpha_5 = \sum_{k \in \mathcal{P}_0} \sigma^2 \frac{\mu_k}{\mu_0} \beta_k \ ; \ \alpha_6 = \sum_{k=0}^{N-1} \frac{\sigma^2}{C_u} \frac{\mu_k}{\mu_0} \beta_k \quad (64)$$

$$\alpha_7 = \sum_{k \in \mathcal{P}_0 \backslash 0} \frac{\mu_k^2}{\mu_0} \beta_k^2 + \frac{C_u}{C_u - 1} \sum_{k=0}^{N-1} \mu_k \left(\frac{c_k}{C_u} + \frac{1}{M} d_k\right) \ . \quad (65)$$

Given that $s = \mathbb{E}\{|s_t|^2\} = \mu_0\rho^2 M^2\beta_0^2$, substituting (62) into (57), a lower bound on the UL ergodic capacity is obtained as

$$R_0 = \frac{C_u - 1}{C_u}\Psi\left(\frac{\mu_0\beta_0^2}{\frac{\rho^2}{\lambda^2}\alpha_1 + \frac{\rho^2}{M\lambda^2}\alpha_2 + \frac{\lambda^2}{\rho^2}\frac{\alpha_3}{M} + \frac{1}{\rho^2\lambda^2}\frac{\alpha_4}{M} + \frac{1}{\rho^2}\frac{\alpha_5}{M} + \frac{1}{\lambda^2}\frac{\alpha_6}{M} + \alpha_7}\right) \qquad (66)$$

which is maximized when the denominator inside $\Psi\left(\cdot\right)$ is minimized. To obtain ρ_{opt}^2, we set $\lambda^2 = 1 - \rho^2$ in (66), differentiate the denominator with respect to ρ^2, and set the result to zero. We then get,

$$\rho_{\text{opt}}^2 = \left(1 + \sqrt{M}\kappa\right)^{-1}, \quad \lambda_{\text{opt}}^2 = \left(1 + \frac{1}{\kappa\sqrt{M}}\right)^{-1}. \qquad (67)$$

where

$$\kappa \triangleq \sqrt{\frac{\alpha_1 + \frac{\alpha_2}{M} + \frac{\alpha_4}{M} + \frac{\alpha_6}{M}}{\alpha_3 + \alpha_4 + \alpha_5}}. \qquad (68)$$

APPENDIX C

CRLB for Channel Estimates Obtained From SP Pilots

To derive the CRLB, the received signal when using SP pilots can be written as [20]

$$\boldsymbol{Y} = \boldsymbol{H}_d\left(\rho\boldsymbol{X}_d + \lambda\boldsymbol{P}_d\right) + \boldsymbol{H}_i\left(\rho\boldsymbol{X}_i + \lambda\boldsymbol{P}_i\right) + \boldsymbol{W} \qquad (69)$$

where $\boldsymbol{H}_d \triangleq [\boldsymbol{h}_{j,j,0}, \ldots, \boldsymbol{h}_{j,j,K-1}]$ are the channel vectors of the desired users and $\boldsymbol{X}_d \triangleq [\boldsymbol{x}_{j,0}, \ldots, \boldsymbol{x}_{j,K-1}]^T$ are the data symbols from the desired users. Similarly, $\boldsymbol{H}_i \in \mathbb{C}^{M\times(r^{\text{SP}}-1)K}$ and $\boldsymbol{X}_i \in \mathbb{C}^{C_u\times(r^{\text{SP}}-1)K}$ are the data and channel vectors, respectively, of the interfering users. The subscript j has been dropped from $\boldsymbol{H}_d$, $\boldsymbol{H}_i$, $\boldsymbol{X}_d$, $\boldsymbol{X}_i$, and $\boldsymbol{W}$ for notational convenience. In addition, the UL transmit power $\mu_{\ell k}$ for each user is assumed to be absorbed into $\beta_{j\ell k}$. The vectorized form of (69) can be written as

$$\begin{aligned}\overline{\boldsymbol{y}} = \text{vec}\left(\boldsymbol{Y}\right) &= \left(\boldsymbol{I}_{C_u} \otimes \boldsymbol{H}_d\right)\left(\rho\overline{\boldsymbol{x}}_d + \lambda\overline{\boldsymbol{p}}_d\right) + \left(\boldsymbol{I}_{C_u} \otimes \boldsymbol{H}_i\right)\left(\rho\overline{\boldsymbol{x}}_i + \lambda\overline{\boldsymbol{p}}_i\right) + \overline{\boldsymbol{w}} \\ &= \left(\left(\rho\boldsymbol{X}_d + \lambda\boldsymbol{P}_d\right)^T \otimes \boldsymbol{I}_M\right)\overline{\boldsymbol{h}}_d + \left(\left(\rho\boldsymbol{X}_i + \lambda\boldsymbol{P}_i\right)^T \otimes \boldsymbol{I}_M\right)\overline{\boldsymbol{h}}_i + \overline{\boldsymbol{w}} \qquad (70)\end{aligned}$$

where the over-bar denotes the vec operation, i.e., $\overline{\boldsymbol{x}} \triangleq \text{vec}(\boldsymbol{X})$ and the property $\text{vec}\left(\boldsymbol{AB}\right) = \left(\boldsymbol{I}_m \otimes \boldsymbol{A}\right)\overline{\boldsymbol{b}} = \left(\boldsymbol{B}^T \otimes \boldsymbol{I}_n\right)\overline{\boldsymbol{a}}$ has been used.

For the set of unknown parameters $\boldsymbol{\theta} \triangleq \left\{ \overline{\boldsymbol{x}}_d, \overline{\boldsymbol{x}}_i, \overline{\boldsymbol{h}}_d, \overline{\boldsymbol{h}}_i \right\}$, the Fischer information matrix can be defined as [34]

$$\mathcal{J}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{Y},\boldsymbol{\theta}} \left\{ \left[\frac{\partial \ln p(\boldsymbol{Y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right] \left[\frac{\partial \ln p(\boldsymbol{Y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right]^H \right\}$$

$$= \mathbb{E}_{\boldsymbol{\theta}} \left\{ \boldsymbol{J}_{\boldsymbol{\theta}\boldsymbol{\theta}^H} \right\} + \mathbb{E}_{\boldsymbol{\theta}} \left\{ \left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right] \left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right]^H \right\} . \tag{71}$$

where

$$\boldsymbol{J}_{\boldsymbol{\theta}\boldsymbol{\theta}^H} \triangleq \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\theta}} \left\{ \left[\frac{\partial \ln p(\boldsymbol{Y}\mid\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right] \left[\frac{\partial \ln p(\boldsymbol{Y}\mid\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right]^H \bigg| \boldsymbol{\theta} \right\} . \tag{72}$$

Using (70), $\boldsymbol{J}_{\boldsymbol{\theta}\boldsymbol{\theta}^H}$ can be written as

$$\boldsymbol{J}_{\boldsymbol{\theta}\boldsymbol{\theta}^H} = \begin{bmatrix} \boldsymbol{J}_{x_d}\boldsymbol{J}_{x_d}^H & \boldsymbol{J}_{x_d}\boldsymbol{J}_{x_i}^H & \boldsymbol{J}_{x_d}\boldsymbol{J}_{h_d}^H & \boldsymbol{J}_{x_d}\boldsymbol{J}_{h_i}^H \\[6pt] \boldsymbol{J}_{x_i}\boldsymbol{J}_{x_d}^H & \boldsymbol{J}_{x_i}\boldsymbol{J}_{x_i}^H & \boldsymbol{J}_{x_i}\boldsymbol{J}_{h_d}^H & \boldsymbol{J}_{x_i}\boldsymbol{J}_{h_i}^H \\[6pt] \boldsymbol{J}_{h_d}\boldsymbol{J}_{x_d}^H & \boldsymbol{J}_{h_d}\boldsymbol{J}_{x_i}^H & \boldsymbol{J}_{h_d}\boldsymbol{J}_{h_d}^H & \boldsymbol{J}_{h_d}\boldsymbol{J}_{h_i}^H \\[6pt] \boldsymbol{J}_{h_i}\boldsymbol{J}_{x_d}^H & \boldsymbol{J}_{h_i}\boldsymbol{J}_{x_i}^H & \boldsymbol{J}_{h_i}\boldsymbol{J}_{h_d}^H & \boldsymbol{J}_{h_i}\boldsymbol{J}_{h_i}^H \end{bmatrix} \tag{73}$$

where

$$\boldsymbol{J}_{x_d} \triangleq \frac{\rho}{\sigma} \left(\boldsymbol{I}_{C_u} \otimes \boldsymbol{H}_d \right)^H , \quad \boldsymbol{J}_{x_i} \triangleq \frac{\rho}{\sigma} \left(\boldsymbol{I}_{C_u} \otimes \boldsymbol{H}_i \right)^H \tag{74}$$

$$\boldsymbol{J}_{h_d} \triangleq \frac{1}{\sigma} \left(\left(\rho \boldsymbol{X}_d + \lambda \boldsymbol{P}_d \right)^T \otimes \boldsymbol{I}_M \right)^H , \quad \boldsymbol{J}_{h_i} \triangleq \frac{1}{\sigma} \left(\left(\rho \boldsymbol{X}_i + \lambda \boldsymbol{P}_i \right)^T \otimes \boldsymbol{I}_M \right)^H . \tag{75}$$

Using (73) to (75), the first term in (71) can be expressed as

$$\mathbb{E}_{\boldsymbol{\theta}} \left[\boldsymbol{J}_{\boldsymbol{\theta}\boldsymbol{\theta}^H} \right] = \frac{M\rho^2}{\sigma^2} \text{blkdiag} \left[\boldsymbol{I}_{C_u} \otimes \boldsymbol{D}_d , \ \boldsymbol{I}_{C_u} \otimes \boldsymbol{D}_i , \ \frac{C_u}{M\rho^2} \boldsymbol{I}_{MK} , \ \frac{C_u}{M\rho^2} \boldsymbol{I}_{M(N-K)} \right] \tag{76}$$

where $\boldsymbol{D}_d \triangleq \text{diag}\left\{ \beta_{j,j,0}, \ldots, \beta_{j,j,K-1} \right\}$ is the diagonal matrix containing the path-loss coefficients of the desired users and $\boldsymbol{D}_i$ is the diagonal matrix containing the path-loss coefficients of the interfering users. The second term in (71) can be found as

$$\mathbb{E}_{\boldsymbol{\theta}} \left\{ \left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right] \left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right]^H \right\} = \text{blkdiag} \left[\boldsymbol{\Gamma}_{x_d} , \ \boldsymbol{\Gamma}_{\boldsymbol{x}_i} , \ \left(\boldsymbol{I}_M \otimes \boldsymbol{D}_d \right)^{-1} , \ \left(\boldsymbol{I}_M \otimes \boldsymbol{D}_i \right)^{-1} \right] \tag{77}$$

where

$$\boldsymbol{\Gamma}_{x_d} \triangleq \mathbb{E}_{\boldsymbol{x}_d} \left\{ \left[\frac{\partial \ln p_{\boldsymbol{x}_d}(\boldsymbol{x}_d)}{\partial \boldsymbol{x}_d^*} \right] \left[\frac{\partial \ln p_{\boldsymbol{x}_d}(\boldsymbol{x}_d)}{\partial \boldsymbol{x}_d^*} \right]^H \right\} \tag{78}$$

$$\boldsymbol{\Gamma}_{x_i} \triangleq \mathbb{E}_{\boldsymbol{x}_i} \left\{ \left[\frac{\partial \ln p_{\boldsymbol{x}_i}(\boldsymbol{x}_i)}{\partial \boldsymbol{x}_i^*} \right] \left[\frac{\partial \ln p_{\boldsymbol{x}_i}(\boldsymbol{x}_i)}{\partial \boldsymbol{x}_i^*} \right]^H \right\} . \tag{79}$$

Therefore, the CRLB for the parameter $\boldsymbol{H}_d$ and the channel vector $\boldsymbol{h}_{j,j,k}$ are given as

$$\text{CRLB}\left(\boldsymbol{H}_d\right) = \text{trace}\left\{\left(\frac{C_u}{\sigma^2}\boldsymbol{I}_{KM} + \left(\boldsymbol{I}_M \otimes \boldsymbol{D}_d\right)^{-1}\right)^{-1}\right\} \tag{80}$$

$$\text{CRLB}\left(\boldsymbol{h}_{j,j,m}\right) = \text{trace}\left\{\left(\frac{C_u}{\sigma^2}\boldsymbol{I}_M + \frac{1}{\beta_{j,j,m}}\boldsymbol{I}_M\right)^{-1}\right\} = \frac{M}{\frac{C_u}{\sigma^2} + \frac{1}{\beta_{j,j,m}}} \ . \tag{81}$$

This completes the derivation of (11).

REFERENCES

[1] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[4] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.

[6] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[7] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[8] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[9] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

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

[11] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[12] K. Upadhya and S. A. Vorobyov, "An array processing approach to pilot decontamination for massive MIMO," in *Proc. IEEE 6th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Dec. 2015, pp. 453–456.

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

[14] E. Björnson, J. Hoydis, and L. Sanguinetti, "Pilot contamination is not a fundamental asymptotic limitation in massive mimo," in *IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[15] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, "On an achievable rate of large rayleigh block-fading MIMO channels with no CSI," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6517–6541, Oct. 2013.

[16] S. He, J. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3007–3021, Jun. 2007.

[17] M. Coldrey and P. Bohlin, "Training-based MIMO systems – part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.

[18] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.

[19] T. Cui and C. Tellambura, "Pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM), St. Louis*, vol. 4, Dec. 2005, pp. 5 pp.–2233.

[20] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917–2932, Jun. 2017.

[21] D. Verenzuela, E. Björnson, and L. Sanguinetti, "Spectral and energy efficiency of superimposed pilots in uplink massive MIMO," *arXiv Preprint*, 2017. [Online]. Available: http://arxiv.org/abs/1709.07722

[22] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multi-cell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, Mar. 2016.

[23] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Downlink performance of superimposed pilots in massive MIMO systems in the presence of pilot contamination," in *Proc. IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, Washington D.C., Dec. 2016, pp. 665–669.

[24] ——, "Time-multiplexed / superimposed pilot selection for massive MIMO pilot decontamination," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LO, Mar. 2017, pp. 3459 – 3463.

[25] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[26] E. Bjornson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[27] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[28] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 540–553, Jan 2018.

[29] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[30] I. Viering, H. Hofstetter, and W. Utschick, "Spatial long-term variations in urban, rural and indoor environments," in *the 5th Meeting of COST 273*, Lisbon, Portugal, Sep. 2002.

[31] ——, "Validity of spatial covariance matrices over time and frequency," in *Proc. Global Telecommun. Conf.*, Taipei, Taiwan, Nov. 2002, pp. 851–855.

[32] D. Kong, D. Qu, K. Luo, and T. Jiang, "Channel estimation under staggered frame structure for massive MIMO system," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1469–1479, Feb. 2016.

[33] W. A. W. M. Mahyiddin, P. A. Martin, and P. J. Smith, "Performance of synchronized and unsynchronized pilots in finite massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6763–6776, Dec. 2015.

[34] A. Van Den Bos, "A Cramér-Rao lower bound for complex parameters," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2859–, Oct. 1994.

Publication VI

Covariance Matrix Estimation for Massive MIMO

Karthik Upadhya [ID], *Student Member, IEEE* and Sergiy A. Vorobyov [ID], *Fellow, IEEE*

Abstract—We propose a novel pilot structure for covariance matrix estimation in massive multiple-input multiple-output systems in which each user transmits two pilot sequences, with the second pilot sequence multiplied by a random phase shift. The covariance matrix of a particular user is obtained by computing the sample cross-correlation of the channel estimates obtained from the two pilot sequences. This approach relaxes the requirement that all the users transmit their uplink pilots over the same set of symbols. We derive expressions for the achievable rate and the mean-squared error of the covariance matrix estimate when the proposed method is used with staggered pilots. The performance of the proposed method is compared with existing methods through simulations.

Index Terms—Covariance estimation, massive multiple-input multiple-output (MIMO), pilot contamination, staggered pilots.

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) is a variation of multiuser MIMO (MU-MIMO) that has a large number of antennas at the base station (BS), which significantly improves the spectral efficiency through spatial multiplexing [1]–[4] at a low cost of simple linear processing at the BS [1], [5], [6]. However, in practice, the BS needs to obtain channel state information using pilots, which have to be reused in different cells, thereby causing pilot contamination [1]. It has been argued that pilot contamination, in independent and identically distributed (i.i.d) Rayleigh fading, puts a fundamental limit on the asymptotically achievable rate in massive MIMO systems [1], and pilot decontamination algorithms have been designed in many works (see [7]–[12] to mention just a few).

It has been recently shown in [13] that the ceiling on the uplink (UL) and downlink (DL) rates due to pilot contamination can be eliminated under certain loose conditions on the covariance matrices of the users. However, this method requires estimates of the covariance matrix at the BS, which have to be obtained from observations that are made in the presence of pilot contamination. In [14], two methods have been developed wherein the users are assigned unique pilots specifically for estimating the covariance matrices. In [15], a method for jointly performing pilot allocation and estimating the covariance matrix has been proposed. A method for estimating a low-rank covariance matrix has also been proposed in [16]. A common theme in all the earlier works is that they require/assume that the users in

all cells transmit their UL pilots simultaneously, which can be infeasible in practice.

In this letter, we develop a method for estimating the users' covariance matrices using a pair of pilot sequences, with the second pilot sequence multiplied by a random phase shift. The quality of the covariance estimate obtained through the proposed method using staggered pilots is quantified using expressions for its mean-squared error (MSE). The achievable rate is analyzed numerically when the covariance matrices are estimated and used with staggered and regular pilots.

Notation: A vector is denoted as $\mathbf{a}$ and a matrix as $\mathbf{A}$. The notations $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, $(\cdot)^{-1}$, blkdiag $\{\cdot\}$, trace $(\cdot)$, $\lfloor x \rfloor$ represent the transpose, conjugate, conjugate transpose, matrix inverse, block diagonal matrix, matrix trace, and largest integer smaller than x, respectively, whereas $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. System Model and Problem Description

We consider a massive MU-MIMO system with L cells each having M antennas at the BS and containing K users. Denoting a user u in cell c as (c, u), the channel vector between user (c, u) and BS j is represented as $\mathbf{h}_{jcu} \in \mathbb{C}^M$ and is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{jcu})$. The channel is assumed to be constant for C symbols and the second-order statistics $\mathbf{R}_{jcu}$ are assumed to be constant for τ_s blocks each containing C symbols. The coherence time of the channel is divided into C_u and C_d symbols for the UL and DL time-slots, respectively.

In [13] and [17], it was shown that the UL and DL rates increase asymptotically in M when the data are estimated using a linear minimum MSE (LMMSE) or zero-forcing precoder/combiner that is designed using the LMMSE channel estimate. If $\widehat{\mathbf{h}}_{jcu}$ is the least-square (LS) estimate of the channel, the channel estimate obtained using the LMMSE criterion can be written as

$$\widehat{\mathbf{h}}_{jcu}^{\text{LMMSE}} = \mathbb{E}\left\{\mathbf{h}_{jcu}\mathbf{h}_{jcu}^H\right\} \mathbb{E}\left\{\widehat{\mathbf{h}}_{jju}\widehat{\mathbf{h}}_{jju}^H\right\}^{-1} \widehat{\mathbf{h}}_{jju}$$

$$= \mathbf{R}_{jcu}\mathbf{Q}_{ju}^{-1}\widehat{\mathbf{h}}_{jju} \tag{1}$$

where $\mathbf{Q}_{ju} \triangleq \mathbb{E}\left\{\widehat{\mathbf{h}}_{jju}\widehat{\mathbf{h}}_{jju}^H\right\}$ and $\mathbf{R}_{jcu} \triangleq \mathbb{E}\left\{\mathbf{h}_{jcu}\mathbf{h}_{jcu}^H\right\}$.

Utilizing (1), the corresponding multicell LMMSE combining vector is

$$\mathbf{v}_{ju} = \left(\sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1} \widehat{\mathbf{h}}_{j\ell k}^{\text{LMMSE}} \left(\widehat{\mathbf{h}}_{j\ell k}^{\text{LMMSE}}\right)^H + \mathbf{Z}_j\right)^{-1} \widehat{\mathbf{h}}_{jju}^{\text{LMMSE}} \tag{2}$$

where $\mathbf{Z}_j \triangleq \sum_{\ell=0}^{L-1}\sum_{k=0}^{K-1}(\mathbf{R}_{j\ell k} - \mathbf{R}_{j\ell k}\mathbf{Q}_{jk}^{-1}\mathbf{R}_{j\ell k}) + \sigma^2\mathbf{I}_M$. From (1) and (2), it can be observed that obtaining the LMMSE estimate of the channel and data requires the covariance matrices $\mathbf{R}_{j\ell k}$, $\forall \ell, k$ and $\mathbf{Q}_{jk}$, $\forall k$. In practice, $\mathbf{R}_{j\ell k}$ has to be estimated in the presence of pilot contamination, which complicates the

estimation problem since the estimate $\widehat{\mathbf{R}}_{j\ell k}$ is contaminated by the covariance matrices of users in adjacent cells that employ the same pilot.

Existing methods for covariance matrix estimation employ regular pilots for both channel and covariance matrix estimation wherein the channel and covariance matrix estimates are obtained by dedicating a part of the time–frequency resource for pilot transmission. Under the assumption that the pilot transmission from all the cells are synchronized[1], and that every cell transmits the same pilots, the received observations at BS j during pilot transmission in the nth coherence block can be written as

$$\mathbf{Y}_j^{(n)} = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu} \mathbf{h}_{j\ell k}^{(n)} \boldsymbol{\phi}_k^T + \mathbf{W}_j^{(n)} \qquad (3)$$

where $\mathbf{Y}_j^{(n)} \in \mathbb{C}^{M \times \tau}$ are the received observations, $\boldsymbol{\phi}_k \in \mathbb{C}^{\tau}$ is the pilot sequence transmitted by user k, μ is the uplink transmit power, and $\mathbf{W}_j^{(n)} \in \mathbb{C}^{M \times \tau}$ is the additive noise at the BS with each element i.i.d as $\mathcal{CN}(0, \sigma^2)$. Assuming that the pilots $\boldsymbol{\phi}_k$ are taken from the columns of a scaled unitary matrix $\boldsymbol{\Phi} \in \mathbb{C}^{\tau \times \tau}$ with $\boldsymbol{\Phi}^H \boldsymbol{\Phi} = \tau \mathbf{I}_{\tau}$, the LS estimate of the channel can be obtained as

$$\widehat{\mathbf{h}}_{jcu}^{(n)} = \frac{1}{\tau \sqrt{\mu}} \mathbf{Y}_j^{(n)} \boldsymbol{\phi}_u^* = \mathbf{h}_{jcu}^{(n)} + \sum_{\ell \neq j} \mathbf{h}_{j\ell u}^{(n)} + \frac{\mathbf{W}_j^{(n)} \boldsymbol{\phi}_m^*}{\tau \sqrt{\mu}} . \quad (4)$$

Since $\mathbf{Q}_{ju} = \mathbb{E}\{\widehat{\mathbf{h}}_{jju} \widehat{\mathbf{h}}_{jju}^H\}$, its estimate can be obtained from the sample mean of $\widehat{\mathbf{h}}_{jju}^{(n)}$ over N_Q coherence blocks as

$$\widehat{\mathbf{Q}}_{ju} = \frac{1}{N_Q} \sum_{n=0}^{N_Q-1} \widehat{\mathbf{h}}_{jju}^{(n)} \left(\widehat{\mathbf{h}}_{jju}^{(n)}\right)^H . \qquad (5)$$

However, as mentioned earlier, estimating individual covariance matrices $\mathbf{R}_{j\ell k}$, $\forall \ell, k$ is challenging, since channel observations are made in the presence of pilot contamination. In [14], $\mathbf{R}_{j\ell k}$ is estimated indirectly through $\mathbf{Q}_{j\ell,-k} \triangleq \mathbf{Q}_{jk} - \mathbf{R}_{j\ell k}$, which is the sum covariance matrix of the channels of all the interfering users using the same pilot as user (ℓ, k). $\widehat{\mathbf{Q}}_{j\ell,-k}$ is estimated separately using N_R unique orthogonal pilots for each k and then subtracted from $\widehat{\mathbf{Q}}_{jk}$ to obtain $\widehat{\mathbf{R}}_{j\ell k}$, i.e.,

$$\widehat{\mathbf{R}}_{j\ell k} = \widehat{\mathbf{Q}}_{jk} - \widehat{\mathbf{Q}}_{j\ell,-k} . \qquad (6)$$

When M is larger than N_Q and N_R, the resulting estimates of $\widehat{\mathbf{R}}_{j\ell k}$ and $\widehat{\mathbf{Q}}_{jk}$ have to be regularized in order to ensure full-rank and positive (semi) definiteness [14]. For a massive MIMO system with L cells and K users per cell, estimating both $\mathbf{R}_{j\ell k}$, $\forall \ell, k$ and $\mathbf{Q}_{jk}$ using this approach would require $LKN_R + KN_Q$ UL training symbols. In addition, utilizing unique pilots for estimating $\widehat{\mathbf{Q}}_{j\ell,-k}$ implicitly assumes and requires that the users in all the L cells transmit UL pilots simultaneously. While such an assumption is common in massive MIMO literature, it may not be practically feasible since it requires that the BSs coordinate the UL pilot transmissions of their users.

III. PROPOSED PILOT STRUCTURE AND METHOD FOR ESTIMATING COVARIANCE MATRICES

In the proposed approach, we assume that the L cells are divided into T subsets with the tth subset containing L_t

contiguous cells. Here, L_t is chosen such that $1 \leq L_t \leq \lfloor C_u/2K \rfloor$. The L_t cells within each of the T subsets are assumed to be able to coordinate their UL pilot transmissions, whereas the cells in two different subsets transmit pilot and data asynchronously. Let $\mathcal{M}: \{1, \ldots, L\} \rightarrow \{1, \ldots, T\}$ be the mapping between a cell and its corresponding subset, and let $t = \mathcal{M}_\ell$.

User (ℓ, k) transmits the symbol vector $\mathbf{s}_{\ell k}^{(n)} \triangleq \rho \mathbf{x}_{\ell k}^{(n)} + \lambda \overline{\mathbf{p}}_{\ell k}^{(n)}$ in the UL in the nth coherence block, where $\overline{\mathbf{p}}_{\ell k}^{(n)} \in \mathbb{C}^{C_u}$ is the UL pilot, $\mathbf{x}_{\ell k}^{(n)} \in \mathbb{C}^{C_u}$ is the UL data, and ρ^2 and λ^2 are the fractions of power with which data and pilots are transmitted, respectively. Then, we assume that the pilot sequence $\overline{\mathbf{p}}_{\ell k}^{(n)}$ is comprised of two subsequences and can be written as

$$\overline{\mathbf{p}}_{\ell k}^{(n)} \triangleq \left[\mathbf{p}_{\ell k}^T, e^{\mathrm{j}\theta_{t,n}} \mathbf{p}_{\ell k}^T\right]^T \qquad (7)$$

where $\{\theta_{t,n}\}_{n=1}^N$ are N realizations of a random variable Θ_t. The random variable Θ_t is assumed to be independent of the channel and data vectors and distributed such that $\{\Theta_t\}_{t=1}^T$ are mutually independent and $\mathbb{E}\left[e^{\mathrm{j}\Theta_t}\right] = 0$, $\forall t$. In addition, we also assume that $\theta_{t,n}$, $\forall t, n$ are known to all the L BSs, and that the subsequence $\mathbf{p}_{\ell k}$ is chosen from the columns of a scaled unitary matrix $\mathbf{P}$, where $\mathbf{P}$ is such that $\mathbf{P}^H \mathbf{P} = K\mathbf{I}_{C_u/2}$. Also, since $C_u \geq 2L_t K$, each user in the L_t cells from subset t can be assigned a unique pilot $\overline{\mathbf{p}}_{\ell k}$. It has to be noted that the symbol vector $\mathbf{s}_{\ell k}^{(n)}$ can either contain regular,[2] staggered, or superimposed pilots depending on the contents of $\mathbf{p}_{\ell k}$ and $\mathbf{x}_{\ell k}^{(n)}$. With staggered pilots, the users in different cells stagger their UL pilot transmissions [18], [19], and with superimposed pilots, the users transmit UL pilots alongside data [11].

Let $\mathbf{Y}_j^{(n,p)} \in \mathbb{C}^{M \times C_u/2}$ for $p = 1, 2$ be the received observations at BS j when the first and second pilot subsequences are transmitted in the nth coherence block. Then, $\mathbf{Y}_j^{(n,p)}$, $\forall p = 1, 2$ can be written as

$$\mathbf{Y}_j^{(n,p)} = \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \sqrt{\mu} \mathbf{h}_{j\ell k} \left(\mathbf{s}_{\ell k}^{(n,p)}\right)^T + \mathbf{W}_j^{(n,p)} \qquad (8)$$

where $\mathbf{s}_{\ell k}^{(n,p)}$ and $\mathbf{W}_j^{(n,p)}$ are the transmitted symbols and additive noise at the BS during the transmission of the pth pilot subsequence.

Dropping the index n, for an arbitrary user u in cell c at BS j, consider the cross-correlation between the LS estimates of the channel obtained from the first and second pilot subsequences, that is,

$$\mathbf{R}_{j,\widehat{\mathbf{h}}_{cu}^{(1)} \widehat{\mathbf{h}}_{cu}^{(2)}} \triangleq \mathbb{E}\left[\left\{\mathbf{Y}_j^{(1)} \left(\lambda^2 \mathbf{p}_{cu}^T \mathbf{p}_{cu}^*\right)^{-1} \lambda \mathbf{p}_{cu}^*\right\} \right.$$
$$\left. \times \left\{\mathbf{Y}_j^{(2)} \left(\lambda^2 \mathbf{p}_{cu}^T \mathbf{p}_{cu}^*\right)^{-1} \lambda e^{-\mathrm{j}\Theta_{\mathcal{M}_c}} \mathbf{p}_{cu}^*\right\}^H \right] . \quad (9)$$

Substituting (8) and the definition of $\mathbf{s}_{cu}$ into (9), we obtain

$$\mathbf{R}_{j,\widehat{\mathbf{h}}_{cu}^{(1)} \widehat{\mathbf{h}}_{cu}^{(2)}} = \mathbb{E}\left[\left\{\mathbf{h}_{jcu} + \boldsymbol{\alpha}_{cu}^{(1)} + \boldsymbol{\epsilon}_{cu}^{(1)} + \mathbf{w}_j^{(1)}\right\} \right.$$
$$\left. \times \left\{\mathbf{h}_{jcu} + e^{-\mathrm{j}\Theta_{\mathcal{M}_c}} \boldsymbol{\alpha}_{cu}^{(2)} + e^{-\mathrm{j}\Theta_{\mathcal{M}_c}} \boldsymbol{\epsilon}_{cu}^{(2)} + e^{-\mathrm{j}\Theta_{\mathcal{M}_c}} \mathbf{w}_j^{(2)}\right\}^H\right] = \mathbf{R}_{jcu} \tag{10}$$

---

[1]Pilots that are transmitted simultaneously by the users in all cells are henceforth referred to as regular pilots.

[2]Note that with regular pilots, users in all the L cells transmit pilots simultaneously and the condition $C_u \geq 2L_t K$ does not apply.

where, for $p = 1, 2$, we have

$$\boldsymbol{\alpha}_{cu}^{(p)} \triangleq \frac{\rho}{K\lambda} \sum_{\ell \in \mathcal{T}_c} \sum_{k=0}^{K-1} \mathbf{h}_{j\ell k} \left(\mathbf{x}_{\ell k}^{(p)}\right)^T \mathbf{p}_{cu}^* \quad (11)$$

$$\boldsymbol{\epsilon}_{cu}^{(p)} \triangleq \frac{1}{K\lambda} \sum_{\ell \in \mathcal{T}_{\backslash c}} \sum_{k=0}^{K-1} \mathbf{h}_{j\ell k} \left(\mathbf{z}_{\ell k}^{(p)}\right)^T \mathbf{p}_{cu}^* \quad (12)$$

$$\mathbf{w}_j^{(p)} \triangleq \mathbf{W}_j^{(p)} \mathbf{p}_{cu}^* / \left(K\lambda\sqrt{\mu}\right) . \quad (13)$$

Here, $\mathcal{T}_c = \{\ell \mid \mathcal{M}_\ell = \mathcal{M}_c\}$ is the set of cells that are in the same subset as cell c and $\mathcal{T}_{\backslash c}$ is its complement, $\mathbf{z}_{\ell k}$ is the vector of symbols (either pilots or data) transmitted asynchronously by a user[3] in $\mathcal{T}_{\backslash c}$. In (10), $e^{j\Theta_{\mathcal{M}_c}}$ decorrelates the channel estimation errors resulting from the transmissions from the users in $\mathcal{T}_{\backslash c}$, which in turn causes the cross-correlation of the channel estimates to become equal to $\mathbf{R}_{jcu}$.

Using the result in (10), an estimate of $\mathbf{R}_{jcu}$ can be obtained by the sample cross-correlation of both the channel estimates averaged over N coherence blocks, i.e.,

$$\widehat{\mathbf{R}}_{jcu} = \frac{1}{N} \sum_{n=1}^{N} \widehat{\mathbf{h}}_{jcu}^{(n,1)} \left(\widehat{\mathbf{h}}_{jcu}^{(n,2)}\right)^H \quad (14)$$

where

$$\widehat{\mathbf{h}}_{jcu}^{(n,1)} = \mathbf{Y}_j^{(n,1)} \left(\lambda^2 \mathbf{p}_{cu}^T \mathbf{p}_{cu}^*\right)^{-1} \lambda \mathbf{p}_{cu}^* \quad (15)$$

$$\widehat{\mathbf{h}}_{jcu}^{(n,2)} = \mathbf{Y}_j^{(n,2)} \left(\lambda^2 \mathbf{p}_{cu}^T \mathbf{p}_{cu}^*\right)^{-1} \lambda e^{-j\theta_{\mathcal{M}_c,n}} \mathbf{p}_{cu}^* . \quad (16)$$

It is straightforward to show that the sample cross-correlation converges in probability to the true correlation, i.e., $\widehat{\mathbf{R}}_{jcu} \xrightarrow[N \to \infty]{P} \mathbf{R}_{jcu}$.

However, for a finite N, the estimate $\widehat{\mathbf{R}}_{jcu}$ is not necessarily Hermitian symmetric. Therefore, this matrix can be regularized by approximating it with a positive semidefinite matrix. Thus, we approximate $\widehat{\mathbf{R}}_{jcu}$ with the positive semidefinite matrix closest in Frobenius norm, which can be easily shown to be $\widehat{\mathbf{R}}_{jcu}^{\text{PSD}} \triangleq \mathbf{U}\mathbf{D}_+\mathbf{U}^H$, where $\mathbf{D}_+$ is a diagonal matrix that contains only the positive eigenvalues of the symmetric part of $\widehat{\mathbf{R}}_{jcu}$, i.e., $\widehat{\mathbf{R}}_{jcu}^{\text{sym}} \triangleq (\widehat{\mathbf{R}}_{jcu} + \widehat{\mathbf{R}}_{jcu}^H)/2$, and $\mathbf{U}$ contains the corresponding eigenvectors.

In order to estimate the channel covariance matrix of a user (c, u) at an arbitrary BS j, the BS requires only the knowledge of $\{\theta_{\mathcal{M}_c,n}\}_{n=1}^{N}$ and the symbol and subcarrier indices in which user (c, u) transmits its UL pilots. As a result, unlike in [14] and [15], the proposed method does not require that all the users transmit pilots simultaneously.[4] In fact, as will be shown in Section V, the proposed method performs well even when $L_t = 1$, $\forall t$, i.e., when none of the BSs coordinate the UL pilot/data transmissions of their users.

IV. ESTIMATING COVARIANCE MATRICES USING STAGGERED PILOTS

The proposed method in Section III can be employed with either regular, superimposed, or staggered pilots. However, we will restrict our attention in this section to staggered pilots since it is a particular case of superimposed pilots and provides an

---

additional degree of freedom by allowing the pilot and data powers to be varied [18]–[20]. The estimated channel will be used in a regularized zero-forcing combiner for data detection, which is given as

$$\mathbf{v}_{cu} = \left(\frac{\sigma^2}{\mu}\mathbf{I} + \sum_{k=0}^{K-1} \widehat{\mathbf{h}}_{cck}^{\text{LMMSE}} \left(\widehat{\mathbf{h}}_{cck}^{\text{LMMSE}}\right)^H\right)^{-1} \widehat{\mathbf{h}}_{ccu}^{\text{LMMSE}} . \quad (17)$$

For covariance matrix estimation using staggered or regular pilots, each user transmits two K length pilot sequences within a coherence block, with the second pilot sequence multiplied by a random phase shift. For each pilot sequence, $\mathbf{P} = \text{blkdiag}\{\boldsymbol{\Phi}, \ldots, \boldsymbol{\Phi}\}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{K \times K}$. A lower bound on the capacity for user $(0, u)$ can be obtained as

$$R \triangleq \eta_1 \Psi \left(\frac{s_u}{\bar{\imath}_u}\right) + \eta_2 \Psi \left(\frac{s_u}{\tilde{\imath}_u}\right) \quad (18)$$

where $\Psi(x) \triangleq \log_2(1 + x)$, and

$$\eta_1 \triangleq \frac{(L-1)K}{C_u \tau_s}(\tau_s + N); \quad \eta_2 \triangleq \left(1 - \frac{LK}{C_u} - \frac{NLK}{C_u \tau_s}\right) \quad (19)$$

$$s_u \triangleq \rho^2 |\mathbb{E}\{\mathbf{v}_{0u}^H \mathbf{h}_{00u}\}|^2 \quad (20)$$

$$\bar{\imath}_u \triangleq \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \theta_\ell \mathbb{E}\{|\mathbf{v}_{0u}^H \mathbf{h}_{0\ell k}|^2\} - \rho^2 |\mathbb{E}\{\mathbf{v}_{0u}^H \mathbf{h}_{00u}\}|^2$$
$$+ \frac{\sigma^2}{\mu}\mathbb{E}\{\|\mathbf{v}_{0u}\|^2\} + \frac{(\lambda^2 - \rho^2)}{(|\mathcal{T}_0| - 1)} \sum_{\mathcal{T}_0 \ni c \neq 0} \sum_{k=0}^{K-1} \mathbb{E}\{|\mathbf{v}_{0u}^H \mathbf{h}_{0ck}|^2\} \quad (21)$$

$$\tilde{\imath}_u \triangleq \sum_{\ell=0}^{L-1} \sum_{k=0}^{K-1} \theta_\ell \mathbb{E}\{|\mathbf{v}_{0u}^H \mathbf{h}_{0\ell k}|^2\} + \frac{\sigma^2}{\mu}\mathbb{E}\{\|\mathbf{v}_{0u}\|^2\}$$
$$- \rho^2 |\mathbb{E}\{\mathbf{v}_{0u}^H \mathbf{h}_{00u}\}|^2 . \quad (22)$$

Here, $\theta_\ell \triangleq (\rho^2 \mathbf{1}_{\{\ell \in \mathcal{T}_0\}} + \max\{\rho^2, \lambda^2\} \mathbf{1}_{\{\ell \notin \mathcal{T}_0\}})$. The derivation of (18) is detailed in [21]. Defining $\mathbf{E}_{jcu} \triangleq \mathbf{R}_{jcu} - \widehat{\mathbf{R}}_{jcu}$, the MSE expressions of the covariance matrix estimates can be obtained in a straightforward manner as

$$\mathbb{E}\{\|\mathbf{E}_{jcu}\|_F^2\} = \frac{1}{N} \sum_{r=0}^{M-1} \sum_{s=0}^{M-1} (g_r g_s + f_{rs}) \quad (23)$$

where

$$f_{rs} \triangleq \sum_{\mathcal{T}_c \ni \ell \neq c} \sum_{k=0}^{K-1} \frac{\rho^4}{K^2 \lambda^4} |[\mathbf{R}_{\ell k}]_{rs}|^2 + \left|\frac{1}{K\lambda^2} \sum_{\ell \in \mathcal{T}_{\backslash c}} \sum_{k=0}^{K-1} [\mathbf{R}_{\ell k}]_{rs}\right|^2 \quad (24)$$

$$g_r \triangleq [\mathbf{R}_{cu}]_{rr} + \sum_{\mathcal{T}_c \ni \ell \neq c} \sum_{k=0}^{K-1} \frac{\rho^2}{K\lambda^2} |[\mathbf{R}_{\ell k}]_{rr}|^2$$
$$+ \sum_{\ell \in \mathcal{T}_{\backslash c}} \sum_{k=0}^{K-1} \frac{1}{K\lambda^2} |[\mathbf{R}_{\ell k}]_{rr}|^2 + \frac{\sigma^2}{K\lambda^2 \mu} . \quad (25)$$

V. SIMULATION RESULTS

We compare the normalized MSE of the channel estimate and the achievable rate of the proposed method with that in
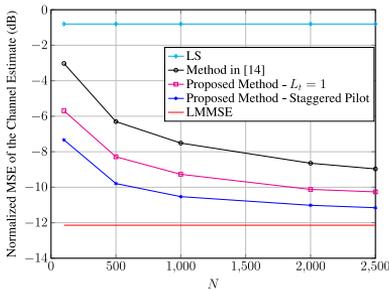
Fig. 1. Normalized MSE of channel estimate of an average user vs. N.
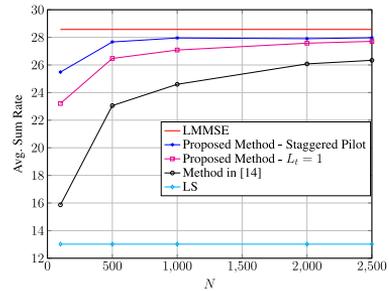
Fig. 2. Achievable rate in the UL vs. N.
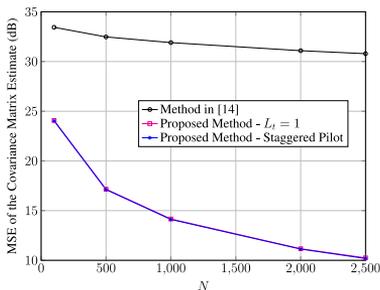
Fig. 3. MSE of the estimated covariance matrix of an average user vs. N.
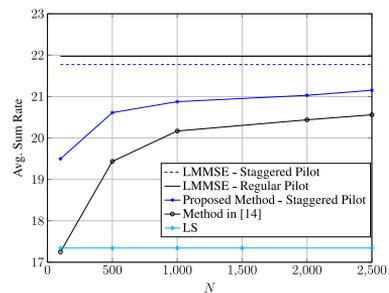
Fig. 4. Achievable rate in the UL vs. N.

[14].[5] Both the methods are simulated for one tier of $L = 7$ hexagonal cells with the BSs separated by 300 m. The performance of the methods are evaluated for the users in the central cell. The SNR at the receiver is $78.6 - 37.6 \log_{10} d$, where d is the distance from the BS in meters. The channel statistics are assumed to be constant over $\tau_s = 25\,000$ coherence blocks. The received paths from a user are assumed to be uniformly distributed over an angular spread of $20°$, with mean angle of arrival given by the geographical locations of the users. In all the plots, the performance metrics are plotted against N. Then, $N = N_Q$ ensures that the same number of coherence blocks are utilized for estimating the covariance matrix for both the proposed method and [14]. For the method in [14], N_R is chosen as $N_R = N_Q/10$. The normalized MSE is defined as $\mathbb{E}\left\{\|\widehat{\mathbf{h}}_{jcu}^{\mathrm{LMMSE}} - \mathbf{h}_{jcu}\|^2 / \mathrm{trace}\left\{\mathbf{R}_{jcu}\right\}\right\}$. For the proposed method, $\rho^2 = \lambda^2 = 1$.

In Figs. 1–3, the performance of the proposed method is compared with that in [14]. In order to compare only the covariance matrix estimates, the proposed method utilizes staggered pilots for N coherence blocks and regular pilots for the remaining $\tau_s - N$ blocks. Consequently, $\widehat{\mathbf{Q}}_{ju}$ is estimated for the proposed method as $\widehat{\mathbf{Q}}_{ju} = \sum_{\ell=0}^{L-1} \widehat{\mathbf{R}}_{j\ell u} + (\sigma^2/K\mu)\mathbf{I}$. Note that we have used the same simulation setup as in [14] in which $C_u = 100$ symbols, each BS has $M = 100$ antennas and contains $K = 10$ users in its cell, which are equispaced on a circle of radius 120 m from the BS. In addition, for staggered pilots L_t is chosen as 7, and for the sake of simplicity, $L_t = 1$ is

simulated using regular pilots, although the proposed method would still work if the pilot and data transmissions of different cells would overlap. From Fig. 3, it can be seen that the MSE of the covariance matrix is significantly lower for the proposed method. Consequently, in Figs. 1 and 2, the MSE and sum-rate performance of the proposed method is significantly better than the method in [14], despite not requiring all users to transmit the pilots over the same set of symbols.

In Fig. 4, $K = 5$ users are uniformly distributed across the entire cell. Each BS has $M = 50$ antennas and the UL time slot has $C_u = 100$ symbols. With the LMMSE method, the sum-rate with regular pilots is marginally higher than that for staggered pilots. However, the proposed method offers a higher throughput in comparison with the method in [14]. Note that the pre-log factor contributes to the small difference between the achievable rates of the LMMSE and estimated covariance matrices.

VI. CONCLUSION

We proposed a novel pilot structure for estimating the covariance matrix in the presence of pilot contamination, which has the advantage of not requiring simultaneous UL pilot transmissions. Using the proposed method along with staggered pilots, we showed that the proposed method offers a higher UL throughput and lower MSE than existing schemes.

The performance of the proposed method could be further improved by optimizing ρ and λ, which we leave as a problem for future research.

REFERENCES

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[4] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] H. Yang and T. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[6] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[7] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[8] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

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

[10] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, Mar. 2016, pp. 3366–3370.

[11] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots are superior for mitigating pilot contamination in massive MIMO," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2917–2932, Jun. 2017.

[12] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[13] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.,* vol. 17, no. 1, pp. 574–590, Jan. 2018.

[14] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. 50th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2016, pp. 974–978.

[15] D. Neumann, K. Shibli, M. Joham, and W. Utschick, "Joint covariance matrix estimation and pilot allocation in massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[16] S. Haghighatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8316–8332, Dec. 2017.

[17] E. Björnson, J. Hoydis, and L. Sanguinetti, "Pilot contamination is not a fundamental asymptotic limitation in massive MIMO," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[18] W. A. W. M. Mahyiddin, P. A. Martin, and P. J. Smith, "Performance of synchronized and unsynchronized pilots in finite massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6763–6776, Dec. 2015.

[19] D. Kong, D. Qu, K. Luo, and T. Jiang, "Channel estimation under staggered frame structure for massive MIMO system," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1469–1479, Feb. 2016.

[20] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Downlink performance of superimposed pilots in massive MIMO systems," *IEEE Trans. Wireless Commun.*, to be published. [Online]. Available: https://arxiv.org/pdf/1606.04476.pdf

[21] K. Upadhya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," arXiv:1802.01513, 2018. [Online]. Available: https://arxiv.org/abs/1802.01513

Publication VII

K. Upadhya, R. W. Heath Jr, S. A. Vorobyov. Tracking abruptly changing channels in mmWave systems using overlaid data and training. *IEEE 7th Int. Workshop Computational Advances Multi-Sensor Adaptive Process.*, pp. 1-5, Dec. 2017.

Tracking Abruptly Changing Channels in mmWave Systems Using Overlaid Data and Training

Karthik Upadhya[*], Robert W. Heath Jr.[†], and Sergiy A. Vorobyov[*]

[*]Department of Signal Processing and Acoustics, Aalto University, Espoo, FI-00076 AALTO, Finland

Emails: karthik.upadhya@aalto.fi and svor@ieee.org

[†] Wireless Networking and Communications Group, The University of Texas at Austin, Austin, TX 78712

Email: rheath@utexas.edu

Abstract—**Millimeter-wave (mmWave) multiple-input multiple-output (MIMO) links are sensitive to abrupt changes in the channel due to blockage and node mobility. We propose to estimate the channel by overlaying pilot and data transmissions. The data transmission is performed over the signal subspace of the channel matrix, while the training, for estimating the parameters of newly appearing paths, is performed over the null-space of the channel matrix. A sparse Bayesian learning-based approach is employed for jointly estimating the channel and data at the receiver. Simulations are used to validate the performance of the proposed method in abruptly changing channel scenarios.**

Index Terms—**mmWave, massive MIMO, channel tracking, blockage**

I. Introduction

Millimeter-wave (mmWave) multiple-input multiple-output (MIMO) systems are affected by blockages and abrupt variations in the channel [1], [2]. Resuming communication after blockage requires re-estimating the channel, leading to an increased overhead and added latency.

There is some prior work on studied channel tracking for mmWave MIMO systems. Methods in [3]–[6] rely on models for the temporal variation of the channel to perform beam selection and switching. However, these methods have been developed for analog beamforming and cannot be extended easily to systems that employ hybrid beamforming. In addition, these methods cannot be applied in scenarios in which accurate knowledge of the temporal variation of the channel is not available. Approaches for channel tracking using compressive sensing with hybrid beamforming have been described in [7]–[9], but these methods require the channels to vary slowly between subsequent training blocks.

In this paper, we propose a method to track abrupt changes in the channel by overlaying pilot and data transmissions. Overlaid pilot and data transmissions have been previously used in conventional sub-6 GHz MIMO for channel estimation in rapidly time-varying channel scenarios [10]. However, unlike conventional sub-6 GHz MIMO, wherein overlaying pilots and data simply involves transmitting pilots alongside data at a reduced power, in mmWave systems, we propose transmitting the data in the signal-space of the channel and the pilots, for identifying new paths, in the null-space of the channel. A data-aided channel estimation method based on the sparse Bayesian learning (SBL) framework is developed for jointly estimating the channel and the data. The proposed method is validated using simulations.

Notation: The transpose, conjugate, and conjugate transpose are denoted by $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$. The Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is $\mathbf{A} \otimes \mathbf{B}$, the Kronecker delta function is $\delta(\cdot)$, $\mathbf{I}$ is the identity matrix, $\mathrm{vec}(\mathbf{A})$ is the vectorization operator, $\mathrm{span}\{\mathbf{A}\}$ represents the column space of $\mathbf{A}$, and $\mathbf{1}_{\{\mathcal{S}\}}$ denotes the indicator function over the set $\mathcal{S}$.

II. System Model

We consider a narrowband mmWave MIMO system with N_{BS} antennas at the base station (BS) and N_{MS} antennas at the mobile station (MS). The BS (MS) is assumed to have $N_{\mathrm{BS}}^{\mathrm{RF}}$ ($N_{\mathrm{MS}}^{\mathrm{RF}}$) RF chains such that $N_{\mathrm{BS}}^{\mathrm{RF}} \ll N_{\mathrm{BS}}$ ($N_{\mathrm{MS}}^{\mathrm{RF}} \ll N_{\mathrm{MS}}$). Without loss of generality, we consider the downlink, where $1 \le D_s \le \min\left\{N_{\mathrm{BS}}^{\mathrm{RF}}, N_{\mathrm{MS}}^{\mathrm{RF}}\right\}$ data streams are transmitted from the BS to the MS. The data transmission interval is assumed to be divided into M blocks with K data symbols per block.[1] The channel is assumed to remain constant within each of these data blocks, but may change across different blocks. If at the kth instant of the mth block, $\mathbf{W}_{m,k} \in \mathbb{C}^{N_{\mathrm{MS}} \times N_{\mathrm{MS}}^{\mathrm{RF}}}$ is chosen as the combining matrix at the MS, $\mathbf{f}_{m,k} \in \mathbb{C}^{N_{\mathrm{BS}}}$ is the precoded vector of symbols transmitted by the BS, and $\mathbf{H}_m \in \mathbb{C}^{N_{\mathrm{MS}} \times N_{\mathrm{BS}}}$ is the channel matrix, then the received vector of symbols can be written as

$$\mathbf{y}_{m,k} = \mathbf{W}_{m,k}^H \mathbf{H}_m \mathbf{f}_{m,k} + \mathbf{W}_{m,k}^H \mathbf{q}_{m,k} \qquad (1)$$

where $\mathbf{q}_{m,k} \in \mathbb{C}^{N_{\mathrm{MS}}}$ is the additive noise distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$.

We assume that the BS and MS use a one-dimensional uniform linear array (ULA). The channel matrix $\mathbf{H}_m$ is assumed to be composed of L distinct paths with the parameters $\alpha_{\ell,m} \in \mathbb{C}$, $\phi_{\ell,m} \in \mathbb{R}$ and $\theta_{\ell,m} \in \mathbb{R}$ representing the complex gain, angle of arrival (AoA), and angle of departure (AoD) of the ℓth path. If $\mathbf{a}_{\mathrm{MS}}(\cdot)$ and $\mathbf{a}_{\mathrm{BS}}(\cdot)$ are the steering vectors corresponding to the MS and BS, the channel matrix $\mathbf{H}_m$ can be written as

$$\mathbf{H}_m = \sum_{\ell=0}^{L-1} \alpha_{\ell,m} \mathbf{a}_{\mathrm{MS}}(\phi_{\ell,m}) \mathbf{a}_{\mathrm{BS}}(\theta_{\ell,m})^H . \qquad (2)$$

---

[1]Each data symbol can be assumed to be a sub-carrier of an orthogonal frequency division multiplexing (OFDM) symbol dedicated for tracking the channel. Therefore a data block consists of K OFDM symbols.

A compact representation of the channel in (2) can be obtained by defining the matrices $\overline{\mathbf{A}}_{\mathrm{BS}_m} \in \mathbb{C}^{N_{\mathrm{BS}} \times L}$, $\overline{\mathbf{A}}_{\mathrm{MS}_m} \in \mathbb{C}^{N_{\mathrm{MS}} \times L}$, and $\overline{\mathbf{D}}_m \in \mathbb{C}^{L \times L}$. The columns of $\overline{\mathbf{A}}_{\mathrm{BS}_m}$ are the steering vectors $\{\mathbf{a}_{\mathrm{BS}}(\theta_{\ell,m})\}_{\ell=0}^{L-1}$ and the columns of $\overline{\mathbf{A}}_{\mathrm{MS}_m}$ are $\{\mathbf{a}_{\mathrm{MS}}(\phi_{\ell,m})\}_{\ell=0}^{L-1}$. The diagonal matrix $\overline{\mathbf{D}}_m$ has diagonal elements $\{\alpha_{\ell,m}\}_{\ell=0}^{L-1}$. Then, (2) can be rewritten as

$$\mathbf{H}_m = \overline{\mathbf{A}}_{\mathrm{MS}_m} \overline{\mathbf{D}}_m \overline{\mathbf{A}}_{\mathrm{BS}_m}^H \tag{3}$$

Since mmWave channels are sparse [1], the channel can be estimated by formulating a corresponding sparse recovery problem under the assumption that ϕ and θ are drawn from a quantized grid, with G_{MS} and G_{BS} points, as in [11]. Then, given the matrices $\mathbf{A}_{\mathrm{BS}} \in \mathbb{C}^{N_{\mathrm{BS}} \times G_{\mathrm{BS}}}$ and $\mathbf{A}_{\mathrm{MS}} \in \mathbb{C}^{N_{\mathrm{MS}} \times G_{\mathrm{MS}}}$, whose columns are $\mathbf{a}_{\mathrm{BS}}(\theta)$ and $\mathbf{a}_{\mathrm{MS}}(\phi)$, and the sparse matrix $\mathbf{D}_m \in \mathbb{C}^{G_{\mathrm{MS}} \times G_{\mathrm{BS}}}$, the channel matrix can be written as

$$\mathbf{H}_m = \mathbf{A}_{\mathrm{BS}} \mathbf{D}_m \mathbf{A}_{\mathrm{MS}}^H \tag{4}$$

where $\| \mathrm{vec}(\mathbf{D}_m) \|_0 \ll G_{\mathrm{BS}} G_{\mathrm{MS}}$.

III. Existing Approach for Channel Estimation

In the channel estimation methods proposed in [11], [12], the BS transmits M_{BS} training symbols for channel estimation while the MS makes M_{MS} measurements for each transmitted training symbol. The BS designs its training precoding matrix $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{BS}} \times M_{\mathrm{BS}}}$ such that $[\mathbf{F}]_{i,j} = e^{\mathrm{j}\xi}$, where ξ is chosen randomly from the set of N_{Q} quantized angles $\{0, 2\pi/N_{\mathrm{Q}}, \dots, 2\pi(N_{\mathrm{Q}}-1)/N_{\mathrm{Q}}\}$. In a similar manner, the MS generates its combining matrix $\mathbf{W} \in \mathbb{C}^{N_{\mathrm{MS}} \times M_{\mathrm{MS}}}$. Then, the received observations at the MS can be written as [12]

$$\mathbf{Y} = \mathbf{W}^H \mathbf{H} \mathbf{F} + \mathbf{Q} \tag{5}$$

where $\mathbf{Q}$ is the matrix corresponding to the additive white noise at the MS and the subscript m is been dropped for the initial channel estimation block. Vectorizing the observation matrix and using (4), (5) becomes

$$\begin{aligned} \mathbf{y} = \mathrm{vec}(\mathbf{Y}) &= \left(\mathbf{F}^T \otimes \mathbf{W}^H\right) \mathrm{vec}(\mathbf{H}) + \mathbf{q} \\ &= \left(\mathbf{F}^T \mathbf{A}_{\mathrm{BS}}^* \otimes \mathbf{W}^H \mathbf{A}_{\mathrm{MS}}\right) \mathbf{d} + \mathbf{q} \end{aligned} \tag{6}$$

where $\mathbf{d} \triangleq \mathrm{vec}(\mathbf{D})$, and $\mathbf{q} \triangleq \mathrm{vec}(\mathbf{Q})$. Since the vector $\mathbf{d}$ is sparse, its estimate $\widehat{\mathbf{d}}$ can be obtained from the observations $\mathbf{y}$ using a sparse-recovery algorithm. Then, the non-zero values of $\widehat{\mathbf{d}}$ and their locations are fed back to the BS for designing the precoding and combining matrices, and the subsequent symbols are used for data transmission.

IV. Channel Estimation Using Overlaid Data and Training

The approach described in Section III is susceptible to abrupt changes in the channel due to blockage and node-mobility. As an alternative, we propose simultaneous data and pilot transmission so that the BS and MS have the latest channel state information and can use alternative path-clusters to maintain the communication link in the event that the LOS path or the dominant NLOS cluster is blocked.

Channel Estimation	Data + Pilot Block 0	Data + Pilot Block 1	...	Data + Pilot Block M-1	Data + Pilot Block M
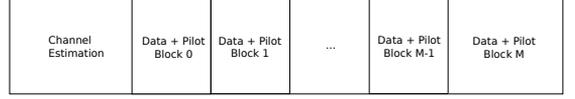
Fig. 1. Frame structure of proposed channel estimation approach.

A. Channel Mobility Model

To model abrupt changes, we propose a generic channel model which may be useful for mmWave MIMO systems. We associate each of the L paths with a state $S_\ell \in \{0, 1\}$, which indicates whether the path is blocked or visible to the MS. Defining $\mathcal{S} \triangleq \{S_\ell, \forall \ell = 1, \dots, L\}$ as the set containing the states of the L paths, the channel matrix can be written as

$$\mathbf{H}(\mathcal{S}) = \sum_{\ell=0}^{L-1} \alpha_\ell S_\ell \mathbf{a}_{\mathrm{MS}}(\phi_\ell) \mathbf{a}_{\mathrm{BS}}(\theta_\ell)^H . \tag{7}$$

In addition, we define a change point τ at which the states of some of the L paths change. If $\mathcal{S}_0$ and $\mathcal{S}_1$ are the state vectors before and after the τth block, the channel in the mth block can be written as

$$\mathbf{H}_m = \mathbf{H}(\mathcal{S}_0) \mathbf{1}_{\{m < \tau\}} + \mathbf{H}(\mathcal{S}_1) \mathbf{1}_{\{m \geq \tau\}} . \tag{8}$$

We consider two abrupt changes in the channel in which the LOS path becomes blocked or visible, which are represented by the state of the LOS path being toggled from active to inactive (or inactive to active).

B. The Approach

In the kth symbol interval of the mth data block, the BS transmits a linear combination of the pilot and data vectors $\rho_{\mathrm{d}} \mathbf{F}_m^{\mathrm{d}} \mathbf{s}_{m,k} + \rho_{\mathrm{t}} \mathbf{f}_{m,k}^{\mathrm{t}}$ as shown in Fig. 1. The data vector $\mathbf{s}_{m,k}$, with symbols drawn from the constellation χ, is precoded using the matrix $\mathbf{F}_m^{\mathrm{d}}$ and is linearly combined with the training vector $\mathbf{f}_{m,k}^{\mathrm{t}}$. The fraction of the normalized power reserved for data ρ_{d}^2 and training ρ_{t}^2 is chosen such that $\rho_{\mathrm{d}}^2 + \rho_{\mathrm{t}}^2 = 1$. In addition, if $N_{\mathrm{MS}}^{\mathrm{RF\text{-}T}}$ RF chains at the MS are reserved for training, the receive combining matrix in the kth symbol of the mth block can be written as $\mathbf{W}_{m,k} = \left[\mathbf{W}_m^{\mathrm{d}}, \mathbf{W}_{m,k}^{\mathrm{t}}\right]$, where $\mathbf{W}_{m,k}^{\mathrm{t}} \in \mathbb{C}^{N_{\mathrm{MS}}, N_{\mathrm{MS}}^{\mathrm{RF\text{-}T}}}$ and $\mathbf{W}_m^{\mathrm{d}} \in \mathbb{C}^{N_{\mathrm{MS}}, N_{\mathrm{MS}}^{\mathrm{RF}} - N_{\mathrm{MS}}^{\mathrm{RF\text{-}T}}}$ are the RF combining matrices corresponding to training and data. Since the analog training precoders $\mathbf{f}_{m,k}^{\mathrm{t}}$ and combiners $\mathbf{W}_{m,k}^{\mathrm{t}}$ cannot be changed for every transmitted symbol [13], they are changed at the same rate as the one during the initial channel training.

If $\widehat{\mathbf{H}}_m$ is the current estimate of the channel, the received vector of symbols in (1) can be written as

$$\begin{aligned} \mathbf{y}_{m,k} = \mathbf{W}_{m,k}^H &\left(\widehat{\mathbf{H}}_m + \Delta \mathbf{H}_m\right)\left(\rho_{\mathrm{d}} \mathbf{F}_m^{\mathrm{d}} \mathbf{s}_{m,k} + \rho_{\mathrm{t}} \mathbf{f}_{m,k}^{\mathrm{t}}\right) \\ &+ \mathbf{W}_{m,k}^H \mathbf{q}_{m,k} \end{aligned} \tag{9}$$

where $\Delta \mathbf{H}_m \triangleq \mathbf{H}_m - \widehat{\mathbf{H}}_m$. The singular-value decomposition (SVD) of $\widehat{\mathbf{H}}_m$ is $\widehat{\mathbf{H}}_m = \widehat{\mathbf{U}} \widehat{\boldsymbol{\Sigma}} \widehat{\mathbf{V}}^H$, where $\widehat{\mathbf{U}} \triangleq [\mathbf{U}_{\mathrm{s}}, \mathbf{U}_{\mathrm{n}}]$, $\widehat{\mathbf{V}} \triangleq [\mathbf{V}_{\mathrm{s}}, \mathbf{V}_{\mathrm{n}}]$, and $\widehat{\boldsymbol{\Sigma}}$ is the diagonal matrix containing the singular values of $\widehat{\mathbf{H}}$. $\mathbf{U}_{\mathrm{s}}$ and $\mathbf{V}_{\mathrm{s}}$ are the matrices of the left and right singular vectors corresponding to the signal

space, while $\mathbf{U}_n$ and $\mathbf{V}_n$ correspond to the null space. Given $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$, the optimal precoder and combiner pair for data transmission is $\{\mathbf{F}_m^d = \mathbf{V}_s, \mathbf{W}_m^d = \mathbf{U}_s\}$.

MmWave MIMO transceivers have large arrays and form narrow beams during data transmission. From (3), the left and right signal subspaces of the channel are spanned by the vectors $\{\mathbf{a}_{BS}(\theta_{\ell,m})\}_{\ell=0}^{L-1}$ and $\{\mathbf{a}_{MS}(\phi_{\ell,m})\}_{\ell=0}^{L-1}$. For arrays that have a large number of elements N, we have the property [14]

$$\frac{1}{N}\mathbf{a}(\theta_1)^H\mathbf{a}(\theta_2) \approx \delta(\theta_1, \theta_2) . \qquad (10)$$

As a result, a new path that has a different AoA and AoD pair, when compared with paths that form the current channel estimate, will have steering vectors that are almost orthogonal to the vectors $\{\mathbf{a}_{BS}(\theta_{\ell,m})\}_{\ell=0}^{L-1}$ and $\{\mathbf{a}_{MS}(\phi_{\ell,m})\}_{\ell=0}^{L-1}$. Therefore, this new path will lie in the null-space of the channel. To estimate the parameters of these new paths, the training precoder and combiner pairs are chosen such that $\mathbf{f}_{m,k}^t \in \text{span}\{\mathbf{V}_n\} \ \forall \ k$ and span $\{\mathbf{W}_{m,k}^t\} \subset \text{span}\{\mathbf{U}_n\} \ \forall \ k$.

To keep the notation simple, we drop the subscript m and assume implicitly that the channel and data correspond to the mth block. Then, from (9), the received observations at the output of $\mathbf{W}^d$ and $\mathbf{W}_k^t$ at the MS are given as

$$\mathbf{y}_d \triangleq (\mathbf{W}^d)^H(\widehat{\mathbf{H}} + \Delta\mathbf{H})(\rho_d\mathbf{F}^d\mathbf{s}_k + \rho_t\mathbf{f}_k^t) + (\mathbf{W}^d)^H\mathbf{q}_k$$
$$= \rho_d\widehat{\boldsymbol{\Sigma}}\mathbf{s}_k + (\mathbf{W}^d)^H\Delta\mathbf{H}(\rho_d\mathbf{F}^d\mathbf{s}_k + \rho_t\mathbf{f}_k^t) + (\mathbf{W}^d)^H\mathbf{q}_k \quad (11)$$

$$\mathbf{y}_t \triangleq (\mathbf{W}_k^t)^H(\widehat{\mathbf{H}} + \Delta\mathbf{H})(\rho_d\mathbf{F}^d\mathbf{s}_k + \rho_t\mathbf{f}_k^t) + (\mathbf{W}_k^t)^H\mathbf{q}_k$$
$$= (\mathbf{W}_k^t)^H\Delta\mathbf{H}(\rho_d\mathbf{F}^d\mathbf{s}_k + \rho_t\mathbf{f}_k^t) + (\mathbf{W}_k^t)^H\mathbf{q}_k . \quad (12)$$

If the new path that constitutes $\Delta\mathbf{H}$ has a different AoA and AoD pair than the paths that constitute $\widehat{\mathbf{H}}$, (11) and (12) can be written as

$$\mathbf{y}_d = \rho_d\widehat{\boldsymbol{\Sigma}}\mathbf{s}_k + (\mathbf{W}^d)^H\mathbf{q}_k \qquad (13)$$
$$\mathbf{y}_t = \rho_t(\mathbf{W}_k^t)^H\Delta\mathbf{H}\mathbf{f}_k^t + (\mathbf{W}_k^t)^H\mathbf{q}_k . \qquad (14)$$

The channel and data estimation problems are decoupled if the data is transmitted in the column-space, while the training sequence is transmitted in the null-space of the channel. Since mmWave transceivers use quantized analog-phase shifters with only a few RF chains, exact orthogonality between the training and data precoders/combiners cannot be achieved. This results in crosstalk between the channel and data estimates, necessitating joint channel and data estimation.

C. Data-Aided Sparse Channel Estimation

For data-aided estimation, we use the sparse Bayesian learning (SBL) based approach [15], wherein sparse recovery is performed by assuming a sparsifying prior on the channel beamspace. From (1) and (9), we can write

$$\mathbf{y}_k = \mathbf{W}_k^H\mathbf{H}\mathbf{f}_k + \mathbf{W}_k^H\mathbf{q}_k$$
$$= (\mathbf{f}_k^T \otimes \mathbf{W}_k^H)(\mathbf{A}_{BS}^* \otimes \mathbf{A}_{MS})\text{vec}(\mathbf{D}) + \mathbf{W}_k^H\mathbf{q}_k$$
$$= \boldsymbol{\psi}_k(\mathbf{s}_k)\mathbf{d} + \mathbf{e}_k \qquad (15)$$

where $\boldsymbol{\psi}_k(\mathbf{s}_k) \triangleq (\mathbf{f}_k^T\mathbf{A}_{BS}^* \otimes \mathbf{W}_k^H\mathbf{A}_{MS})$ and $\mathbf{e}_k \triangleq \mathbf{W}_k^H\mathbf{q}_k$. Stacking the K received observations of a block,

$$\mathbf{y} \triangleq [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T = \boldsymbol{\psi}(\mathbf{S})\mathbf{d} + \mathbf{e} \qquad (16)$$

where $\boldsymbol{\psi}(\mathbf{S}) \triangleq [\boldsymbol{\psi}_1^T(\mathbf{s}_1), \ldots, \boldsymbol{\psi}_K^T(\mathbf{s}_K)]^T$ and $\mathbf{e} \triangleq [\mathbf{e}_1^T, \ldots, \mathbf{e}_K^T]^T$. We assume a Bayesian prior on $\mathbf{d}$, i.e., $\mathbf{d} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma} \triangleq \text{diag}(\boldsymbol{\gamma})$ and $\boldsymbol{\gamma} \triangleq (\gamma_1, \ldots, \gamma_{G_{BS}G_{MS}})$. The value of γ_p determines the amplitude of the pth element of $\mathbf{d}$, with the amplitude approaching 0 when $\gamma_p \to 0$. Then, the ML estimate of $\boldsymbol{\gamma}$ and $\mathbf{S}$ is

$$\{\boldsymbol{\gamma}_{ML}, \mathbf{S}_{ML}\} = \arg\max_{\boldsymbol{\gamma}, \mathbf{S}}\log(p(\mathbf{y}; \boldsymbol{\gamma}, \mathbf{S})) . \qquad (17)$$

The channel and data are estimated using the expectation-maximization (EM) algorithm. The E-step at the rth iteration is given as

$$Q(\boldsymbol{\gamma}, \mathbf{S}|\boldsymbol{\gamma}^{(r)}, \mathbf{S}^{(r)}) = \mathbb{E}_{\mathbf{d}|\boldsymbol{\gamma}^{(r)}, \mathbf{S}^{(r)}}\{\log(p(\mathbf{y}; \boldsymbol{\gamma}, \mathbf{S}))\}$$
$$= -\text{trace}(\boldsymbol{\psi}^H(\mathbf{S})\mathbf{C}^{-1}\boldsymbol{\psi}(\mathbf{S})\widetilde{\boldsymbol{\Sigma}}^{(r)}) - \log\det\boldsymbol{\Gamma}$$
$$+ 2\Re\{\mathbf{y}^H\mathbf{C}^{-1}\boldsymbol{\psi}(\mathbf{S})\boldsymbol{\mu}^{(r)}\} - \text{trace}\{\boldsymbol{\Gamma}^{-1}\widetilde{\boldsymbol{\Sigma}}^{(r)}\} \quad (18)$$

where

$$\widetilde{\boldsymbol{\Sigma}}^{(r)} = \boldsymbol{\Sigma}^{(r)} + \boldsymbol{\mu}^{(r)}(\boldsymbol{\mu}^{(r)})^H \qquad (19)$$
$$\boldsymbol{\Sigma}^{(r)} = (\boldsymbol{\Gamma}^{(r)})^{-1} + \boldsymbol{\psi}^H(\mathbf{S})\mathbf{C}^{-1}\boldsymbol{\psi}(\mathbf{S}) \qquad (20)$$
$$\boldsymbol{\mu}^{(r)} = \boldsymbol{\Sigma}^{(r)}\boldsymbol{\psi}^H(\mathbf{S})\mathbf{C}^{-1}\mathbf{y} \qquad (21)$$
$$\mathbf{C} \triangleq \mathbb{E}\{\mathbf{e}\mathbf{e}^H\} = \text{diag}\{\sigma^2\mathbf{W}_1^H\mathbf{W}_1, \ldots, \sigma^2\mathbf{W}_K^H\mathbf{W}_K\}. \quad (22)$$

The derivation of these expressions have been omitted for the sake of brevity. The initial values of $\boldsymbol{\Gamma}$ and $\mathbf{S}$ are chosen as $\boldsymbol{\Gamma}^{(0)} = \mathbf{I}$ and $\mathbf{S}^{(0)} = \widetilde{\mathbf{S}}$, where $\widetilde{\mathbf{S}}$ is the estimate of the data obtained at the output of the RF combiner $\mathbf{W}^d$. The M-step is then given as

$$\{\boldsymbol{\gamma}^{(r+1)}, \mathbf{S}^{(r+1)}\} = \arg\max_{\boldsymbol{\gamma}, \mathbf{S}}Q(\boldsymbol{\gamma}, \mathbf{S}|\boldsymbol{\gamma}^{(r)}, \mathbf{S}^{(r)}) . \quad (23)$$

From (18) and (23), the optimization over the parameters $\boldsymbol{\gamma}$ and $\mathbf{S}$ is separable. Thus, the optimal values of $\boldsymbol{\gamma}^{(r+1)}$ and $\mathbf{S}^{(r+1)}$ can be obtained separately as

$$\boldsymbol{\gamma}^{(r+1)} = \text{diag}\{\widetilde{\boldsymbol{\Sigma}}^{(r+1)}\} \qquad (24)$$
$$\mathbf{s}_k^{(r+1)} = \arg\min_{\mathbf{s}_k \in \chi}\{-\text{trace}(\boldsymbol{\psi}_k^H(\mathbf{s}_k)\mathbf{C}^{-1}\boldsymbol{\psi}_k(\mathbf{s}_k)\widetilde{\boldsymbol{\Sigma}}^{(r)})$$
$$+ 2\Re\{\mathbf{y}^H\mathbf{C}^{-1}\boldsymbol{\psi}_k(\mathbf{s}_k)\boldsymbol{\mu}^{(r)}\}\} \quad \forall k . \qquad (25)$$

At the end of every Jth data block, the MS returns the location and values of the non-zero entries of $\widehat{\mathbf{d}}$ back to the BS using a side channel, as in [12].

D. Implementation with Hybrid Architecture

The precoding and combining vectors $\mathbf{f}_k^t$ and $\mathbf{W}_k^t$ are obtained by projecting $\widetilde{\mathbf{f}}_k^t$ and $\widetilde{\mathbf{W}}_k^t$, which are obtained by setting random phase shifts in the analog phase-shifters, onto the left and right null spaces of the estimated channel matrix. Then, using the channel state information $\widehat{\mathbf{H}} = $
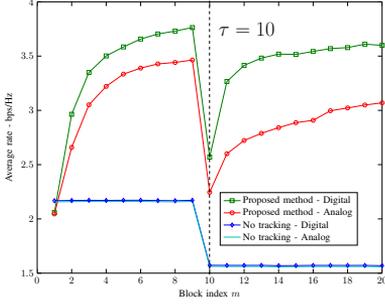
Fig. 2. Plot of average rate when the LOS path is blocked.

Fig. 3. Plot of average rate when the LOS path becomes visible.

$\mathbf{A}_{\text{BS}}\widehat{\mathbf{D}}\mathbf{A}_{\text{MS}}^H = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}^H$ at both the BS and MS, the precoding and combining matrices are chosen as

$$\mathbf{F}^{\text{d}} = \mathbf{V}_{\text{s}} \tag{26}$$

$$\mathbf{f}_k^{\text{t}} = \mathbf{V}_{\text{n}}\mathbf{V}_{\text{n}}^H\widetilde{\mathbf{f}}_k^{\text{t}} \tag{27}$$

$$\mathbf{W}^{\text{d}} = \mathbf{U}_{\text{s}} \tag{28}$$

$$\mathbf{W}_k^{\text{t}} = \mathbf{U}_{\text{n}}\mathbf{U}_{\text{n}}^H\widetilde{\mathbf{W}}_k^{\text{t}} \tag{29}$$

To implement the proposed method with the hybrid architecture, $N_{\text{BS}}^{\text{RF-T}} < N_{\text{BS}}^{\text{RF}}$ and $N_{\text{MS}}^{\text{RF-T}} < N_{\text{MS}}^{\text{RF}}$ RF chains are reserved for training at the transmitter and receiver. The precoding matrices for data and training are obtained by solving the following problems

$$\left\{\left[\mathbf{F}_{\text{RF}}^{\text{d}}\right]^{\star}, \left[\mathbf{F}_{\text{BB}}^{\text{d}}\right]^{\star}\right\} = \arg\min_{\mathbf{F}_{\text{RF}}^{\text{d}} \in \mathcal{A}_{\text{can}}, \mathbf{F}_{\text{BB}}^{\text{d}}} \left\|\mathbf{F}^{\text{d}} - \mathbf{F}_{\text{RF}}^{\text{d}}\mathbf{F}_{\text{BB}}^{\text{d}}\right\|_F^2$$
$$\text{subject to } \left\|\mathbf{F}_{\text{RF}}^{\text{d}}\left[\mathbf{F}_{\text{BB}}^{\text{d}}\right]_{*,p}\right\|^2 = 1 \,\forall p \tag{30}$$

$$\left\{\left[\mathbf{F}_{\text{RF},k}^{\text{t}}\right]^{\star}, \left[\mathbf{F}_{\text{BB},k}^{\text{t}}\right]^{\star}\right\} = \arg\min_{\mathbf{F}_{\text{RF},k}^{\text{t}} \in \mathcal{A}_{\text{can}}, \mathbf{f}_{\text{BB},k}^{\text{t}}} \left\|\mathbf{f}_k^{\text{t}} - \mathbf{F}_{\text{RF},k}^{\text{t}}\mathbf{f}_{\text{BB},k}^{\text{t}}\right\|_F^2$$
$$\text{subject to } \left\|\mathbf{F}_{\text{RF},k}^{\text{t}}\mathbf{f}_{\text{BB},k}^{\text{t}}\right\|^2 = 1 \,. \tag{31}$$

where $\mathcal{A}_{\text{can}}$ is the set of candidate codebook vectors which satisfy the unit modulus and quantized phase-shift constraints [11]. A similar optimization problem can also be formulated for the combining vectors at the MS. The optimization problem can be solved using orthogonal matching pursuit as in [16]. Then, the BS sends the indices of the vectors in $\mathcal{A}_{\text{can}}$ and the values of $\left[\mathbf{F}_{\text{BB}}^{\text{d}}\right]^{\star}$ and $\left[\mathbf{f}_{\text{BB},k}^{\text{t}}\right]^{\star}$ to the MS using a side channel to construct the measurement matrix $\boldsymbol{\psi}\left(\cdot\right)$.

V. SIMULATION RESULTS

We compare the average rate achieved by the existing method without channel tracking [12] with the proposed method in two scenarios. The average rate is calculated for a 16-QAM constellation. In both scenarios, the channel is composed of one LOS path and two NLOS clusters each containing $P = 20$ paths with a cluster width of 0.3 radians ($17°$) and path-loss coefficient that is 10 dB lower than that of the LOS path. S_{LOS} and S_{NLOS} denote the state of the LOS and NLOS paths and the state of these paths changes at the $\tau = 10$th block. For both methods, the initial channel is estimated using the SBL-based sparse recovery algorithm with $M_{\text{BS}} = 12$ training symbols and $M_{\text{MS}} = 6$

measurements per training symbol. The BS is equipped with $N_{\text{BS}} = 32$ antennas and $N_{\text{BS}}^{\text{RF}} = 4$ RF chains, and the MS is equipped with $N_{\text{MS}} = 16$ antennas and $N_{\text{MS}}^{\text{RF}} = 2$ RF chains. The number of grid points for AoA and AoD quantization are set as $G_{\text{MS}} = 16$ and $G_{\text{BS}} = 32$. For training, $N_{\text{BS}}^{\text{RF-T}} = N_{\text{MS}}^{\text{RF-T}} = 1$ RF chain is reserved at the BS and MS. The data transmission is performed over $M = 20$ blocks with $K = 484$ data symbols per block, at SNR of -10dB. For the proposed method ρ_{d} is set to 0.8 and the feedback interval J is set to 1, whereas for the case where there is no channel tracking, all the transmit power is allocated for data transmission. In addition, we assume that the LOS path is attenuated by 20 dB when blocked.

In Fig. 2, the average rate is plotted for the LOS path blockage scenario. Both S_{LOS} and S_{NLOS} are 1 when $m < \tau$, whereas S_{LOS} becomes 0 when $m \geq \tau$. While both the proposed and existing methods suffer a loss in throughput due to blockage, the proposed method estimates the parameters of the remaining clusters in the channel without the need for a separate channel training interval. In addition, the higher SNR during data transmission leads to a lower mean-squared error for the channel estimate and a corresponding higher throughput, since the channel and data are estimated jointly.

Fig. 3 shows results for the scenario when $S_{\text{NLOS}} = 1$ and $S_{\text{LOS}} = 0$ when $m < \tau$, whereas for $m \geq \tau$, the LOS path is no longer blocked. While both Figs. 2 and 3 represent abrupt changes to the channel, the latter change is not detrimental and results in the availability of the stronger LOS path. As a result, the rate achieved by the proposed method increases when $m \geq \tau$ since it estimates the parameters of the LOS path and orients the beam in that direction. However, the method without tracking identifies the new LOS path only during the subsequent channel training interval.

VI. CONCLUSION

We have proposed an approach for channel tracking in mmWave MIMO systems that involves overlaying the training and data transmissions. The proposed approach uses the SBL framework and is capable of tracking abrupt changes in the channel without requiring any prior information on the temporal variation of the channel. The overlaid channel training has been simulated in two extreme scenarios and is shown to be capable of tracking abrupt changes in the channel.

REFERENCES

[1] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[2] M. Jacob, S. Priebe, T. Krner, M. Peter, M. Wisotzki, R. Felbecker, and W. Keusgen, "Extension and validation of the ieee 802.11ad 60 ghz human blockage model," in *Proc. 7th European Conf. on Antennas and Propagation (EuCAP)*, Gothenburg, Sweden, April 2013, pp. 2806–2810.

[3] X. An, C. S. Sum, R. V. Prasad, J. Wang, Z. Lan, J. Wang, R. Hekmat, H. Harada, and I. Niemegeers, "Beam switching support to resolve link-blockage problem in 60 GHz WPANs," in *Proc. IEEE 20th Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Tokyo, Japan, Sep. 2009, pp. 390–394.

[4] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "Beamspy: Enabling robust 60 GHz links under blockage," in *Proc. 13th USENIX Symp. on Networked Systems Design and Implementation (NSDI 16)*, Santa Clara, CA, 2016, pp. 193–206.

[5] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[6] A. Patra, L. Simić, and P. Mähönen, "Smart mm-Wave beam steering algorithm for fast link re-establishment under node mobility in 60 GHz indoor WLANs," in *Proc. 13th ACM Int. Symp. on Mobility Management and Wireless Access*, Cancun, Mexico, 2015, pp. 53–62.

[7] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

[8] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, Dec. 2014, pp. 416–421.

[9] J. Palacios, D. D. Donno, D. Giustiniano, and J. Widmer, "Speeding up mmwave beam training through low-complexity hybrid transceivers," in *Proc. IEEE 27th Annu. Int. Symp. on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 1–7.

[10] M. Coldrey and P. Bohlin, "Training-based MIMO systems – part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov 2007.

[11] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[12] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Australia, Apr. 2015, pp. 2909–2913.

[13] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, to be published.

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

[15] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, Jul. 2014.

[16] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

Publication VIII

K. Upadhya, S. A. Vorobyov, R. W. Heath Jr. Low-overhead receiver-side channel tracking for mmWave MIMO. *in Proc. IEEE Int. Conf. Acoustics Speech Signal Process.*, pp. 3859–3863, Apr. 2018.

LOW-OVERHEAD RECEIVER-SIDE CHANNEL TRACKING FOR MMWAVE MIMO

Karthik Upadhya*   *Sergiy A. Vorobyov**   *Robert W. Heath Jr.* [†]

*Department of Signal Processing and Acoustics, Aalto University, Espoo 02150, Finland
[†]Wireless Networking and Communications Group, The University of Texas at Austin, Austin, TX 78712
E-mails: karthik.upadhya@aalto.fi, svor@ieee.org, rheath@utexas.edu

ABSTRACT

Millimeter wave (mmWave) multiple-input multiple-output (MIMO) transceivers employ narrow beams to obtain a large array-gain, rendering them sensitive to changes in the angles of arrival and departure of the paths. Since the singular vectors that span the channel subspace are used to design the precoder and combiner, we propose a method to track the receiver-side channel subspace during data transmission using a separate radio frequency (RF) chain dedicated for channel tracking. Under certain conditions on the transmit precoder, we show that the receiver-side channel subspace can be estimated during data transmission without knowing the structure of the precoder or the transmitted data. The performance of the proposed method is evaluated through simulations.

Index Terms— Massive MIMO, channel subspace tracking, mmWave communication.

1. INTRODUCTION

Channel estimation and tracking is essential for realizing high bandwidth communication links for millimeter-wave (mmWave) multiple-input multiple-output (MIMO) systems [1–3]. Since mmWave transceivers employ large number of antennas, channel estimation requires a large communication overhead. Channel tracking algorithms minimize this overhead by using the temporal correlation of the channel to narrow the search range of the unknown parameters. Furthermore, in mmWave transceivers that employ hybrid precoding and combining, the analog phase-shifters/switches have a non-negligible settling time when their values are changed, adding to the estimation/tracking overhead [4–6].

Several methods for mmWave channel tracking have been recently proposed. In [7], the precoding and combining matrices have been optimized to minimize the Cramér-Rao lower bound when the angle-of-arrivals (AoAs) and angle-of-departures (AoDs) lie in a confidence interval around the previously estimated values. However, this method requires the receiver to know the precoding matrices used at the transmitter for channel estimation. The methods in [5, 8, 9] track a single AoA and AoD with low overhead, but they have been

designed for analog beamforming and extensions to hybrid architecture are not straightforward.

In this paper, we propose a low-overhead method to track the channel subspace at the receiver during data transmission under the assumption that the receiver has extra radio frequency (RF) chains dedicated for training. The subspace is estimated using the angular distribution of the energy of the received signal. Therefore, this method can be viewed as a generalization of beam-training methods which use the energy of the received signal to select/track the beam. The channel subspace estimation during data transmission results in a significantly lower overhead for the following reasons: (i) tracking the receiver-side channel does not require dedicated pilot transmission and (ii) the settling time of the phase shifters/switches when using different transmission beams for estimating the channel is subsumed within data transmission. The tracking performance of the proposed method is demonstrated by means of simulations.

Notation: Vector and matrix are denoted as $\mathbf{a}$ and $\mathbf{A}$, respectively, $\mathbf{I}$ is the identity matrix, $\mathrm{span}\{\mathbf{A}\}$ is the column space of $\mathbf{A}$. The diagonal elements of $\mathbf{A}$ are denoted by $\mathrm{diag}(\mathbf{A})$ and $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with the values of $\mathbf{a}$ on its diagonal. Subspace dimension is denoted as $\dim\{\cdot\}$, $\mathrm{rank}\{\mathbf{A}\}$ is the rank of $\mathbf{A}$, and $\mathrm{vec}(\mathbf{A})$ is the vectorization of $\mathbf{A}$. Operations $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^\dagger$ are the transpose, conjugate, Hermitian transpose, and pseudo-inverse respectively, while $\otimes$ is the Kronecker product, $\|\mathbf{x}\|_0$ is the number of non-zero elements in $\mathbf{x}$, and $\|\mathbf{x}\|_2$ is the ℓ_2 norm. Complex Gaussian distribution of independent variables with zero mean and variance σ^2 is denoted as $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$.

2. SYSTEM MODEL AND INITIAL CHANNEL ESTIMATION

Consider a system with an AP having N_{AP} antennas with $N_{\mathrm{AP}}^{\mathrm{RF}}$ RF chains. The AP communicates with a user equipment (UE) having N_{UE} antennas and $N_{\mathrm{UE}}^{\mathrm{RF}}$ RF chains. Out of the $N_{\mathrm{UE}}^{\mathrm{RF}}$ RF chains at the UE, $N_{\mathrm{UE}}^{\mathrm{RF-T}}$ RF chains are allocated for training at the receiver during data transmission. Likewise, $N_{\mathrm{AP}}^{\mathrm{RF-T}}$ out of $N_{\mathrm{AP}}^{\mathrm{RF}}$ RF chains at the AP are reserved for training. We assume that N_{s} data streams are trans-

mitted from the AP to the UE and the channel is estimated in the downlink (DL). If the AP uses the precoder $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{AP}} \times N_{\mathrm{s}}}$ and UE uses the combiner $\mathbf{W} \in \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{s}}}$, the received observation vector at the UE can be written as

$$\mathbf{y} = \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{s} + \mathbf{W}^H \mathbf{q} \qquad (1)$$

where $\mathbf{s} \in \mathbb{C}^{N_{\mathrm{s}}}$ is the transmitted data vector and $\mathbf{q} \in \mathbb{C}^{N_{\mathrm{UE}}}$ is the additive noise vector distributed as $\mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$. The precoder $\mathbf{F}$ consists of the analog $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{AP}} \times N_{\mathrm{AP}}^{\mathrm{RF}}}$ and digital $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{AP}}^{\mathrm{RF}} \times N_{\mathrm{s}}}$ precoders such that $\mathbf{F} = \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}$. Similarly, the combiner $\mathbf{W}$ is the product of $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{UE}}^{\mathrm{RF}}}$ and $\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{UE}}^{\mathrm{RF}} \times N_{\mathrm{s}}}$.

The channel is assumed to be composed of P paths, and the AP and UE employ one-dimensional uniform linear arrays (ULA). If α_p, ϕ_p, and ψ_p denote the path gain, AoA, and AoD of the pth path, respectively, the channel matrix from the AP to UE can be written as

$$\mathbf{H} = \sum_{p=0}^{P-1} \alpha_p \mathbf{a}_{\mathrm{UE}}\left(\phi_p\right) \mathbf{a}_{\mathrm{AP}}^H\left(\psi_p\right) = \mathbf{A}_{\mathrm{UE}} \mathbf{D} \mathbf{A}_{\mathrm{AP}}^H \qquad (2)$$

where $\mathbf{a}_{\mathrm{UE}}\left(\cdot\right)$ and $\mathbf{a}_{\mathrm{AP}}\left(\cdot\right)$ are the steering vectors at the UE and AP, respectively, $\mathbf{A}_{\mathrm{AP}} \triangleq \left[\mathbf{a}_{\mathrm{AP}}\left(\psi_0\right), \ldots, \mathbf{a}_{\mathrm{AP}}\left(\psi_{P-1}\right)\right]$, $\mathbf{A}_{\mathrm{UE}} \triangleq \left[\mathbf{a}_{\mathrm{UE}}\left(\phi_0\right), \ldots, \mathbf{a}_{\mathrm{UE}}\left(\phi_{P-1}\right)\right]$, and $\mathbf{D} \triangleq \mathrm{diag}\left(\left[\alpha_0, \ldots \alpha_{P-1}\right]^T\right)$. In addition, we define matrices $\widetilde{\mathbf{A}}_{\mathrm{AP}} \in \mathbb{C}^{N_{\mathrm{AP}} \times P_{\mathrm{AP}}}$ and $\widetilde{\mathbf{A}}_{\mathrm{UE}} \in \mathbb{C}^{N_{\mathrm{UE}} \times P_{\mathrm{UE}}}$ that contain the steering vectors corresponding to $P_{\mathrm{AP}} \leq P$ unique AoDs at the AP and $P_{\mathrm{UE}} \leq P$ unique AoAs at the UE, respectively.

Assuming that the AoAs and AoDs are on a grid with G_{UE} and G_{AP} points, respectively, the channel matrix $\mathbf{H}$ can be approximated as

$$\mathbf{H} \approx \bar{\mathbf{A}}_{\mathrm{UE}} \bar{\mathbf{D}} \bar{\mathbf{A}}_{\mathrm{AP}}^H \qquad (3)$$

where $\bar{\mathbf{A}}_{\mathrm{AP}} \triangleq \left[\mathbf{a}_{\mathrm{AP}}\left(\psi_0\right), \ldots, \mathbf{a}_{\mathrm{AP}}\left(\psi_{G_{\mathrm{AP}}-1}\right)\right] \in \mathbb{C}^{N_{\mathrm{AP}} \times G_{\mathrm{AP}}}$, $\bar{\mathbf{A}}_{\mathrm{UE}} \triangleq \left[\mathbf{a}_{\mathrm{UE}}\left(\phi_0\right), \ldots, \mathbf{a}_{\mathrm{UE}}\left(\phi_{G_{\mathrm{UE}}-1}\right)\right] \in \mathbb{C}^{N_{\mathrm{UE}} \times G_{\mathrm{UE}}}$, and $\bar{\mathbf{D}} \in \mathbb{C}^{G_{\mathrm{UE}} \times G_{\mathrm{AP}}}$ is a sparse matrix with non-zero locations corresponding to the AoA and AoD pairs.

Initial channel training is performed in the DL, wherein the UE estimates the channel and feeds back the channel parameters to the AP [2, 10]. For estimating the channel during initial access, the AP transmits M_{AP} training symbols. For each of the M_{AP} symbols, the UE makes M_{UE} measurements. Therefore, the total overhead for initial channel estimation is $J = M_{\mathrm{AP}} M_{\mathrm{UE}}$ symbols. The AP designs its training precoding matrix $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{AP}} \times M_{\mathrm{AP}}}$ such that $\left[\mathbf{F}\right]_{i,j} = e^{j\xi}$ with ξ randomly chosen from the set of N_Q quantized phase shifts $\left\{0, 2\pi/N_Q, \ldots, 2\pi\left(N_Q - 1\right)/N_Q\right\}$ [10]. The elements of the combining matrix $\mathbf{W} \in \mathbb{C}^{N_{\mathrm{UE}} \times M_{\mathrm{UE}}}$ are also chosen randomly. The J received observations can be written as

$$\mathbf{Y} = \mathbf{W}^H \mathbf{H} \mathbf{F} + \mathbf{Q} \approx \mathbf{W}^H \bar{\mathbf{A}}_{\mathrm{UE}} \bar{\mathbf{D}} \bar{\mathbf{A}}_{\mathrm{AP}}^H \mathbf{F} + \mathbf{Q} \qquad (4)$$

where $\mathbf{Q} \triangleq \left[\mathbf{W}_0^H \mathbf{q}_0, \ldots, \mathbf{W}_{J-1}^H \mathbf{q}_{J-1}\right]$. Vectorizing the received observation matrix, we obtain

$$\mathbf{y} = \mathrm{vec}\left(\mathbf{Y}\right) = \left(\mathbf{F}^T \bar{\mathbf{A}}_{\mathrm{AP}}^* \otimes \mathbf{W}^H \bar{\mathbf{A}}_{\mathrm{UE}}\right) \mathbf{d} + \mathbf{q} \qquad (5)$$

where $\mathbf{d} = \mathrm{vec}\left(\bar{\mathbf{D}}\right)$ and $\mathbf{q} = \mathrm{vec}\left(\mathbf{Q}\right)$. Since $\mathbf{d}$ is a sparse vector, its elements can be estimated from (5) by solving the following optimization problem

$$\widehat{\mathbf{d}} = \min_{\mathbf{d}} \|\mathbf{d}\|_0 \quad \text{subject to} \quad \|\mathbf{y} - \boldsymbol{\Psi} \mathbf{d}\|_2 \leq \epsilon \qquad (6)$$

where $\boldsymbol{\Psi} \triangleq \mathbf{F}^T \bar{\mathbf{A}}_{\mathrm{AP}}^* \otimes \mathbf{W}^H \bar{\mathbf{A}}_{\mathrm{UE}}$. Then, the estimate of the channel can be obtained as $\widehat{\mathbf{H}} = \bar{\mathbf{A}}_{\mathrm{UE}} \widehat{\mathbf{d}} \bar{\mathbf{A}}_{\mathrm{AP}}^H$ and used to design the precoding and combining matrices $\mathbf{F}^{\mathrm{d}}$ and $\mathbf{W}^{\mathrm{d}}$.

3. CHANNEL SUBSPACE TRACKING

3.1. Proposed Approach

We consider channel estimation in the DL. The channel is assumed to remain constant for $M = M_{\mathrm{AP}} M_{\mathrm{UE}}$ blocks, each containing N symbols. To index a given block, we either use the single index ℓ or the pair of indexes $m \in \{1, \ldots, M_{\mathrm{AP}}\}$ and $n \in \{1, \ldots, M_{\mathrm{UE}}\}$, where ℓ and (m, n) are related to each other as $\ell = (m - 1)M_{\mathrm{UE}} + n$. The AP and UE use precoder $\mathbf{F}_m$ and combiner $\mathbf{W}_n$ in block (m, n). If $\mathbf{s}_{m,n}\left[k\right] \in \mathbb{C}^{N_{\mathrm{s}}}$ is the data transmitted in the DL in symbol k of block (m, n), the received observation can be written as

$$\mathbf{y}_{m,n}\left[k\right] = \mathbf{W}_n^H \mathbf{H} \mathbf{F}_m \mathbf{s}_{m,n}\left[k\right] + \mathbf{W}_n^H \mathbf{q}_{m,n}\left[k\right] \qquad (7)$$

where the precoding and combining matrices $\mathbf{F}_m \in \mathbb{C}^{N_{\mathrm{AP}} \times N_{\mathrm{s}}}$ and $\mathbf{W}_n \in \mathbb{C}^{N_{\mathrm{UE}} \times N_{\mathrm{UE}}^{\mathrm{RF}}}$ consist of both the training and data RF chains and are written, respectively, as $\mathbf{W}_n \triangleq \left[\mathbf{W}^{\mathrm{d}}, \mathbf{W}_n^{\mathrm{t}}\right]$ and $\mathbf{F}_m \triangleq \left[\mathbf{F}^{\mathrm{d}}, \mathbf{F}_m^{\mathrm{t}}\right]$. The phase-shifters corresponding to $\mathbf{F}^{\mathrm{d}}$ and $\mathbf{W}^{\mathrm{d}}$ are assumed to remain constant for the entire duration of the M blocks, whereas the phase-shifters corresponding to training remain constant within a block and are changed across each data block.

Computing the autocorrelation of the received observations in the (m, n)th block, we obtain

$$\begin{aligned} \mathbf{R}_{m,n} &\triangleq \mathbb{E}\left\{\mathbf{y}_{m,n}\left[k\right] \mathbf{y}_{m,n}^H\left[k\right]\right\} \\ &= \mathbf{W}_n^H \mathbf{H} \mathbf{F}_m \mathbf{F}_m^H \mathbf{H}^H \mathbf{W}_n + \sigma^2 \mathbf{W}_n^H \mathbf{W}_n \end{aligned} \qquad (8)$$

where the expectation is taken only over $\mathbf{s}_{m,n}\left[k\right]$. It has to be noted that $\mathbf{R}_{m,n}$ is computed within a single coherence block and the averaging is performed only over the received data. Therefore, unlike the matrices estimated in [11, 12], $\mathbf{R}_{m,n}$ is not an estimate of the second order statistics of the channel. Defining $\mathbf{F} \triangleq \left[\mathbf{F}_1, \ldots, \mathbf{F}_{M_{\mathrm{AP}}}\right]$, the sum of the autocorrelation matrices obtained for a given combining matrix $\mathbf{W}_n$, when M_{AP} different precoding matrices are employed, can be written as

$$\mathbf{R}_n \triangleq \sum_{m=1}^{M_{\mathrm{AP}}} \mathbf{R}_{m,n} = \mathbf{W}_n^H \mathbf{X} \mathbf{W}_n + \sigma^2 M_{\mathrm{AP}} \mathbf{W}_n^H \mathbf{W}_n \qquad (9)$$

where $\mathbf{X} \triangleq \mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H$.

Let r be the rank of the channel matrix $\mathbf{H}$, $\boldsymbol{\Sigma}_{\mathrm{s}} \in \mathbb{C}^{r \times r}$ be the diagonal matrix containing the non-zero singular values, and $\mathbf{U}_{\mathrm{s}} \in \mathbb{C}^{N_{\mathrm{UE}} \times r}$ and $\mathbf{V}_{\mathrm{s}} \in \mathbb{C}^{N_{\mathrm{AP}} \times r}$ be the matrices consisting of the left and right singular vectors corresponding to the signal space of the channel matrix $\mathbf{H}$. Then, we have $\mathbf{H} = \mathbf{U}_{\mathrm{s}}\boldsymbol{\Sigma}_{\mathrm{s}}\mathbf{V}_{\mathrm{s}}^H$. In addition, from (2), we have $\mathbf{H} = \mathbf{A}_{\mathrm{UE}}\mathbf{D}\mathbf{A}_{\mathrm{AP}}^H$. Therefore, we have $\mathrm{span}\{\mathbf{V}_{\mathrm{s}}\} \subseteq \mathrm{span}\{\mathbf{A}_{\mathrm{AP}}\} = \mathrm{span}\{\widetilde{\mathbf{A}}_{\mathrm{AP}}\}$ and $\mathrm{span}\{\mathbf{U}_{\mathrm{s}}\} \subseteq \mathrm{span}\{\mathbf{A}_{\mathrm{UE}}\} = \mathrm{span}\{\widetilde{\mathbf{A}}_{\mathrm{UE}}\}$.

Consider now the term $\mathbf{X} = \mathbf{U}_{\mathrm{s}}\boldsymbol{\Sigma}_{\mathrm{s}}\mathbf{V}_{\mathrm{s}}^H\mathbf{F}\mathbf{F}^H\mathbf{V}_{\mathrm{s}}\boldsymbol{\Sigma}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}^H$. We then have the following proposition

Proposition 1. $\mathrm{span}\{\mathbf{X}\} = \mathrm{span}\{\mathbf{U}_{\mathrm{s}}\}$ *if and only if $\mathbf{F}$ is chosen such that $\mathbf{V}_{\mathrm{s}}^H\mathbf{F}$ has full row-rank.*

Proposition 1 is intuitive and its proof is relatively straightforward, and is therefore, omitted because of the space limitation. The proposition implies that the basis vectors of the space $\mathrm{span}\{\mathbf{U}_{\mathrm{s}}\}$ can be obtained without knowing $\mathbf{F}$ provided $\mathbf{F}$ is chosen such that the AP transmits data/training symbols in the directions of all the AoDs of the channel. In Section 3.2, an algorithm for estimating $\widetilde{\mathbf{A}}_{\mathrm{UE}}$, which is a basis for $\mathrm{span}\{\mathbf{U}_{\mathrm{s}}\}$, from $\mathbf{R}_n$ is described. Note that $\mathbf{R}_n$ is obtained at the output of the analog precoder. Therefore, the proposed method obtains the subspace of the full-dimension channel matrix from a reduced-dimension observation space.

3.2. Receiver-side Channel Subspace Estimation

Using the fact that the matrix $\mathbf{A}_{\mathrm{AP}}$ can have repeated columns when multiple paths have the same AoA with different AoDs, the channel model in (2) and its quantized version can be rewritten using $\widehat{\mathbf{A}}_{\mathrm{AP}}$ as

$$\mathbf{H} = \mathbf{A}_{\mathrm{UE}}\widetilde{\mathbf{D}}\widetilde{\mathbf{A}}_{\mathrm{AP}}^H \approx \bar{\mathbf{A}}_{\mathrm{UE}}\check{\mathbf{D}}\widetilde{\mathbf{A}}_{\mathrm{AP}}^H \qquad (10)$$

where $\widetilde{\mathbf{D}} \in \mathbb{C}^{P \times P_{\mathrm{AP}}}$ is a non-diagonal matrix containing the path gains and $\check{\mathbf{D}} \in \mathbb{C}^{G_{\mathrm{UE}} \times P_{\mathrm{AP}}}$ is the corresponding matrix when only the AoAs are quantized. The tall matrix $\check{\mathbf{D}}$ has sparse columns such that the locations of the non-zero entries correspond to the AoAs of the received paths.

Substituting (10) into (9), and defining $\mathbf{G} \triangleq \widetilde{\mathbf{A}}_{\mathrm{AP}}^H\mathbf{F}\mathbf{F}^H\widetilde{\mathbf{A}}_{\mathrm{AP}}$, (9) can be approximated as

$$\mathbf{R}_n \approx \mathbf{W}_n^H\bar{\mathbf{A}}_{\mathrm{UE}}\check{\mathbf{D}}\mathbf{G}\check{\mathbf{D}}^H\bar{\mathbf{A}}_{\mathrm{UE}}^H\mathbf{W}_n + \sigma^2 M_{\mathrm{AP}}\mathbf{W}_n^H\mathbf{W}_n . \quad (11)$$

Vectorizing (11) and defining $\mathbf{q}_n \triangleq \sigma^2 M_{\mathrm{AP}}\mathrm{vec}\left(\mathbf{W}_n^H\mathbf{W}_n\right)$ and $\boldsymbol{\Psi}_n \triangleq \left(\left(\bar{\mathbf{A}}_{\mathrm{UE}}^H\mathbf{W}_n\right)^T \otimes \mathbf{W}_n^H\bar{\mathbf{A}}_{\mathrm{UE}}\right)$, we obtain

$$\mathbf{r}_n = \mathrm{vec}\left(\mathbf{R}_n\right) = \boldsymbol{\Psi}_n\mathrm{vec}\left(\check{\mathbf{D}}\mathbf{G}\check{\mathbf{D}}^H\right) + \mathbf{q}_n . \quad (12)$$

Stacking the columns $\mathbf{r}_n$ in a long vector, we finally obtain

$$\mathbf{r} \triangleq \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_{M_{\mathrm{UE}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}_1 \\ \vdots \\ \boldsymbol{\Psi}_{M_{\mathrm{UE}}} \end{bmatrix} \mathrm{vec}\left(\check{\mathbf{D}}\mathbf{G}\check{\mathbf{D}}^H\right) + \begin{bmatrix} \mathbf{q}_1 \\ \vdots \\ \mathbf{q}_{M_{\mathrm{UE}}} \end{bmatrix} . \quad (13)$$

Since $\check{\mathbf{D}}$ has sparse columns, the matrix $\check{\mathbf{D}}\mathbf{G}\check{\mathbf{D}}^H$ and its vectorized version has sparse entries. In fact, it is straightforward to show that the diagonal values of $\check{\mathbf{D}}\mathbf{G}\check{\mathbf{D}}^H$ are sufficient to obtain the support of the AoAs if $\widehat{\mathbf{A}}_{\mathrm{AP}}^H\mathbf{F}$ has full row-rank. Therefore, the problem of receiver-side channel subspace estimation can be solved efficiently using any of the multiple available sparse recovery methods.

3.3. Choice of $\mathbf{F}_m^{\mathrm{t}}$ and $\mathbf{W}_n^{\mathrm{t}}$

Note that $\mathrm{span}\{\mathbf{V}_{\mathrm{s}}\} \subseteq \mathrm{span}\{\widetilde{\mathbf{A}}_{\mathrm{AP}}\}$. Then, the condition that the matrix $\mathbf{V}_{\mathrm{s}}^H\mathbf{F}$ is full row-rank is satisfied when $\mathbf{F}$ is chosen such that $\mathrm{span}\{\mathbf{V}_{\mathrm{s}}\} \subseteq \mathrm{span}\{\widetilde{\mathbf{A}}_{\mathrm{AP}}\} \subseteq \mathrm{span}\{\mathbf{F}\}$. However, since subspace estimation is performed during data transmission, $N_{\mathrm{AP}}^{\mathrm{RF}} - N_{\mathrm{AP}}^{\mathrm{RF-T}}$ RF chains at the AP and $N_{\mathrm{UE}}^{\mathrm{RF}} - N_{\mathrm{UE}}^{\mathrm{RF-T}}$ RF chains at the UE will be utilized for data transmission, implying that $\mathbf{F}^{\mathrm{d}}$ and $\mathbf{W}^{\mathrm{d}}$ in $\mathbf{F}_m$ and $\mathbf{W}_n$ are fixed. As a result, only the components $\mathbf{F}_m^{\mathrm{t}}$ and $\mathbf{W}_n^{\mathrm{t}}$ in $\mathbf{F}_m$ and $\mathbf{W}_n$ are variable.

The analog phase shifters can be set to random values, as in Section 2, allowing for parameter estimation in the entire angular range $[0, \pi]$ provided M_{AP} is sufficiently large. However, $\mathbf{F}_m^{\mathrm{t}}$ has to be chosen under the additional constraint that the training sequence does not interfere with the received data. Therefore, denoting the matrix containing random phase shifts as $\bar{\mathbf{F}}_m^{\mathrm{t}}$, the training precoder can be obtained by projecting $\bar{\mathbf{F}}_m^{\mathrm{t}}$ on the space that is orthogonal to the column-space of $\mathbf{F}^{\mathrm{d}}$ [13]

$$\mathbf{F}_m^{\mathrm{t}} = \boldsymbol{\Pi}_{\mathbf{F}^{\mathrm{d}}}^{\perp}\bar{\mathbf{F}}_m^{\mathrm{t}} \qquad (14)$$

where $\boldsymbol{\Pi}_{\mathbf{F}^{\mathrm{d}}}^{\perp} \triangleq \left[\mathbf{I} - \mathbf{F}^{\mathrm{d}}\left(\mathbf{F}^{\mathrm{d}H}\mathbf{F}^{\mathrm{d}}\right)^{-1}\mathbf{F}^{\mathrm{d}H}\right]$ is the projection matrix onto the space orthogonal to the columns space of $\mathbf{F}^{\mathrm{d}}$. The resulting $\mathbf{F}_m^{\mathrm{t}}$ can be implemented in a hybrid architecture using methods such as in [14].

3.4. Updating $\mathbf{W}^{\mathrm{d}}$

After estimating a basis for $\mathrm{span}\{\mathbf{U}_{\mathrm{s}}\}$ from $\mathbf{X}$, the objective is to design the receive combiner using this basis. If $\mathbf{B}$ correspond to any basis for the space $\mathrm{span}\{\mathbf{U}_{\mathrm{s}}\}$, note that there exists a matrix $\mathbf{P}$ with full column rank such that $\mathbf{U}_{\mathrm{s}} = \mathbf{B}\mathbf{P}$. Then, the receive combiner can be chosen to satisfy the zero-forcing condition $(\mathbf{W}^{\mathrm{d}})^H\mathbf{H}\mathbf{F}^{\mathrm{d}} = \mathbf{I}$, which will result in $\mathbf{W}^{\mathrm{d}}$ being obtained as[1]

$$\mathbf{W}^{\mathrm{d}} = \mathbf{B}\mathbf{P}^\dagger \qquad (15)$$

where $\mathbf{P} \triangleq \mathbf{B}^H\mathbf{H}\mathbf{F}^{\mathrm{d}}$.

As described in Section 3.1, the estimated angular support $\widehat{\mathbf{A}}_{\mathrm{UE}}$ is an estimate of the basis of $\mathrm{span}\{\mathbf{U}_{\mathrm{s}}\}$. In order to estimate $\mathbf{P}$, setting $\mathbf{B} = \widehat{\mathbf{A}}_{\mathrm{UE}}$ as the combiner, the received observations can be expressed as $\mathbf{Y} = \widehat{\mathbf{A}}_{\mathrm{UE}}^H\mathbf{H}\mathbf{F}^{\mathrm{d}}\mathbf{S} + \widehat{\mathbf{A}}_{\mathrm{UE}}^H\mathbf{Q}$.

---

[1]Note that if $\mathbf{F}^{\mathrm{d}}$ is known at the receiver, the condition $(\mathbf{W}^{\mathrm{d}})^H\mathbf{H}\mathbf{F}^{\mathrm{d}} = \mathbf{I}$ can be satisfied at the receiver without channel state information at the transmitter by choosing $\mathbf{W}^{\mathrm{d}}$ as in (15).
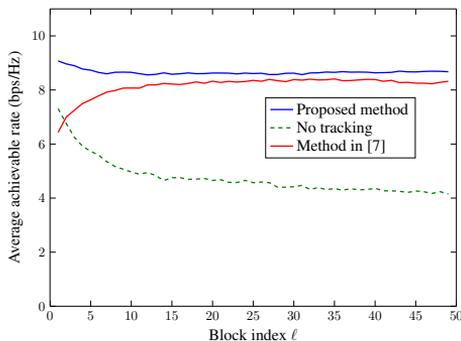
Fig. 1. Plot of the average achievable rate vs block index at SNR = 0 dB, $\sigma_\phi = 2°$ and $\sigma_\psi = 0$.

Using an $N_{\text{AP}}^{\text{RF}}$ length orthogonal pilot sequence $\mathbf{S}$, the least-squares (LS) estimate of $\mathbf{P}$ can be obtained as

$$\widehat{\mathbf{P}} = \mathbf{Y} \left(\mathbf{S}^H \mathbf{S}\right)^{-1} \mathbf{S}^H = \mathbf{P} + \widehat{\mathbf{A}}_{\text{UE}}^H \mathbf{Q} \left(\mathbf{S}^H \mathbf{S}\right)^{-1} \mathbf{S}^H . \quad (16)$$

Then, the receive combining matrix can be obtained from (15) as $\mathbf{W}^{\text{d}} = \widehat{\mathbf{A}}_{\text{UE}} \widehat{\mathbf{P}}^\dagger$. The resulting digital receive combiner $\mathbf{W}^{\text{d}}$ in is realized using the $N_{\text{UE}}^{\text{RF}} - N_{\text{UE}}^{\text{RF}-\text{T}}$ RF chains. Then, the corresponding baseband and RF combiners, $\mathbf{W}_{\text{BB}}^{\text{d}}$ and $\mathbf{W}_{\text{RF}}^{\text{d}}$, can be easily obtained using the approach in [14].

4. SIMULATION RESULTS

We evaluate here the performance of the proposed method in a time-varying channel in which only AoAs vary with time. The performance of the proposed method are compared with the receiver-side tracking method described in [7, Section IV.C]. We consider a system with $N_{\text{AP}} = 64$ antennas, $N_{\text{UE}} = 32$ antennas, $N_{\text{AP}}^{\text{RF}} = N_{\text{UE}}^{\text{RF}} = 4$ RF chains, and analog phase shifters with 7-bit phase shifts. At the UE, $N_{\text{UE}}^{\text{RF}-\text{T}} = 1$ RF chains are reserved for training, whereas at the AP, all the RF chains are used for data transmission. For initial channel estimation, $M_{\text{AP}} = 12$ and $M_{\text{UE}} = 3$ symbols are used. However, for channel subspace estimation during data transmission, $M_{\text{AP}} = 1$ and $M_{\text{UE}} = 20$ data blocks are employed. The number of data symbols in a data block N is chosen as 256, thereby implying that the channel is constant for a frame of $M_{\text{UE}} N = 20 \cdot 256 = 5120$ symbols. The AoDs and AoAs are quantized into grids with $G_{\text{AP}} = 128$ and $G_{\text{UE}} = 64$ grid points.

The channel is assumed to consist of a line-of-sight (LOS) path with AoA $\phi = 90°$ (the broad-side of the array) and a non-LOS (NLOS) cluster with AoA $\phi = 45°$. The NLOS cluster has 100 paths within an angular spread of $10°$ and 10 dB lower power than the LOS path. The angles of the paths change in every block and the angular difference between blocks is distributed as $\mathcal{N}\left(0, \sigma_\phi^2\right)$ for the AoAs and $\mathcal{N}\left(0, \sigma_\psi^2\right)$ for the AoDs. The AoDs are assumed to be
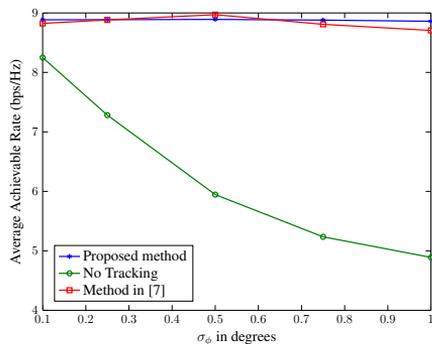
Fig. 2. Plot of the average achievable rate vs σ_ϕ at SNR = 0 dB and $\sigma_\psi = 0$ at the $\ell = 50$th block.

static implying that $\sigma_\psi = 0$, whereas σ_ϕ is varied from $0.1°$ to $2°$. The path amplitudes vary according to the Gauss-Markov model with forgetting factor $\eta = 0.8$, i.e., if α_p^ℓ is the path amplitude of the pth path in the ℓth block, then $\alpha_p^{\ell+1} = \eta \alpha_p^\ell + (1 - \eta) u^{\ell+1}$ where $u^\ell \sim \mathcal{CN}(0, 1)$.

In Fig. 1, the average achievable rate is plotted for different data blocks, at 0 dB SNR and $\sigma_\phi = 2°$. The achievable rate is almost constant for the proposed method despite the large value of σ_ϕ. In Fig. 2, the average achievable rate at 50th block is plotted against σ_ϕ at 0 dB SNR. In can be seen that the proposed method offers almost constant throughput in both slowly and rapidly varying channel scenarios. In both Figs. 1 and 2, the proposed method offers a throughput close to that of the method in [7] despite a significantly lower overhead.

If τ_{s} symbols are required for the phase-shifter values to settle after they are changed, and if $\widehat{\mathbf{A}}_{\text{UE}}$ has $\widehat{P}_{\text{UE}}$ columns, the proposed method requires an overhead of $\lceil \widehat{P}_{\text{UE}}/N_{\text{UE}}^{\text{RF}} \rceil (\tau_{\text{s}} + N_{\text{s}})$ symbols to estimate $\mathbf{P}$. However, the method in [7] requires an overhead of $N_{\text{s}} \lceil r/N_{\text{UE}}^{\text{RF}} \rceil \tau_{\text{s}}$ symbols, in addition to the knowledge of $\mathbf{F}$, to estimate the channel. As an example, if $\lceil r/N_{\text{UE}}^{\text{RF}} \rceil = \lceil \widehat{P}_{\text{UE}}/N_{\text{UE}}^{\text{RF}} \rceil = 1$, $\tau_{\text{s}} = 64$ symbols (as in [15]), and $N_{\text{s}} = 4$ streams, the training overhead is 70 symbols for the proposed method and 256 symbols for the method in [7].

5. CONCLUSION

We have developed a method for tracking the channel-subspace at the receiver with negligible overhead, and derived conditions on the transmit precoder under which the channel subspace can be tracked without knowing the transmit precoder or data. The proposed method is particularly useful in scenarios wherein the AoAs vary significantly faster than the AoDs, which is the case with handheld devices communicating with a fixed AP. We have shown by simulations that the proposed method is capable of tracking rapidly varying channels.

6. REFERENCES

[1] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[2] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[3] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[4] R. Mndez-Rial, C. Rusu, N. Gonzlez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016.

[5] V. Va, X. Zhang, and R. W. Heath, "Beam switching for millimeter wave communication to support high speed trains," in *Proc. IEEE 82nd Veh. Technol. Conf.*, Boston, MA, Sept. 2015, pp. 1–5.

[6] A. Natarajan, S. K. Reynolds, M. D. Tsai, S. T. Nicolson, J. H. C. Zhan, D. G. Kam, D. Liu, Y. L. O. Huang, A. Valdes-Garcia, and B. A. Floyd, "A fully-integrated 16-element phased-array receiver in SiGe BiCMOS for 60-GHz communications," *IEEE J. Solid-State Circuits*, vol. 46, no. 5, pp. 1059–1075, May 2011.

[7] N. Garcia, H. Wymeersch, and D. T. M. Slock, "Optimal robust precoders for tracking the AoD and AoA of a mm-Wave path," *ArXiv*, 2017, http://arxiv.org/abs/1703.10978.

[8] J. Li, Y. Sun, L. Xiao, S. Zhou, and C. E. Koksal, "Analog Beam Tracking in Linear Antenna Arrays: Convergence, Optimality, and Performance," *ArXiv*, May 2017, http://arxiv.org/abs/1705.09378.

[9] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[10] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, South Brisbane, Australia, Apr. 2015, pp. 2909–2913.

[11] S. Park and R. W. Heath, "Spatial channel covariance estimation for mmWave hybrid MIMO architecture," in *Proc. 50th Asilomar Conf. Signals, Syst. and Comput.*, Pacific Grove, CA, Nov. 2016, pp. 1424–1428.

[12] P. A. Eliasi, S. Rangan, and T. S. Rappaport, "Low-rank spatial channel estimation for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2748–2759, May 2017.

[13] K. Upadhya, R. W. Heath Jr., and S. A. Vorobyov, "Tracking abruptly changing channels in mmWave systems using overlaid data and training," in *Proc. IEEE 7th Int. Workshop Computational Advances in Multi-Sensor Adaptive Processing*, Curacao, Dutch Antilles, Dec. 2017.

[14] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[15] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 GHz band," *IEEE Std 802.11ad-2012*, pp. 1–628, Dec. 2012.

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.

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS