The Dissertation Committee for Ahmed Abdulkareem Nageeb Youssef Alkhateeb certifies that this is the approved version of the following dissertation:

Millimeter Wave and Massive MIMO Communications for Next-Generation Wireless Systems

Committee:

Robert W. Heath Jr., Supervisor

Jeffrey G. Andrews

Geert Leus

Evdokia Nikolova

Sanjay Shakkottai

Millimeter Wave and Massive MIMO Communications for Next-Generation Wireless Systems

by

Ahmed Abdulkareem Nageeb Youssef Alkhateeb, B.E.; M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2016

Dedicated to my parents, my wife, and my children.

Acknowledgments

First of all, I thank Allah for all his blessings, and for giving me strength and ability to complete this work. Foremost, I would like to express my deep and sincere gratitude to my supervisor Prof. Robert W. Heath Jr. for his patience, encouragement, and technical guidance throughout this work. The truth is that this work would not have been done without his support on both the technical and personal levels. So, I really feel fortunate to have this chance of working under his supervision in my PhD. I would also like to thank my committee members Prof. Jeffrey G. Andrews, Prof. Geert Leus, Prof. Evdokia Nikolova, and Prof. Sanjay Shakkottai for their comments, insights, and guidance that greatly improved this work. I would like also to express my deep gratitude to my wife, for her continued support and encouragement, and to my children Youssef and Laila—my ultimate motivation. Lastly, but most importantly, I thank my great parents for their tremendous support and endless self-sacrifices.

Ahmed Alkhateeb

The University of Texas At Austin

July 2016

Millimeter Wave and Massive MIMO Communications for Next-Generation Wireless Systems

Publication No. _____

Ahmed Abdulkareem Nageeb Youssef Alkhateeb, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Robert W. Heath Jr.

Multiple-input multiple-output (MIMO) communication is expected to play a central role in future wireless systems through the deployment of a large number of antennas at the transmitters and receivers. In low-frequency systems, massive MIMO offers high multiplexing gains that boost system spectral efficiency. In millimeter wave (mmWave) systems, the deployment of large antenna arrays at both the base station and mobile users is necessary to guarantee sufficient received signal power. Realizing these systems in practice, however, requires addressing several key challenges: (i) fully-digital solutions are costly and power hungry, (ii) channel training and estimation process has high overhead, and (iii) precoders design optimization problems are non-trivial. In this dissertation, precoding and channel estimation strategies that address these challenges are proposed for both mmWave and massive MIMO systems. The proposed solutions adopt hybrid analog/digital architectures that divide

precoding/combining processing between RF and baseband domains and lead to savings in cost and power consumption. Further, the developed techniques leverage the structure and characteristics of mmWave and massive MIMO channels to reduce the training overhead and precoders design complexity. The main contributions of this dissertation are (i) developing a channel estimation solution for hybrid architecture based mmWave systems, exploiting the sparse nature of the mmWave channels, (ii) designing hybrid precoding algorithm for multi-user mmWave and massive MIMO systems, (iii) proposing a multi-layer precoding framework for massive MIMO cellular systems, and (iv) developing hybrid precoding and codebook solutions for frequency selective mmWave systems. Mathematical analysis as well as numerical simulations illustrate the promising performance of the proposed solutions, marking them as enabling technologies for mmWave and massive MIMO systems.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

MIMO communication with large numbers of antennas is a principal feature of next-generation wireless systems [1–4]. At low-frequency, base stations equipped with massive antenna numbers can simultaneously serve large numbers of users [1, 2]. This boosts system sum spectral efficiency thanks to the orders of magnitude increase in multiplexing gains. Further, massive MIMO focuses the transmission into small dimensional subspaces, which reduces the required transmit power [1, 5]. Moreover, the asymptotic orthogonality characteristics of massive MIMO systems allows simplified multi-user processing to achieve near-optimal performance.

Antennas also play a critical but different role at millimeter wave (mmWave) frequencies. In mmWave systems, deploying large antenna arrays at both the base station and mobile users is necessary to guarantee sufficient received signal power [4, 6]. This allows signal transmission with ultra high data rates thanks to large bandwidths available at the mmWave frequency bands [7–11]. Large-dimensional arrays also enable mmWave cellular systems to achieve good coverage performance [7, 9], rendering it as a key candidate for 5G systems.

The efficient realization of mmWave and massive MIMO gains in practice, however, face several challenges. Next, we discuss three key challenges, namely, the

hardware constraints, the channel acquisition overhead, and the precoding design complexity. Then, we will introduce our contributions for addressing these challenges.

1.1 Massive MIMO Hardware Constraints

Traditional MIMO systems normally perform the required signal processing for precoding, combining, and channel estimation, entirely at baseband. This implicitly assumes that a complete RF chain is dedicated for each antenna. In large antenna array systems, however, the high cost and power consumption of mixed signal components makes fully digital processing difficult to realize in practice [12]. This is particularly critical at mmWave systems. To overcome this challenge, an initial proposed solution was to move all the required processing to the analog domain, resulting in analog-only beamforming/combining architectures [13–15]. Analog-only beamforming is usually implemented using networks of phase shifters [16, 17]. These analog-only solutions, though, are mainly limited to single-stream transmission. Further, the hardware constraints on the RF components, like the difficulty of having variable gains and the availability of only quantized phase shifters, make it difficult to perform sophisticated multi-stream or multi-user processing using analog-only architectures.

To compromise between system performance and hardware limitations, hybrid analog/digital architectures were proposed [18–20]. As shown in Fig. 1.1, hybrid architectures divide the precoding processing between the analog and digital domains, which reduces the required number of RF chains. The hardware constraints associated with the hybrid architectures such as the limitations on the RF components and the

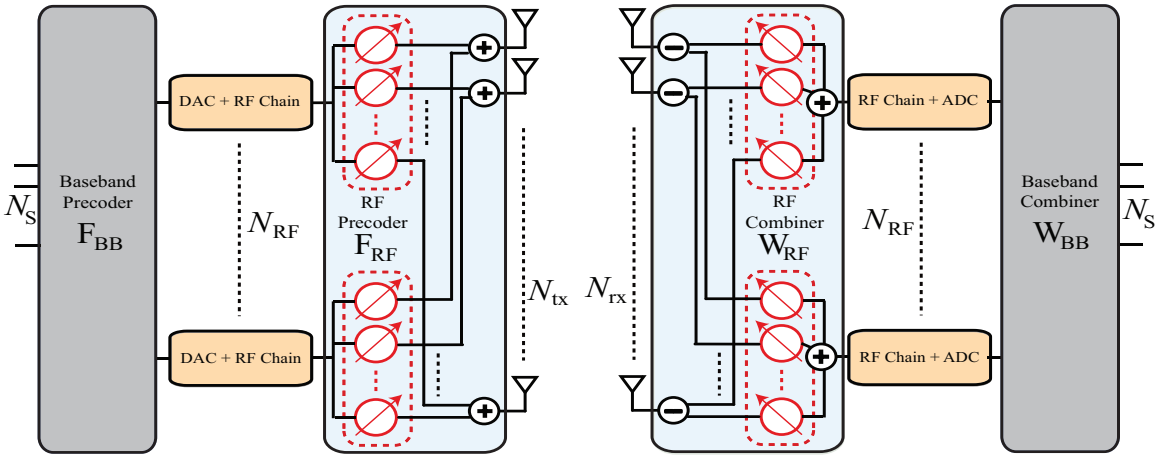


Figure 1.1: Hybrid analog/digital architectures divide the precoding/combining processing between analog and digital domains.

coupling between analog and digital precoders, however, impose new constraints on the precoding and channel estimation design problems. Initial work considered hybrid precoding design for point-to-point narrow-band MIMO channels. In [18, 19], joint analog/digital precoder design problem was considered for both spatial diversity and multiplexing systems. In large-scale mmWave systems, [20] leveraged the sparse structure of the channels [21], and proposed a low-complexity hybrid precoding/combining solution on the algorithmic concept of basis pursuit. More work, though, is required to extend these solutions to multi-user systems and frequency selective channels.

1.2 Massive MIMO Channel Aquisition Overhead

Acquiring channel state information is critical for efficient operation in wireless communication systems. The overhead associated with obtaining this channel knowledge, though, represents a major challenge. This overhead also scales when more

users are served or more antennas are deployed at the transmitters and receivers. The time/frequency resources overhead accompanying channel estimation/feedback and the possible error in channel estimation define the fundamental limits in massive MIMO systems. In [22], it was found that performance could be increased by reducing the number of transmit antennas in FDD massive MIMO systems. In [1], error due to channel estimation is shown to limit the performance in massive MIMO, even with very large antenna numbers. At lower frequency, some solutions have been proposed to overcome the massive MIMO channel acquisition overhead. In [23], large channel statistics were leveraged to reduce the dimensions of the effective channels in FDD massive MIMO systems, and hence save some training/feedback overhead. In [1], channel reciprocity was exploited to relax the training and feedback overhead in TDD massive MIMO systems.

Channel training and estimation is even more challenging at mmWave systems. In addition to the large training overhead associated with the large antennas, the SNR is typically low before beamforming design. Further, the hardware constraints, that results from RF/hybrid precoding, makes the channels at the baseband seen only through the RF lens. Initial solutions employed analog-only architectures and attempted to overcome the channel estimation problem by resting to beam training algorithms. In beam training, the BS and MS iteratively design the analog beamforming coefficients without channel knowledge at the transmitter. In [13, 15, 24], adaptive beamwidth beamforming algorithms and multi-stage codebooks were developed by which the transmitter and receiver jointly design their beamforming vectors. In [14], multiple beams with unique signatures were simultaneously used to mini-

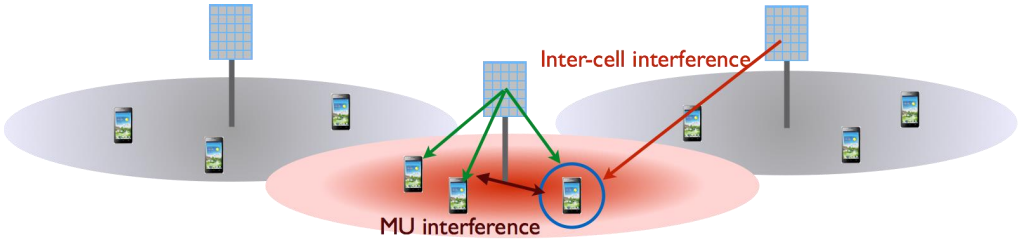


Figure 1.2: Designing precoders to manage different sources of interference in massive MIMO systems is of high complexity and requires training of large channel matrices.

mize the required beam training time. Despite the reduced complexity of [13–15, 24], they generally share the disadvantage of converging towards only one communication beam. Hence, these techniques are not capable of achieving multiplexing gains by sending multiple parallel streams. This motivates devising new channel estimation solutions for low-frequency and mmWave massive MIMO systems.

1.3 Massive MIMO Precoding Design Complexity

Harvesting the massive MIMO gains requires performing some sort of precoding/combining processing. The general objective of this precoding is to manage the different sources of interference and maximize the achievable rates of the serving users, see Fig. 1.2. In fact, this interference in the uplink and downlink transmission is the main limiting factors for the performance of massive MIMO systems. For example, the interference between the same-pilot users in the uplink channel training cause channel estimation errors, which ultimately limits the performance of TDD massive MIMO systems [1]. Further, handling the inter-cell interference usually requires some collaboration between the neighboring base stations. The overhead associated

with this collaboration, though, can be a limiting factor for system performance [25]. Moreover, the design of the precoders commonly leads to non-trivial and non-convex optimization problems [26], where closed-form solutions are difficult to obtain. This becomes more complex at massive MIMO systems because of the large-dimensional precoding matrices associated with the large channels. The hardware constraints resulting from employing practical architectures, such as hybrid analog/digital architectures, add more constraints and difficulties to the precoding design problems. All that motivates the need to develop low-complexity precoding solutions for massive MIMO systems.

1.4 Overview of Contributions

Enabling low-frequency and mmWave massive MIMO systems in practice requires addressing the key challenges discussed in Section 1.1-1.3. With this motivation, the problems tackled in this dissertation lie on the intersection of these challenges. We propose low-complexity precoding and channel estimation solutions that yield efficient performance while respecting the hardware constraints. The primary contributions of this dissertation can be summarized as follows.

1. We propose efficient channel estimation algorithms for mmWave systems with hybrid architectures. Leveraging the mmWave channel characteristics, we develop a sparse formulation of the mmWave channel estimation problem. Based on this formulation, we propose adaptive channel training algorithms that estimate the defining parameters of the multi-path mmWave channels. We also propose a hierarchical beamforming/combining codebook that constructs training

beamforming beams with arbitrary beamwidth, employing hybrid analog/digital architectures [27–29].

2. We propose and analyze a hybrid precoding algorithm for downlink multi-user massive MIMO systems. The proposed algorithm attempts to construct the precoding and combining matrices with low-training overhead. This solution is analytically proved to achieve asymptotically optimal performance compared to fully-digital precoding techniques. Further, we analyze the proposed hybrid precoding strategy under limited feedback settings, and draw insights into the quantization impact of both the analog and digital precoders on the system performance [30, 31].

3. We propose and analyze a general precoding framework, called multi-layer precoding, for full-dimensional and massive MIMO systems. Multi-layer precoding extends the hybrid precoding ideas to cellular systems, where both the inter-cell interference and multi-user interference are considered. Leveraging the directional structure of massive MIMO channels, the proposed solutions manages the different sources of interference while requiring low training overhead. We analyze multi-layer precoding and prove that it can achieve an asymptotic optimal performance for some special yet important channel models [32, 33].

4. We propose hybrid precoding and codebook designs for frequency selective mmWave MIMO systems. Considering an OFDM-based mmWave system, we find the optimal hybrid precoders for any given RF codebook. Based on that, we develop efficient codebooks for the analog and digital precoders and draw

insights into their performance at limited feedback systems. Further, we devise a low-complexity greedy solution for constructing near-optimal hybrid analog and digital precoders for frequency selective mmWave channels [34].

1.5 Notations and Abbreviations

We use the following notation throughout this thesis: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar, and $\mathcal{A}$ is a set. $|\mathbf{A}|$ is the determinant of $\mathbf{A}$, $\|\mathbf{A}\|_F$ is its Frobenius norm, whereas $\mathbf{A}^T$, $\mathbf{A}^*$,$\mathbf{A}^c$, $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$ are its transpose, Hermitian (conjugate transpose), conjugate, inverse, and pseudo-inverse respectively. $[\mathbf{A}]_{\mathcal{R},:}$ ($[\mathbf{A}]_{:,\mathcal{R}}$) are the rows (columns) of the matrix $\mathbf{A}$ with indices in the set $\mathcal{R}$, and $\text{diag}(\mathbf{a})$ is a diagonal matrix with the entries of $\mathbf{a}$ on its diagonal. $\mathbf{I}$ is the identity matrix and $\mathbf{1}_N$ is the N-dimensional all-ones vector. $\mathbf{A} \circ \mathbf{B}$ is the Khatri-Rao product of $\mathbf{A}$, and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$, and $\mathbf{B}$, and $\mathbf{A} \odot \mathbf{B}$ denotes the Hadamard product of $\mathbf{A}$, and $\mathbf{B}$. $\mathcal{N}(\mathbf{m}, \mathbf{R})$ is a complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{R}$. $\mathbb{E}[\cdot]$ is used to denote expectation.

The abbreviations used in this dissertation are summarized in Table 1.1

1.6 Organization

The rest of this dissertation is organized as follows. In Chapter 2, we propose a channel estimation solution for mmWave channels. Then, we consider multi-user mmWave systems in Chapter 3, and propose a low-complexity hybrid analog/digital precoding algorithm. This is then extended in Chapter 4 for full-dimensional and

massive MIMO systems, where out-ot-cell interference is considered. In Chapter 5, we consider frequency selective massive MIMO systems and develop hybrid precoding and codebook designs. Concluding remarks and future work are finally presented in Chapter 6.

Table 1.1: Summary of Abbreviations

AoA	Angle of Arrival
AoD	Angle of Departure
BD	Block Diagonalization
BS	Base Station
CS	Compressed Sensing
DFT	Discrete Fourier Transform
FDD	Frequency Division Duplexing
FD-MIMO	Full-Dimensional MIMO
LOS	Line of Sight
LTE	Long-Term Evolution
MIMO	Multiple-Input Multiple-Output
mmWave	Millimeter Wave
MS	Mobile Station
NLOS	Non Line of Sight
OFDM	Orthogonal Frequency Division Multiplexing
RVQ	Random Vector Quantization
SINR	Signal-to-Interference-Plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SVD	Single Value Decomposition
TDD	Time Division Duplexing
ULA	Uniform Linear Array
UPA	Uniform Planar Array
3GPP	The 3rd Generation Partnership Project

Chapter 2

Channel Estimation for Hybrid Architectures

2.1 Overview

In this chapter[1], we develop an adaptive algorithm to estimate the mmWave channel parameters when the transmitter and receiver employ hybrid analog/digital architectures. The proposed solution exploits the poor scattering nature of the channel. To enable the efficient operation of this algorithm, a novel hierarchical multi-resolution codebook is designed to construct training beamforming vectors with different beamwidths. For single-path channels, an upper bound on the estimation error probability using the proposed algorithm is derived, and some insights into the efficient allocation of the training power among the adaptive stages of the algorithm are obtained. The adaptive channel estimation algorithm is then extended to the multi-path case relying on the sparse nature of the channel. Using the estimated channel, this chapter proposes a new hybrid analog/digital precoding algorithm that overcomes the hardware constraints on the analog-only beamforming, and approaches the performance of digital solutions. Simulation results show that the proposed low-complexity

[1]This chapter is based on the work published in the journal paper: A. Alkhateeb, O. ElAyach, G. Leus, and R. W. Heath Jr, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 831-846, Oct. 2014. This work was supervised by Prof. Robert Heath. Prof. Geert Leus and Dr. Omar ElAyach provided important ideas for the channel estimation and hybrid precoding design that greatly improved the work.

channel estimation algorithm achieves comparable precoding gains compared to exhaustive channel training algorithms. The results also illustrate that the proposed channel estimation and precoding algorithms can approach the coverage probability achieved by perfect channel knowledge even in the presence of interference.

2.2 Introduction

MmWave communication is a promising technology for future outdoor cellular systems [6, 21, 35, 36]. Directional precoding with large antenna arrays appears to be inevitable to support longer outdoor links and to provide sufficient received signal power. Fortunately, large antenna arrays can be packed into small form factors at mmWave frequencies [16, 17], making it feasible to realize the large arrays needed for high precoding gains. The high power consumption of mixed signal components, however, makes digital baseband precoding impossible [6]. Moreover, the design of the precoding matrices is usually based on complete channel state information, which is difficult to achieve in mmWave due to the large number of antennas and the small SNR before beamforming. Because of the additional hardware constraints when compared with conventional microwave frequency MIMO systems, new channel estimation and precoding algorithms that are tailored to mmWave cellular systems must be developed.

To overcome the radio frequency (RF) hardware limitations, analog beamforming solutions were proposed in [13–15, 24]. The main idea is to control the phase of the signal transmitted by each antenna via a network of analog phase shifters. Several solutions, known as beam training algorithms, were proposed to iteratively

design the analog beamforming coefficients in systems without channel knowledge at the transmitter. In [13, 15, 24], adaptive beamwidth beamforming algorithms and multi-stage codebooks were developed by which the transmitter and receiver jointly design their beamforming vectors. In [14], multiple beams with unique signatures were simultaneously used to minimize the required beam training time. Despite the reduced complexity of [13–15, 24], they generally share the disadvantage of converging towards only one communication beam. Hence, these techniques are not capable of achieving multiplexing gains by sending multiple parallel streams. Moreover, the performance of analog strategies such as those in [13–15, 24] is sub-optimal compared with digital precoding solutions due to (i) the constant amplitude constraint on the analog phase shifters, and (ii) the potentially low-resolution signal phase control.

To achieve larger precoding gains, and to enable precoding multiple data streams, [18, 19, 37] propose to divide the precoding operations between the analog and digital domains. In [18], the joint analog-digital precoder design problem was considered for both spatial diversity and multiplexing systems. First, optimal unconstrained RF pre-processing signal transformations followed by baseband precoding matrices were proposed, and then closed-form sub-optimal approximations when RF processing is constrained by variable phase-shifters were provided. In [19], hybrid analog/digital precoding algorithms were developed to minimize the received signal's mean-squared error in the presence of interference when phase shifters with only quantized phases are available. The work in [18, 19], however, was not specialized for mmWave systems, and did not account for mmWave channel characteristics. In [20], the mmWave channel's sparse multi-path structure [21, 38], and the algorith-

mic concept of basis pursuit, were leveraged in the design of low-complexity hybrid precoders that attempt to approach capacity assuming perfect channel knowledge is available to the receiver. In [20,39], the hybrid precoding design problem was considered in systems where the channel is partially known at the transmitter. While the developed hybrid precoding algorithms in [18–20,39] overcome the RF hardware limitations and can support the transmission of multiple streams, the realization of these gains require some knowledge about the channel at the transmitter prior to designing the precoding matrices. This motivates developing multi-path mmWave channel estimation algorithms, which enable hybrid precoding to approach the performance of the digital precoding algorithms.

In this chapter, we develop low-complexity channel estimation and precoding algorithms for a mmWave system with large antenna arrays at both the BS and MS. These algorithms account for practical assumptions on the mmWave hardware in which (i) the analog phase shifters have constant modulus and quantized phases, and (ii) the number of RF chains is limited, i.e., less than the number of antennas. The main contributions of the work in this chapter can be summarized as follows:

- We propose a new formulation for the mmWave channel estimation problem. This formulation captures the sparse nature of the channel, and enables leveraging tools developed in the adaptive CS field to design efficient estimation algorithms for mmWave channels.

- We design a novel multi-resolution codebook for the training precoders. The new codebook relies on joint analog/digital processing to generate beamforming

vectors with different beamwidths, which is critical for proper operation of the adaptive channel estimation algorithms presented in the chapter.

- We design an adaptive CS based algorithm that efficiently estimates the parameters of mmWave channels with a small number of iterations, and with high success probability. The advantage of the proposed algorithm over prior beam training work appears in multi-path channels where our algorithm is able to estimate channel parameters. Hence, it enables multi-stream multiplexing in mmWave systems, while prior work [13–15, 24] was limited to the single-beam training and transmission.

- We analyze the performance of the proposed algorithm in single-path channels. We derive an upper bound on the error probability in estimating channel parameters, and find sufficient conditions on the total training power and its allocation over the adaptive stages of the algorithm to estimate the channel parameters with a certain bound on the maximum error probability.

- We propose a new hybrid analog/digital precoding algorithm for mmWave channels. In the proposed algorithm, instead of designing the precoding vectors as linear combinations of the steering vectors of the known angles of arrival/departure as assumed in [20], our design depends only on the quantized beamsteering directions to directly approximate the channel's dominant singular vectors. Hence, it implicitly considers the hardware limitations, and more easily generalizes to arbitrary antenna arrays.

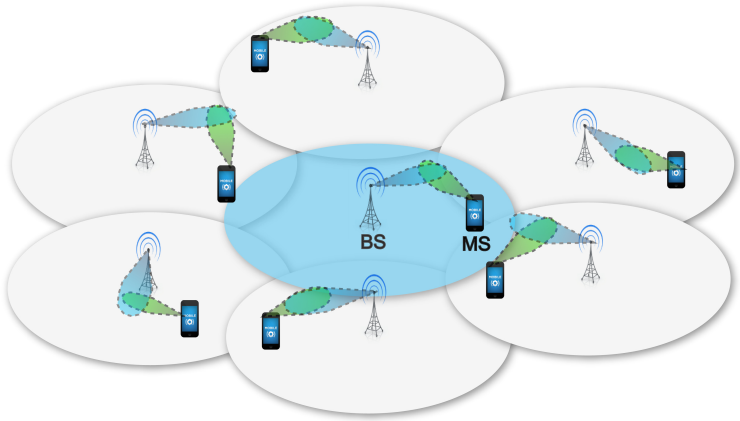


Figure 2.1: A mmWave cellular system model, in which BSs and MSs communicate via directive beamforming using large antenna arrays

- We evaluate the performance of the proposed estimation algorithm by simulations in a mmWave cellular system setting, assuming that both the BS and MS adopt hybrid precoding algorithms.

Simulation results indicate that the precoding gains given by the proposed channel estimation algorithm are close to that obtained when exhaustive search is used to design the precoding vectors. Multi-cell simulations show that the spectral efficiency and coverage probability achieved when hybrid precoding is used in conjunction with the proposed channel estimation strategy are comparable to that achieved when perfect channel knowledge and digital unconstrained solutions are assumed.

2.3 System Model

Consider the mmWave cellular system shown in Fig. 2.1. A BS with N_{BS} antennas and N_{RF} RF chains is assumed to communicate with a single MS with N_{MS}

antennas and N_{RF} RF chains as shown in Fig. 2.2. The number of RF chains at the MSs is usually less than that of the BSs in practice, but we do not exploit this fact in our model. The BS and MS communicate via N_{S} data streams, such that $N_{\mathrm{S}} \leq N_{\mathrm{RF}} \leq N_{\mathrm{BS}}$ and $N_{\mathrm{S}} \leq N_{\mathrm{RF}} \leq N_{\mathrm{MS}}$ [20].

In this chapter, we will focus on the downlink transmission. The BS is assumed to apply an $N_{\mathrm{RF}} \times N_{\mathrm{S}}$ baseband precoder $\mathbf{F}_{\mathrm{BB}}$ followed by an $N_{\mathrm{BS}} \times N_{\mathrm{RF}}$ RF precoder, $\mathbf{F}_{\mathrm{RF}}$. If $\mathbf{F}_{\mathrm{T}} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ is the $N_{BS} \times N_{\mathrm{S}}$ combined BS precoding matrix, the discrete-time transmitted signal is then

$$\mathbf{x} = \mathbf{F}_{\mathrm{T}}\mathbf{s}, \tag{2.1}$$

where $\mathbf{s}$ is the $N_{\mathrm{S}} \times 1$ vector of transmitted symbols, such that $\mathbb{E}\left[\mathbf{s}\mathbf{s}^{*}\right] = \frac{P_{\mathrm{S}}}{N_{\mathrm{S}}}\mathbf{I}_{N_{\mathrm{S}}}$, and P_{S} is the average total transmit power. Since $\mathbf{F}_{\mathrm{RF}}$ is implemented using analog phase shifters, its entries are of constant modulus. We normalize these entries to satisfy $\left|\left[\mathbf{F}_{\mathrm{RF}}\right]_{m,n}\right|^2 = N_{\mathrm{BS}}^{-1}$, where $\left|\left[\mathbf{F}_{\mathrm{RF}}\right]_{m,n}\right|$ denotes the magnitude of the (m,n)th element of $\mathbf{F}_{\mathrm{RF}}$. The total power constraint is enforced by normalizing $\mathbf{F}_{\mathrm{BB}}$ such that $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_{\mathrm{S}}$.

We adopt a narrowband block-fading channel model in which an MS observes the received signal

$$\mathbf{r} = \mathbf{H}\mathbf{F}_{\mathrm{T}}\mathbf{s} + \mathbf{n}, \tag{2.2}$$

where $\mathbf{H}$ is the $N_{\mathrm{MS}} \times N_{\mathrm{BS}}$ matrix that represents the mmWave channel between the BS and MS, and $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian noise corrupting the received signal.

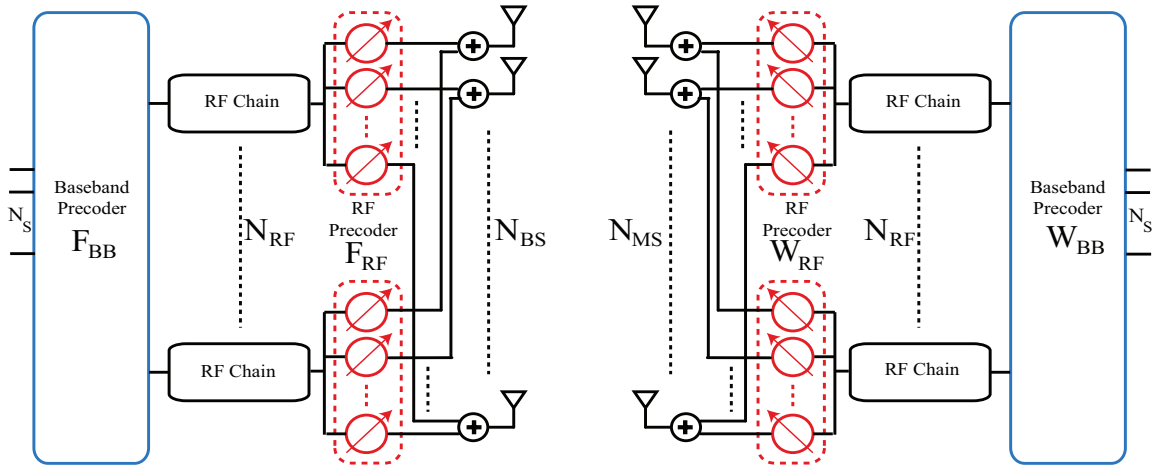


Figure 2.2: Block diagram of BS-MS transceiver that uses RF and baseband beamformers at both ends.

At the MS, the combiner $\mathbf{W}_{\mathrm{T}}$ composed of the RF and baseband combiners $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ is used to process the received signal $\mathbf{r}$ which results in

$$\mathbf{y} = \mathbf{W}^{*}\mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{W}^{*}\mathbf{n}. \tag{2.3}$$

We will explain the proposed algorithms for the downlink model. The same algorithms, however, can be directly applied to the uplink system whose input-output relationship is identical to (2.3) with $\mathbf{H}$ replaced by the uplink channel, and the roles of the precoders ($\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$) and combiners ($\mathbf{W}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{BB}}$) switched.

While the mmWave channel estimation and precoding algorithms developed in the following sections consider only a BS-MS link with no interfering BSs, these algorithms will also be numerically evaluated by simulations in the case of mmWave cellular systems where out-of-cell interference exists in Section 2.8.2.

Since mmWave channels are expected to have limited scattering [21, 40], we

adopt a geometric channel model with L scatterers. Each scatterer is further assumed to contribute a single propagation path between the BS and MS [20]. Under this model, the channel $\mathbf{H}$ can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{\rho}} \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}_{\mathrm{MS}}\left(\theta_\ell\right) \mathbf{a}_{\mathrm{BS}}^{*}\left(\phi_\ell\right), \tag{2.4}$$

where ρ denotes the average path-loss between the BS and MS, and α_ℓ is the complex gain of the ℓ^{th} path. The path amplitudes are assumed to be Rayleigh distributed, i.e., $\alpha_\ell \sim \mathcal{N}\left(0, \bar{P}_{\mathrm{R}}\right), \ell = 1, 2, ..., L$ with $\bar{P}_{\mathrm{R}}$ the average power gain. The variables $\phi_\ell \in [0, 2\pi]$ and $\theta_\ell \in [0, 2\pi]$ are the ℓ^{th} path's azimuth AoDs/AoAs of the BS and MS, respectively. Considering only the azimuth, and neglecting elevation, implies that all scattering happens in azimuth and that the BS and MS implement horizontal (2-D) beamforming only. Extensions to 3-D beamforming are possible [20]. Finally, $\mathbf{a}_{\mathrm{BS}}\left(\phi_\ell\right)$ and $\mathbf{a}_{\mathrm{MS}}\left(\theta_\ell\right)$ are the antenna array response vectors at the BS and MS, respectively. While the algorithms and results developed in this work can be applied to arbitrary antenna arrays, we use ULAs, in the simulations of Section 2.8. If a ULA is assumed, $\mathbf{a}_{\mathrm{BS}}\left(\phi_\ell\right)$ can be written as

$$\mathbf{a}_{\mathrm{BS}}\left(\phi_\ell\right) = \frac{1}{\sqrt{N_{\mathrm{BS}}}} \left[1, e^{j\frac{2\pi}{\lambda} d \sin(\phi_\ell)}, \ldots, e^{j(N_{\mathrm{BS}}-1)\frac{2\pi}{\lambda} d \sin(\phi_\ell)}\right]^T, \tag{2.5}$$

where λ is the signal wavelength, and d is the distance between antenna elements. The array response vectors at the MS, $\mathbf{a}_{\mathrm{MS}}\left(\theta_\ell\right)$, can be written in a similar fashion.

The channel in (2.4) is written in a more compact form as

$$\mathbf{H} = \mathbf{A}_{\mathrm{MS}} \mathrm{diag}\left(\boldsymbol{\alpha}\right) \mathbf{A}_{\mathrm{BS}}^{*}, \tag{2.6}$$

where $\boldsymbol{\alpha} = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{\rho}} \left[\alpha_1, \alpha_2, ..., \alpha_L\right]^T$. The matrices

$$\mathbf{A}_{\mathrm{BS}} = \left[\mathbf{a}_{\mathrm{BS}}\left(\phi_1\right),\ \mathbf{a}_{\mathrm{BS}}\left(\phi_2\right),\ ...,\ \mathbf{a}_{\mathrm{BS}}\left(\phi_L\right)\right], \tag{2.7}$$

and

$$\mathbf{A}_{\mathrm{MS}} = \left[\mathbf{a}_{\mathrm{MS}}\left(\theta_1\right),\ \mathbf{a}_{\mathrm{MS}}\left(\theta_2\right),\ ...,\ \mathbf{a}_{\mathrm{MS}}\left(\theta_L\right)\right], \tag{2.8}$$

contain the BS and MS array response vectors.

In this work, we assume that both the BS and MS have no a priori knowledge of the channel. Hence, in the first part of the chapter, namely, Section 2.4-Section 2.6, the mmWave channel estimation problem is formulated, and an adaptive CS based algorithm is developed and employed at the BS and MS to solve it. In the second part, i.e., Section 2.7, the estimated channel is used to construct the hybrid precoding and decoding matrices.

2.4 Formulation of the MmWave Channel Estimation Problem

Given the geometric mmWave channel model in (2.4), estimating the mmWave channel is equivalent to estimating the different parameters of the L channel paths; namely the AoA, the AoD, and the gain of each path. To do that accurately and with low training overhead, the BS and MS need to carefully design their training precoders and combiners. In this section, we exploit the poor scattering nature of the mmWave channel, and formulate the mmWave channel estimation problem as a sparse problem. We will also briefly show how adaptive CS work invokes some ideas for the

design of the training precoders and combiners. Inspired by these ideas, and using the hybrid analog/digital system architecture, we will develop a novel hierarchical multi-resolution codebook for the training beamforming vectors in Section 2.5. We will then propose algorithms that adaptively use the developed codebook to estimate the mmWave channel along with evaluating their performance in Section 2.6.

2.4.1 mmWave Channel Estimation: A Sparse Formulation

Consider the system and mmWave channel models described in Section 2.3. If the BS uses a beamforming vector $\mathbf{f}_p$, and the MS employs a measurement vector $\mathbf{w}_q$ to combine the received signal, the resulting signal can be written as

$$y_{q,p} = \mathbf{w}_q^* \mathbf{H} \mathbf{f}_p s_p + \mathbf{w}_q^* \mathbf{n}_{q,p}, \tag{2.9}$$

where s_p is the transmitted symbol on the beamforming vector $\mathbf{f}_p$, such that $\mathbb{E}\left[s_p s_p^*\right] = P$, with P the average power used per transmission in the training phase. In Section 2.5, we will develop a hybrid analog/digital design for the beamforming/ measurement vectors, $\mathbf{f}_p$ and $\mathbf{w}_q$. If M_{MS} such measurements are performed by the MS vectors $\mathbf{w}_q, q = 1, 2, ..., M_{\text{MS}}$ at M_{MS} successive instants to detect the signal transmitted over the beamforming vector $\mathbf{f}_p$, the resulting vector will be

$$\mathbf{y}_p = \mathbf{W}^* \mathbf{H} \mathbf{f}_p s_p + \text{diag}(\mathbf{W}^* \left[\mathbf{n}_{1,p}, ..., \mathbf{n}_{M_{\text{MS}},p}\right]), \tag{2.10}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{M_{\text{MS}}}]$ is the $N_{\text{MS}} \times M_{\text{MS}}$ measurement matrix. If the BS employs M_{BS} such beamforming vectors $\mathbf{f}_p, p = 1, ..., M_{\text{BS}}$, at M_{BS} successive time slots, and the MS uses the same measurement matrix $\mathbf{W}$ to combine the received

signal, the resultant matrix can then be written by concatenating the M_{BS} processed vectors $\mathbf{y}_p, p = 1, 2, ..., M_{\text{BS}}$

$$\mathbf{Y} = \mathbf{W}^* \mathbf{H} \mathbf{F} \mathbf{S} + \mathbf{Q}, \tag{2.11}$$

where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_{M_{\text{BS}}}]$ is the $N_{\text{BS}} \times M_{\text{BS}}$ beamforming matrix used by the BS, and $\mathbf{Q}$ is an $M_{\text{MS}} \times M_{\text{BS}}$ noise matrix given by concatenating the M_{BS} noise vectors. The matrix $\mathbf{S}$ is a diagonal matrix carrying the M_{BS} transmitted symbols $s_p, p = 1, ..., M_{\text{BS}}$ on its diagonal. For the training phase, we assume that all transmitted symbols are equal, namely, $\mathbf{S} = \sqrt{P}\mathbf{I}_{M_{\text{BS}}}$ and therefore

$$\mathbf{Y} = \sqrt{P}\mathbf{W}^* \mathbf{H} \mathbf{F} + \mathbf{Q}. \tag{2.12}$$

To exploit the sparse nature of the channel, we first vectorize the resultant matrix $\mathbf{Y}$

$$\mathbf{y}_{\text{v}} = \sqrt{P}\text{vec}\left(\mathbf{W}^* \mathbf{H} \mathbf{F}\right) + \text{vec}\left(\mathbf{Q}\right) \tag{2.13}$$

$$\overset{(a)}{=} \sqrt{P}\left(\mathbf{F}^T \otimes \mathbf{W}^*\right) \text{vec}\left(\mathbf{H}\right) + \mathbf{n}_{\text{Q}} \tag{2.14}$$

$$\overset{(b)}{=} \sqrt{P}\left(\mathbf{F}^T \otimes \mathbf{W}^*\right) \left(\mathbf{A}_{\text{BS}}^c \circ \mathbf{A}_{\text{MS}}\right) \boldsymbol{\alpha} + \mathbf{n}_{\text{Q}}, \tag{2.15}$$

where (a) follows from [41, Theorem 13.26], (b) follows from the channel model in (2.6), and the properties of the Khatri-Rao product, [42]. The matrix $\left(\mathbf{A}_{\text{BS}}^c \circ \mathbf{A}_{\text{MS}}\right)$ is an $N_{\text{BS}}N_{\text{MS}} \times L$ matrix in which each column has the form $\left(\mathbf{a}_{\text{BS}}^c\left(\phi_\ell\right) \otimes \mathbf{a}_{\text{MS}}\left(\theta_\ell\right)\right), \ell = 1, 2, ..., L$, i.e., each column ℓ represents the Kronecker product of the BS and MS array response vectors associated with the AoA/AoD of the ℓth path of the channel.

To complete the problem formulation, we assume that the AoAs, and AoDs are taken from a uniform grid of N points, with $N \gg L$, i.e., we assume that

$\phi_\ell, \theta_\ell \in \left\{0, \frac{2\pi}{N}, ..., \frac{2\pi(N-1)}{N}\right\}, \ell = 1, 2, ..., L$. As the values of the AoAs/AoDs are actually continuous, other off-grid based algorithms like sparse regularized total least squared [43], continuous basis pursuit [44], or Newton refinement ideas [45] can be incorporated to reduce the quantization error. In this work, we consider only the case of quantized AoAs/AoDs, leaving possible improvements for future work. We evaluate the impact of this quantization error on the performance of the proposed algorithms in this work by numerical simulations in Section 2.8.

By neglecting the grid quantization error, we can approximate $\mathbf{y}_{\mathrm{v}}$ in (2.15) as

$$\mathbf{y}_{\mathrm{v}} = \sqrt{P}\left(\mathbf{F}^T \otimes \mathbf{W}^*\right)\mathbf{A}_{\mathrm{D}}\mathbf{z} + \mathbf{n}_{\mathrm{Q}}, \tag{2.16}$$

where $\mathbf{A}_{\mathrm{D}}$ is an $N_{\mathrm{BS}}N_{\mathrm{MS}} \times N^2$ dictionary matrix that consists of the N^2 column vectors of the form $\left(\mathbf{a}^c_{\mathrm{BS}}\left(\bar{\phi}_u\right) \otimes \mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_v\right)\right)$, with $\bar{\phi}_u$, and $\bar{\theta}_v$ the uth, and vth points, respectively, of the angles uniform grid, i.e, $\bar{\phi}_u = \frac{2\pi u}{N}, u = 0, 2, ..., N-1$, and $\bar{\theta}_v = \frac{2\pi v}{N}, v = 0, 2, ..., N-1$. $\mathbf{z}$ is an $N^2 \times 1$ vector which carries the path gains of the corresponding quantized directions. Note that detecting the columns of $\mathbf{A}_{\mathrm{D}}$ that correspond to non-zero elements of $\mathbf{z}$, directly implies the detection of the AoAs and AoDs of the dominant paths of the channel. The path gains can be also determined by calculating the values of the corresponding elements in $\mathbf{z}$.

The formulation of the vectorized received signal $\mathbf{y}_{\mathrm{v}}$ in (2.16) represents a sparse formulation of the channel estimation problem as $\mathbf{z}$ has only L non-zero elements and $L \ll N^2$. This implies that the number of required measurements, $M_{\mathrm{BS}}M_{\mathrm{MS}}$, to detect the non-zero elements of $\mathbf{z}$ is much less than N^2. In other words, this means that the BS does not need to transmit along each vector defined

in the dictionary, nor does the MS need to observe signals using its entire codebook. Given this formulation in (2.16), CS tools can be leveraged to design estimation algorithms to determine the quantized AoAs/AoDs. If we define the sensing matrix $\mathbf{\Psi}$ as $\mathbf{\Psi} = \left(\mathbf{F}^T \otimes \mathbf{W}^*\right) \mathbf{A}_{\mathrm{D}}$, the objective of the CS algorithms will be to efficiently design this sensing matrix to guarantee the recovery of the non-zero elements of the vector $\mathbf{z}$ with high probability, and with a small number of measurements. One common criterion for that is the restricted isometry property (RIP), which requires the matrix $\mathbf{\Psi}^*\mathbf{\Psi}$ to be close to diagonal on average [46].

To simplify the explanation of the BS-MS beamforming vectors' design problem in the later chapters, we prefer to use the Kronecker product properties and write (2.16) as [42]

$$\mathbf{y}_{\mathrm{v}} = \sqrt{P}\left(\mathbf{F}^T \mathbf{A}^c_{\mathrm{BS,D}} \otimes \mathbf{W}^* \mathbf{A}_{\mathrm{MS,D}}\right) \mathbf{z} + \mathbf{n}_{\mathrm{Q}} \tag{2.17}$$

$$= \sqrt{P}\mathbf{F}^T \mathbf{A}^c_{\mathrm{BS,D}} \mathbf{z}_{\mathrm{BS}} \otimes \mathbf{W}^* \mathbf{A}_{\mathrm{MS,D}} \mathbf{z}_{\mathrm{MS}} + \mathbf{n}_{\mathrm{Q}}, \tag{2.18}$$

where $\mathbf{z}_{\mathrm{BS}}$, and $\mathbf{z}_{\mathrm{MS}}$ are two $N \times 1$ sparse vectors that have non-zero elements in the locations that correspond to the AoDs, and AoAs, respectively. $\mathbf{A}_{\mathrm{BS,D}}$, and $\mathbf{A}_{\mathrm{MS,D}}$ are $N_{\mathrm{BS}} \times N$, and $N_{\mathrm{MS}} \times N$ dictionary matrices that consist of column vectors of the forms $\mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_u\right)$, and $\mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_u\right)$, respectively.

In the standard CS theory, the number of measurement vectors required to guarantee the recovery of the L-sparse vector $\mathbf{z}$ with high probability is of order $\mathcal{O}(L \log(N/L))$ [47]. While these results are theoretically proved, their implementations to specific applications and the development of efficient algorithms require further work. We therefore resort to adaptive CS tools which invoke some ideas for

the design of the training beamforming vectors.

2.4.2 Adaptive Compressed Sensing Solution

In adaptive CS [48,49], the training process is divided into a number of stages. The training precoding, and measurement matrices used at each stage are not determined a priori, but rather depend on the output of the earlier stages. More specifically, if the training process is divided into S stages, then the vectorized received signals of these stages are

$$\begin{aligned}
\mathbf{y}_{(1)} &= \sqrt{P_{(1)}} \left(\mathbf{F}_{(1)}^{T} \mathbf{A}_{\mathrm{BS,D}}^{*} \otimes \mathbf{W}_{(1)}^{*} \mathbf{A}_{\mathrm{MS,D}}\right) \mathbf{z} + \mathbf{n}_1 \\
\mathbf{y}_{(2)} &= \sqrt{P_{(2)}} \left(\mathbf{F}_{(2)}^{T} \mathbf{A}_{\mathrm{BS,D}}^{*} \otimes \mathbf{W}_{(2)}^{*} \mathbf{A}_{\mathrm{MS,D}}\right) \mathbf{z} + \mathbf{n}_2 \\
&\vdots \\
\mathbf{y}_{(\mathrm{S})} &= \sqrt{P_{(\mathrm{S})}} \left(\mathbf{F}_{(\mathrm{S})}^{T} \mathbf{A}_{\mathrm{BS,D}}^{*} \otimes \mathbf{W}_{(\mathrm{S})}^{*} \mathbf{A}_{\mathrm{MS,D}}\right) \mathbf{z} + \mathbf{n}_{\mathrm{S}}
\end{aligned} \tag{2.19}$$

The design of the sth stage training precoders and combiners, $\mathbf{F}_{(\mathrm{s})}, \mathbf{W}_{(\mathrm{s})}$, depends on $\mathbf{y}_{(1)}, \mathbf{y}_{(2)}$, ..., $\mathbf{y}_{(s-1)}$. Recent research in [48, 49] shows that adaptive CS algorithms yield better performance than standard CS tools at low SNR, which is the typical case at mmWave systems before beamforming. Moreover, these adaptive CS ideas that rely on successive bisections provide important insights that can be used in the design of the training beamforming vectors.

In our proposed channel estimation algorithm described in Section 2.6, the training beamforming vectors are adaptively designed based on the bisection concept. In particular, the algorithm starts initially by dividing the vector $\mathbf{z}$ in (2.19) into a number of partitions, which equivalently divides the AoAs/AoDs range into a number of intervals, and design the training precoding and combining matrices of the first

stage, $\mathbf{F}_{(1)}, \mathbf{W}_{(1)}$, to sense those partitions. The received signal $\mathbf{y}_{(1)}$ is then used to determine the partition(s) that are highly likely to have non-zero element(s) which are further divided into smaller partitions in the later stages until detecting the non-zero elements, the AoAs/AoDs, with the required resolution. If the number of BS precoding vectors used in each stage of the adaptive algorithm equals K, where K is a design parameter, then the number of adaptive stages needed to detect the AoAs/AoDs with a resolution $\frac{2\pi}{N}$ is $\mathrm{S} = \log_K N$, which we assume to be integer for ease of exposition. Before delving into the details of the algorithm, we will focus in the following section on the design of a multi-resolution beamforming codebook which is essential for the proper operation of the adaptive channel estimation algorithm.

2.5 Hybrid Precoding Based Multi-Resolution Hierarchical Codebook

In this section, we present a novel hybrid analog/digital based approach for the design of a multi-resolution beamforming codebook. Besides considering the RF limitations, namely, the constant amplitude phase shifters with quantized phases, the proposed approach for constructing the beamforming vectors is general for ULAs/non-ULAs, has a very-low complexity, and outperforms the analog-only beamforming codebooks thanks to its additional digital processing layer.

The design of a multi-resolution or variant beamwidth beamforming vector codebook has been studied before in [13–15, 24]. This prior work focused on analog-only beamforming vectors, and on the physical design of the beam patterns. Unfortunately, the design of analog-only multi-resolution codebooks is subject to prac-

tical limitations in mmWave. (1) The existence of quantized phase shifters makes the design of non-overlapping beam patterns difficult, and may require an exhaustive search over a large space given the large number of antennas. (2) The design of analog-only beamforming vectors with certain beamwidths relies mostly on the beamsteering beam patterns of ULAs, and is hard to apply for non-ULAs due to the lack of intuition about their beam patterns.

To simplify explaining the codebook structure and design, we focus on the design of the BS training precoding codebook $\mathcal{F}$; a similar approach can be followed to construct the MS training codebook $\mathcal{W}$.

2.5.1 Codebook Structure

The proposed hierarchical codebook consists of S levels, $\mathcal{F}_s, s = 1, 2, ..., \mathrm{S}$. Each level contains beamforming vectors with a certain beamwidth to be used in the corresponding training stage of the adaptive mmWave channel estimation algorithm. Fig. 2.3 shows the first three levels of an example codebook with $N = 256$, and $K = 2$, and Fig. 2.4 illustrates the beam patterns of the beamforming vectors of each codebook level.

In each codebook level s, the beamforming vectors are divided into K^{s-1} subsets, with K beamforming vectors in each of them. Each subset k, of the codebook level s is associated with a unique range of the AoDs equal to $\{\frac{2\pi u}{N}\}_{u\in\mathcal{I}_{(s,k)}}$, where $\mathcal{I}_{(s,k)} = \left\{\frac{(k-1)N}{K^{s-1}}, ..., \frac{kN}{K^{s-1}}\right\}$. This AoD range is further divided into K sub-ranges, and each of the K beamforming vectors in this subset is designed so as to have an almost equal projection on the vectors $\mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_u\right)$, with u in this sub-range, and zero

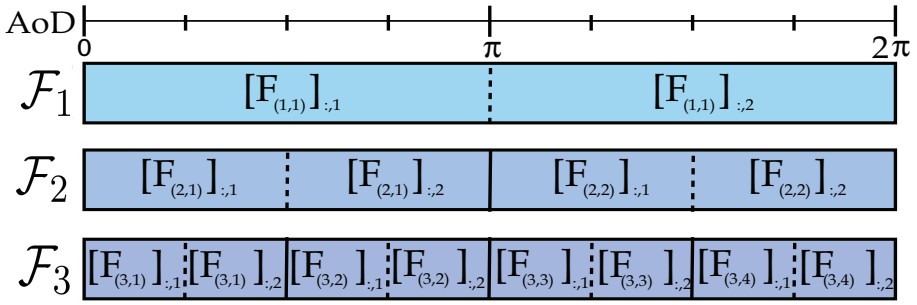


Figure 2.3: An example of the structure of a multi-resolution codebook with a resolution parameter $N = 8$, and with $K = 2$ beamforming vectors in each subset.

projection on the other vectors. Physically, this implies the implementation of a beamforming vector with a certain beamwidth determined by these sub-ranges, and steered in pre-defined directions.

While the proposed codebook structure is similar to the codebooks in [13, 24], which also have multiple levels, each with beamforming vectors with a certain beamwidth, we adopt a different way for defining each beamforming vector in terms of the set of *quantized* angles that it covers. This is different from the previous work in [13, 24] which defined each vector by the central beamforming angle and the beamwidth. This difference leads to a novel formulation of the arbitrary beamwidth beamforming design problem in addition to a completely new way for realizing these vectors using analog/digital architecture as will be explained shortly. To the best of our knowledge, this is the first work that considers implementing beamforming vectors of different beamwidth using joint analog/digital processing; as the previous work relied on analog-only designs. This additional digital processing layer adds more degrees of freedom to the beamforming design problem which can be leveraged

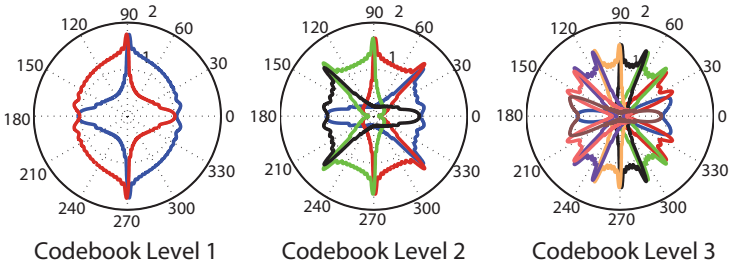


Figure 2.4: The resulting beam patterns of the beamforming vectors in the first three codebook levels of an example hierarchical codebook.

to obtain better characteristics in the beamforming patterns.

2.5.2 Design of the Codebook Beamforming Vectors

In each codebook level s, and subset k, the beamforming vectors $\left[\mathbf{F}_{(s,k)}\right]_{:,m}, m = 1, 2, ..., K$ are designed such that

$$\left[\mathbf{F}_{(s,k)}\right]^*_{:,m} \mathbf{a}_{\text{BS}}\left(\bar{\phi}_u\right) = \begin{cases} C_s & \text{if } u \in \mathcal{I}_{(s,k,m)} \\ 0 & \text{if } u \notin \mathcal{I}_{(s,k,m)} \end{cases}, \tag{2.20}$$

where $\mathcal{I}_{(k,s,m)} = \left\{\frac{N}{K^s}\left(K(k-1)+m-1\right)+1, ..., \frac{N}{K^s}\left(K(k-1)+m\right)\right\}$ defines the sub-range of AoDs associated with the beamforming vector $\left[\mathbf{F}_{(s,k)}\right]_{:,m}$, and C_s is a normalization constant that satisfies $\|\mathbf{F}_{(s,k)}\|_F = K$. For example, the beamforming vector $\left[\mathbf{F}_{(2,1)}\right]_{:,1}$ in Fig. 2.3 is designed such that it has a constant projection on the array response vectors $\mathbf{a}_{\text{BS}}\left(\bar{\phi}_u\right)$, u is in $\{0, 1, ..., 63\}$, i.e., $\bar{\phi}_u$ is in $\left\{0, ..., 2\pi\frac{63}{256}\right\}$, and zero projection on the other directions.

In a more compact form, we can write the design objective of the beamforming

vectors $\mathbf{F}_{(s,k)}$ in (2.20) as the solution of

$$\mathbf{A}^{*}_{\mathrm{BS,D}}\mathbf{F}_{(s,k)} = C_s\mathbf{G}_{(s,k)}, \tag{2.21}$$

where $\mathbf{G}_{(s,k)}$ is an $N \times K$ matrix where each column m containing $1's$ in the locations $u, u \in \mathfrak{I}_{(s,k,m)}$, and zeros in the locations $u, u \notin \mathfrak{I}_{(s,k,m)}$. Now, we note that the BS AoDs matrix $\mathbf{A}_{\mathrm{BS,D}}$ is an over-complete dictionary with $N \geq N_{\mathrm{BS}}$, i.e., (2.21) represents an inconsistent system of which the approximate solution is given by $\mathbf{F}_{(s,k)} = C_s(\mathbf{A}_{\mathrm{BS,D}}\mathbf{A}^{*}_{\mathrm{BS,D}})^{-1}\mathbf{A}_{\mathrm{BS,D}}\mathbf{G}_{(s,k)}$. Further, given the available system model in Section 2.3, the precoding matrix $\mathbf{F}_{(s,k)}$ is defined as $\mathbf{F}_{(s,k)} = \mathbf{F}_{\mathrm{RF},(s,k)}\mathbf{F}_{\mathrm{BS},(s,k)}$. As each beamforming vector will be individually used in a certain time instant, we will design each of them independently in terms of the hybrid analog/digitl precoders. Consequently, the design of the hybrid analog and digital training precoding matrices is accomplished by solving

$$\begin{aligned}
\left\{\mathbf{F}^{\star}_{\mathrm{RF},(s,k)}, \left[\mathbf{F}^{\star}_{\mathrm{BB},(s,k)}\right]_{:,m}\right\} = \\
\arg\min \quad & \| \left[\mathbf{F}_{(s,k)}\right]_{:,m} - \mathbf{F}_{\mathrm{RF},(s,k)} \left[\mathbf{F}_{\mathrm{BB},(s,k)}\right]_{:,m} \|_F, \\
\text{s.t.} \quad & \left[\mathbf{F}_{\mathrm{RF},(s,k)}\right]_{:,i} \in \left\{ [\mathbf{A}_{\mathrm{can}}]_{:,\ell} \,|\, 1 \leq \ell \leq N_{\mathrm{can}} \right\}, i = 1, 2, ..., N_{\mathrm{RF}} \\
& \|\mathbf{F}_{\mathrm{RF},(s,k)} \left[\mathbf{F}_{\mathrm{BB},(s,k)}\right]_{:,m} \|^2_F = 1,
\end{aligned} \tag{2.22}$$

where $\left[\mathbf{F}_{(s,k)}\right]_{:,m} = C_s(\mathbf{A}_{\mathrm{BS,D}}\mathbf{A}^{*}_{\mathrm{BS,D}})^{-1}\mathbf{A}_{\mathrm{BS,D}} \left[\mathbf{G}_{(s,k)}\right]_{:,m}$, and $\mathbf{A}_{\mathrm{can}}$ is an $N_{\mathrm{BS}} \times N_{\mathrm{can}}$ matrix which carries the finite set of possible analog beamforming vectors. Note that the choice of the Frobenious norm in the objective function of (2.22) is based on the point that the unconstrained vectors have semi-unitary structure and they, therefore,

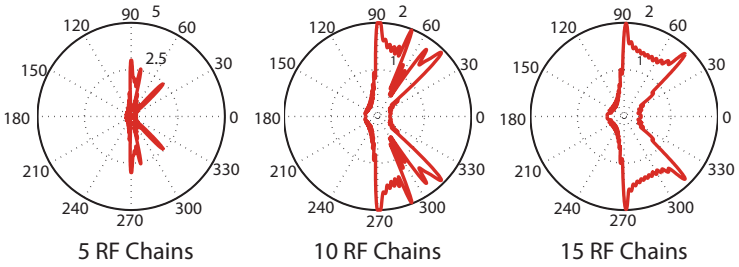


Figure 2.5: Beam patterns approximation of one of the beamforming vectors in the second codebook level with different numbers of RF chains.

represent points on the Grassman manifold. Hence, and due to the manifold's locally Euclidean property, minimizing the Euclidean distance (with the Frobenious norm) becomes a good approximation of minimizing the chordal distance between the unconstrained and hybrid precoders points on the Grassmann manifold. The columns of the candidate matrix $\mathbf{A}_{\text{can}}$ can be chosen to satisfy arbitrary analog beamforming constraints. Two example candidate beamformer designs we consider in the simulations of Section 2.8 are summarized as follows.

1. Equally spaced ULA beam steering vectors [20], i.e., a set of N_{can} vectors of the form $\mathbf{a}_{BS}(\frac{t_{\text{can}}\pi}{N})$ for $t_{\text{can}} = 0,\ 1,\ 2,\ \ldots, N_{\text{can}} - 1$.

2. Beamforming vectors whose elements can be represented as quantized phase shifts. In the case of quantized phase shifts, if each phase shifter is controlled by an N_{Q}-bit input, the entries of the candidate precoding matrix $\mathbf{A}_{\text{can}}$ can all be written as $e^{j\frac{k_{\text{Q}}2\pi}{2^{N_{\text{Q}}}}}$ for some $k_{\text{Q}} = 0,\ 1,\ 2,\ \ldots, 2^{N_{\text{Q}}} - 1$.

Now, given the matrix of possible analog beamforming vectors $\mathbf{A}_{\text{can}}$, the op-

timization problem in (2.22) can be reformulated as a sparse approximation problem similar to the optimization problem in [20, equation(17)], with the matrices $\tilde{\mathbf{F}}_{\mathrm{BB}}^{\mathrm{opt}}$, $\mathbf{F}_{\mathrm{opt}}$, $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{A}_{\mathrm{t}}$ in [20, equation(17)] taking the values $\left[\mathbf{F}_{\mathrm{BB},(s,k)}^{\star}\right]_{:,m}$, $\left[\mathbf{F}_{(s,k)}\right]_{:,m}$, $\left[\mathbf{F}_{\mathrm{BB},(s,k)}\right]_{:,m}$, and $\mathbf{A}_{\mathrm{can}}$, respectively, and with setting $N_{\mathrm{s}} = 1$. This sparse problem can be then solved using Algorithm 1 in [20] which is a variant of orthogonal matching pursuit algorithms.

An example of the beam patterns resulting from applying the proposed algorithm is shown in Fig. 2.5. These patterns are generated by a BS has 32 antennas, and a number of RF chains $N_{\mathrm{RF}} = 5, 10, 15$ to approximate one of the unconstrained beamforming vectors in the second codebook level shown in Fig. 2.4.

After the design of the BS training beamforming vectors for the kth subset of the sth codebook, the following quantities are calculated, as they will be used after that in the channel estimation algorithm in Section 2.6:

- **Beamforming Gain**: Given the channel model in (2.4), and the codebook beamforming design criteria in (2.20), we define the beamforming gain of the BS training vectors at the sth stage as $G_{(s)}^{\mathrm{BS}} = N_{\mathrm{BS}} C_s^2$. A similar definition can be used for the MS beamforming vectors, yielding a total training beamforming gain at the sth stage equal to $G^{(s)} = G_{(s)}^{\mathrm{BS}} G_{(s)}^{\mathrm{MS}}$.

- **Error Matrix**: As the system in (2.21) is inconsistent, the solution given by the pseudo-inverse means that $\mathbf{A}_{\mathrm{BS,D}}^{*}\mathbf{F}_{(s,k)}$ may not be exactly equal to $C_s\mathbf{G}_{(s,k)}$. Moreover, the limitations of the RF beamforming vectors and the approximate solution of the optimization problem in (2.22) result in an additional error

in satisfying (2.21). This error physically means (i) the existence of a spectral leakage of the beamforming vectors outside their supposed AoD sub-ranges, and (ii) the beamforming gain is not exactly uniform over the desired AoD ranges. To take the effect of this error into the performance analysis of the proposed channel estimation algorithm in Section 2.6, we define the error matrix of each subset k of the sth BS beamforming codebook level as

$$\mathbf{E}^{\mathrm{BS}}_{(s,k)} = \mathbf{A}^{*}_{\mathrm{BS,D}}\mathbf{F}_{(s,k)} - C_s\mathbf{G}_{(s,k)}. \tag{2.23}$$

As a similar error exists in the MS combining codebook, we can define the overall beamforming error matrix experienced by the received vector $\mathbf{y}_s$ in (2.18) after applying the Kronecker product as

$$\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} = {\mathbf{E}^{\mathrm{BS}}_{(s,k_{\mathrm{BS}})}}^{T}\otimes{\mathbf{E}^{\mathrm{MS}}_{(s,k_{\mathrm{MS}})}}^{T}+{\mathbf{E}^{\mathrm{BS}}_{(s,k_{\mathrm{BS}})}}^{T}\otimes C^{\mathrm{MS}}_{s}\mathbf{G}^{T}_{(s,k_{\mathrm{MS}})}+C^{\mathrm{BS}}_{s}\mathbf{G}^{T}_{(s,k_{\mathrm{BS}})}\otimes{\mathbf{E}^{\mathrm{MS}}_{(s,k_{\mathrm{MS}})}}^{T}. \tag{2.24}$$

Now, if we also define the overall gain matrix as $\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} = \mathbf{G}_{(s,k_{\mathrm{BS}})}\otimes\mathbf{G}_{(s,k_{\mathrm{MS}})}$, then we can rewrite $\mathbf{y}_s$ in (2.19) as follows assuming the subsets k_{BS} of $\mathcal{F}_s$, and k_{MS} of $\mathcal{W}_s$ are used at the BS and MS

$$\mathbf{y}_{(s)} = \sqrt{P_{(s)}}\left(\mathbf{F}^{T}_{(s,k_{\mathrm{BS}})}\otimes\mathbf{W}^{*}_{(s,k_{\mathrm{MS}})}\right)\mathbf{A}_{\mathrm{D}}\mathbf{z}+\mathbf{n}_{\mathrm{Q}} \tag{2.25}$$

$$= \sqrt{P_{(s)}}\left(\mathbf{F}^{T}_{(s,k_{\mathrm{BS}})}\mathbf{A}_{\mathrm{BS,D}}\otimes\mathbf{W}^{*}_{(s,k_{\mathrm{MS}})}\mathbf{A}_{\mathrm{MS,D}}\right)\mathbf{z}+\mathbf{n}_{\mathrm{Q}} \tag{2.26}$$

$$= \sqrt{P_{(s)}}\left(\sqrt{\frac{G^{(s)}}{N_{\mathrm{BS}}N_{\mathrm{MS}}}}\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}+\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right)\mathbf{z}+\mathbf{n}_{\mathrm{Q}}. \tag{2.27}$$

- **Actual Beamforming Gain**: To include the effect of the previously defined error matrix on the beamforming gain, we will define the actual forward and backward gains of the designed beamforming vectors. First, note that $\left(\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}+\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right)$ is a $K^2\times N^2$ matrix in which each row corresponds to a certain pair of the precoding/measurement vectors, and each column corresponds to a certain quantized AoA/AoD pair. We denote each AoA/AoD pair as a direction d, $d=1,2,...,N^2$, and the set of directions covered by each pair m of the precoder/measurement vectors as $\mathcal{D}\left(m\right), m=1,2,...,K^2$. Now, if d is one of the directions covered by the joint beamforming vector m, i.e., $d\in\mathcal{D}(m)$, then the actual beamforming gain of the vector m in the direction d can be defined as

$$\hat{G}^{(s)}_{d}=N_{\mathrm{BS}}N_{\mathrm{MS}}\left|\left[\sqrt{\frac{G^{(s)}}{N_{\mathrm{BS}}N_{\mathrm{MS}}}}\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}+\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{m,d}\right|^2, \tag{2.28}$$

$$\overset{(a)}{=}\left|\sqrt{G_{(s)}}+\sqrt{N_{\mathrm{BS}}N_{\mathrm{MS}}}\left[\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{m(d),d}\right|^2, \tag{2.29}$$

where (a) is by noting that $\left[\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{m,d}=1$ if $d\in\mathcal{D}(m)$. We can also define the side lobe gain introduced by each one of the other precoding/measurement vectors, $\bar{m}$, in the direction d as $\Delta^{(s)}_{\bar{m},d}=N_{\mathrm{BS}}N_{\mathrm{MS}}\left|\left[\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{\bar{m},d}\right|^2$, $\bar{m}=1,2,...,K^2$, $\bar{m}\neq m$. To simplify the notation, we omitted the beamforming subset symbols $k_{\mathrm{BS}},k_{\mathrm{MS}}$ from the previous definition; as each direction d is covered by only one beamforming/combining vector on each codebook level.

In the analysis of the proposed channel estimation algorithm in Section 2.6, we will be interested in the ratio between the previously defined gains. Hence, we

define the ratio between the actual forward gain in a certain direction d and the side lobe gain, in the same direction, induced by beamforming vector $\bar{m}$ as

$$\beta_{d,\bar{m}}^{(s)} = \frac{\hat{G}_d^{(s)}}{\Delta_{d,\bar{m}}^{(s)}}. \tag{2.30}$$

Note that one disadvantage of the proposed approach for constructing the beamforming vectors is the shown ripples in the main lobes in Fig. 2.4. This comes from the approximate solution of the inconsistent system in (2.21), and from the fact that we design over a finite set of directions in $\mathbf{A}_{\mathrm{BS,D}}$. These patterns, however, are acceptable for the *sparse* channel estimation problem that we consider. The main reason is that this ripple is in the main lobe, while the side lobes of these patterns are very small except for the end-fire points, i.e., the patterns containing the 0 and 180 angles for the half-wavelength ULA adopted in Fig. 2.4. If the channel has only one path in a certain direction d, then it will be affected by only one sample of this main lobe, which is in the direction d. Hence, this ripple in the main lobe just affects the forward beamforming gain. As we will show in the analysis of the proposed adaptive channel estimation algorithm in Section 2.6, the performance of the proposed algorithm depends mainly on the ratio of the backward to forward gain which is very small in the designed patterns except for small set of directions around the end-fire angles. Further, these patterns have very small overlapping which also decreases the probability of error in estimating the right AoA/AoD range as will be shown in Section 2.6.

2.6 Adaptive Estimation Algorithms for MmWave Channels

In this section, we consider the sparse channel estimation problem formulated in (2.19) of Section 2.4, and propose algorithms that adaptively use the hierarchical codebook developed in Section 2.5 to estimate the mmWave channel. We firstly address this problem for the rank-one channel model, i.e., when the channel has only one-path, in Section 2.6.1. We then extend the proposed algorithm for the multi-path case in Section 2.6.2.

2.6.1 Adaptive Channel Estimation Algorithm for Single-Path Channels

Given the problem formulation in (2.19), the single-path channel implies that the vector $\mathbf{z}$ has only one non-zero element. Hence, estimating the single-path channel is accomplished by determining the location of this non-zero element, which in turn defines the AoA/AoD, and the value of this element, which decides the channel path gain. To efficiently do that with low training overhead, we propose Algorithm 1 which adaptively searches for the non-zero element of $\mathbf{z}$ by using the multi-resolution beamforming vectors designed in Section 2.5.

Algorithm 1 operates as follows. In the initial stage, the BS uses the K training precoding vectors of the first level of the codebook $\mathcal{F}$ in Section 2.5. For each of those vectors, the MS uses the K measurement vectors of the first level of $\mathcal{W}$ to combine the received signal. Note that the first level of the hierarchical codebook in Section 2.5 has only one subset of beamforming vectors. After the K^2 precoding-measurement steps of this stage, the MS compares the power of the K^2 received signals to determine the one with the maximum received power. As each one of the precoding/measurement

Algorithm 1 Adaptive Estimation Algorithm for Single-Path MmWave Channels

Input: BS and MS know N, K, and have $\mathcal{F}, \mathcal{W}$.

Initialization: $k_1^{\text{BS}} = 1, k_1^{\text{MS}} = 1$ // Initialize the subsets to be used of codebooks $\mathcal{F}, \mathcal{W}$

$\text{S} = \log_K N$ // The number of adaptive stages

for $s \leq \text{S}$ **do**

  for $m_{\text{BS}} \leq K$ **do**

    BS transmits a training symbol using $\left[\mathbf{F}_{(s,k_s^{\text{BS}})}\right]_{:,m_{\text{BS}}}$

    for $m_{\text{MS}} \leq K$ **do**

      MS makes a measurement using $\left[\mathbf{W}_{(s,k_s^{\text{MS}})}\right]_{:,m_{\text{MS}}}$

    After MS measurements: $\mathbf{y}_{m_{\text{BS}}} = \sqrt{P_s}\left[\mathbf{W}_{(s,k_s^{\text{MS}})}\right]\mathbf{H}\left[\mathbf{F}_{(s,k_s^{\text{BS}})}\right]_{:,m_{\text{BS}}} + \mathbf{n}_{m_{\text{BS}}}$

  $\mathbf{Y}_{(s)} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_K]$

  $(m_{\text{BS}}^\star, m_{\text{MS}}^\star) = \arg\max_{\forall m_{\text{BS}}, m_{\text{MS}}=1,2,...,K}\left[\mathbf{Y}_{(s)} \odot \mathbf{Y}_{(s)}^*\right]_{m_{\text{MS}},m_{\text{BS}}}$

  $k_{s+1}^{\text{BS}} = K(m_{\text{BS}}^\star - 1) + 1, k_{s+1}^{\text{MS}} = K(m_{\text{MS}}^\star - 1) + 1$

$\hat{\phi} = \bar{\phi}_{k_{\text{S}+1}^{\text{BS}}}, \hat{\theta} = \bar{\theta}_{k_{\text{S}+1}^{\text{MS}}}$

$\hat{\alpha} = \sqrt{\frac{\rho}{P_{(\text{S})}G_{(\text{S})}}}\left[\mathbf{Y}_{(\text{S})}\right]_{m_{\text{MS}}^\star, m_{\text{BS}}^\star}$

vectors is associated with a certain range of the quantized AoA/AoD, the operation of the first stage divides the vector $\mathbf{z}$ in (2.19) into K^2 partitions, and compares between the power of the sum of each of them. Hence, the selection of the maximum power received signal implies the selection of the partition of $\mathbf{z}$, and consequently the range of the quantized AoA/AoD, that is highly likely to contain the single path of the channel. The output of the maximum power problem is then used to determine the subsets of the beamforming vectors of level $s+1$ of $\mathcal{F}$, and $\mathcal{W}$ to be used in the next stage. The MS then feeds back the selected subset of the BS precoders to the BS to use it in the next stage, which needs only $\log_2 K$ bits. As the beamforming vectors of the next levels have higher and higher resolution, the AoA/AoD ranges are further refined adaptively as we proceed in the algorithm stages until the desired resolution,

$\frac{2\pi}{N}$, is achieved. Note that the training powers in the S stages are generally different as will be discussed shortly.

Based on the proposed algorithm, the total number of stages required to estimate the AoA/AoD with a resolution $\frac{2\pi}{N}$ is $\log_K N$. Also, since we need K beamforming vectors, and K measurement vectors for each beamforming vector in each stage, the total number of steps needed to estimate the mmWave channel using the proposed algorithm becomes $K^2 \log_K N$ steps. Moreover, since N_{RF} RF chains can be simultaneously used at the MS to combine the measurements, the required number of steps can be further reduced to be $K\lceil \frac{K}{N_{\text{RF}}} \rceil \log_K N$.

In the following theorem, we characterize the performance of the proposed algorithm for the case of single dominant path channels, i.e., assuming that the channel model in (2.4) has $L = 1$. We find an upper bound of the probability of error in estimating the AoA/AoD with a certain resolution using Algorithm 1. We will then use Theorem 1 to derive sufficient conditions on the total training power and its distribution over the adaptive stages of Algorithm 1 to guarantee estimating the AoA/AoD of the mmWave channel with a desired resolution, and a certain bound on the maximum error probability.

Theorem 1 *Algorithm 1 succeeds in estimating the correct AoA and AoD of the single-path channel model in* (2.4)*, for a desired resolution* $\frac{2\pi}{N}$*, with an average prob-*

ability of error $\bar{p}$ which is upper bounded by

$$\bar{p} \leq \frac{K^2-1}{2N^2} \sum_{s=1}^{\mathrm{S}} \sum_{d=1}^{N^2} \left(1 - \frac{\left(1 - \frac{1}{\overline{\beta}_d^{(s)}}\right) P_{(s)} \hat{G}_d^{(s)} \bar{\gamma}}{4\sqrt{1 + \frac{1}{2}\left(1 + \frac{1}{\overline{\beta}_d^{(s)}}\right) P_{(s)} \hat{G}_d^{(s)} \bar{\gamma} + \frac{1}{16} P_{(s)}^2 \hat{G}_d^{(s)^2} \bar{\gamma}^2 \left(1 - \frac{1}{\overline{\beta}_d^{(s)}}\right)^2}} \right) \tag{2.31}$$

where $\overline{\beta}_d^{(s)} = \min_{\substack{\forall \bar{m}=1,2,\ldots,K^2, \\ d \notin \mathcal{D}(\bar{m})}} \beta_{d,\bar{m}}^{(s)}$ is the minimum forward to backward ratio in the direction d, $\hat{G}_d^{(s)}$ is the actual beamforming gain in this direction, and $\bar{\gamma}$ is the average channel SNR defined as $\bar{\gamma} = \frac{\bar{P}_R}{\rho \sigma^2}$.

Proof: See [27]. □

While the central idea of the proof of Theorem 1 is similar to the beam misalignment analysis in [24], this theorem has a number of important contributions over the prior work. One key difference is that Theorem 1 considers the case when the channel path gains are taken from a fading distribution; the analysis in [24] assumes constant LOS channels. Both the LOS and NLOS cases are important for mmWave systems, and using blockage models, e.g., the model in [50], we can combine the probability of estimation error in the two cases to get more accurate performance evaluation in mmWave systems. Theorem 1 also characterizes the AoA/AoD estimation error as a function of the different parameters of the designed hierarchical codebook in Section 2.5: the forward to backward gain ratios and the actual beamforming gains with imperfect realization errors. Hence, it provides a realistic evaluation of the proposed adaptive channel estimation algorithm with more practical beamforming patterns. Further, while the beam alignment analysis in [24] focused on the exhaustive search

case, i.e., when only the highest resolution beams are used, Theorem 1 considers the case when the AoAs/AoDs are estimated using adaptive algorithms. This is of particular interest in this channel estimation work, as the bound in Theorem 1 hints that the training power can be distributed among the different adaptive stages to reduce the probability of estimation error. These insights into the power allocation are explained in the remainder of this section.

Now, for the case when $\beta_d^{(s)} \to \infty$, i.e., when the backward gain in negligible and $\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \to \mathbf{0}$, we can proceed further, and obtain a sufficient condition on the training power distribution to guarantee estimating the AoA/AoD of the channel with a certain bound on the maximum probability of error.

Corollary 2 *Consider using Algorithm 1 to estimate the AoA and AoD of the single-path mmWave channel of model* (2.4)*, with a resolution* $\frac{2\pi}{N}$*, with* K *precoding and measurement vectors of* $\mathcal{F}, \mathcal{W}$ *used at each stage, and with* $\beta_d^{(s)} \to \infty$*, and* $\mathbf{E}^{(s)}_{k_{\mathrm{BS}},\mathrm{MS}} \to 0$*. If the power at each stage* $P_{(s)}, s = 1, 2, ..., \mathrm{S}$ *satisfies:*

$$P_{(s)} \geq \frac{\Gamma}{G_{(s)}} \tag{2.32}$$

with

$$\Gamma = \frac{2}{\bar{\gamma}} \left(\frac{(K^2 - 1)\mathrm{S}}{\delta} - 2 \right), \tag{2.33}$$

then, the AoA and AoD are guaranteed to be estimated with an average probability of error $\bar{p} \leq \delta$*.*

To prove Corollary 2, it is sufficient to substitute with the given $P_{(s)}$, and Γ in (2.31) to get $\bar{p} \leq \delta$.

We can also note that the power allocation strategy described in Corollary 2 makes the probability of AoA/AoD estimation error equal for the different stages. The intuition behind this result is that the stages with narrower, i.e., higher resolution, beamforming vectors have higher beamforming gains, and hence need less training power to achieve the same estimation success probability compared with the staged with wider beamforming vectors. Another advantage of Corollary 2 is that it characterizes an upper bound on the required training power to achieve a certain success probability. From Corollary 2, it is easy to show that a total training power P_{T}, with $P_{\mathrm{T}} \geq K^2 \Gamma \sum_{s=1}^{\mathrm{S}} \frac{1}{G_{(s)}}$ is sufficient to estimate the AoA/AoD of the single-path mmWave channel with $\bar{p} \leq \delta$ if it is distributed according to the way described in Corollary 2.

Finally, if we have a bound on the total training power, we can use Theorem 1 to get an upper bound on the error probability.

Corollary 3 *Consider using Algorithm 1 to estimate the AoA and AoD of the single-path mmWave channel of model* (2.4)*, with a resolution* $\frac{2\pi}{N}$*, with* K *precoding and measurement vectors of* $\mathcal{F}, \mathcal{W}$ *used at each stage, and with* $\beta_d^{(s)} \to \infty$*, and* $\mathbf{E}_{k_{\mathrm{BS}},\mathrm{MS}}^{(s)} \to 0$*. If the total training power is* P_{T}*, and if this power is distributed over the adaptive stages of Algorithm 1 such that:*

$$P_{(s)} = \frac{P_{\mathrm{T}}}{K^2 \sum_{n=1}^{\mathrm{S}} \frac{G_{(s)}}{G_{(n)}}}, s = 1, 2, ..., \mathrm{S} \tag{2.34}$$

Then, the AoA and AoD are guaranteed to be estimated with an average prob-

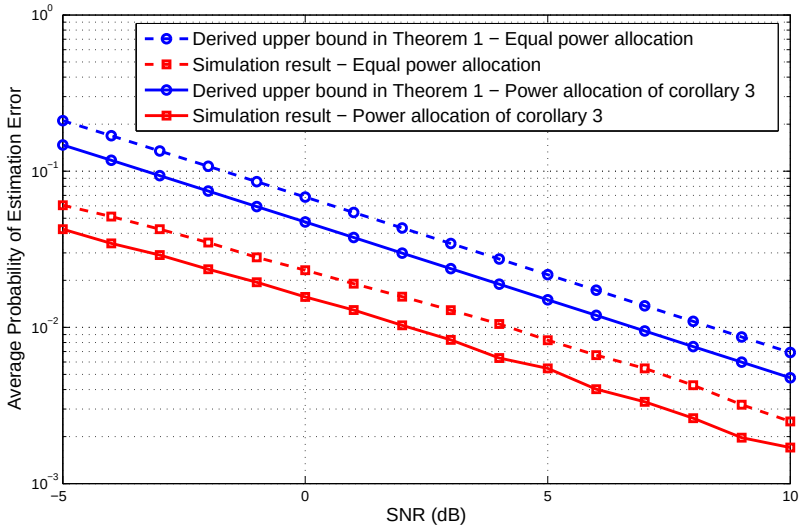


Figure 2.6: Average probability of error in estimating the AoA/AoD of single-path channels using Algorithm 1.

ability of error $\bar{p}$ *where*

$$\bar{p} \leq \frac{(K^2-1)S}{\frac{P_{\mathrm{T}}\bar{\gamma}}{2K^2\sum_{s=1}^{\mathrm{S}}\frac{1}{G_{(s)}}}+2}. \tag{2.35}$$

To prove Corollary 3, it is sufficient to substitute with the given $P_{(s)}$ in (2.31) to get the bound on $\bar{p}$. Corollary 3 is important as it gives an indication of the reliability of the AoA/AoD estimation when a certain training power is used.

To study the accurateness of the derived bound in Theorem 1 and the power allocation strategy in Corollary 3, Fig. 2.6 compares the average probability of the AoA/AoD estimation error when Algorithm 1 is used with the bound in Theorem 1. In this figure, we assume a BS with 64 antennas and 10 RF chains is communicating

with a MS that has 32 antennas and 6 RF chains. The actual average estimation error and the bound in Theorem 1 are plotted for the case when an equal power allocation over the adaptive stages is used and when the power allocation strategy in Corollary 3 is adopted. The results in Fig. 2.6 indicate that although the derived bound is not very tight, it gives very useful insights on how the training power can be distributed among the adaptive stages of the proposed channel estimation algorithm to reduce the average probability of estimation error.

In this section, the main idea of the proposed adaptive mmWave channel estimation algorithm was explained and analyzed for the single-path channels. Now, we extend this algorithm to the general case of multi-path mmWave channels.

2.6.2 Adaptive Channel Estimation Algorithm for Multi-Path Channels

Consider the case when multiple paths exist between the BS and MS. Thanks to the poor scattering nature of the mmWave channels, the channel estimation problem can be formulated as a sparse compressed sensing problem as discussed in Section 2.4. Consequently, a modified matching pursuit algorithm can be used to estimate the AoAs and AoDs along with the corresponding path gains of L_{d} paths of the channel, where L_{d} is the number of dominant paths need to be resolved. Given the problem formulation in (2.19), the objective now is to determine the L_{d} non-zero elements of $\mathbf{z}$ with the maximum power. Based on the single-path case, we propose Algorithm 2 to adaptively estimate the different channel parameters.

Modified Hierarchical Codebook: For the multi-path case, we need to make a small modification to the structure of the hierarchical codebook described

Algorithm 2 Adaptive Estimation Algorithm for Multi-Path MmWave Channels

Input: BS and MS know N, K, L_d, and have $\mathcal{F}, \mathcal{W}$

Initialization: $\mathbf{T}^{\text{BS}}_{(1,1)} = \{1, ..., 1\}, \mathbf{T}^{\text{MS}}_{(1,1)} = \{1, ..., 1\}, \text{S} = \log_K (N/L_d)$

for $\ell \leq L_d$ **do**

  for $s \leq \text{S}$ **do**

    for $m_{\text{BS}} \leq KL_d$ **do**

      BS transmits a training symbol using $\left[\mathbf{F}_{(s,\mathbf{T}^{\text{BS}}_{(\ell,s)})}\right]_{:,m_{\text{BS}}}$

      for $m_{\text{MS}} \leq KL_d$ **do**

        MS makes a measurement using $\left[\mathbf{W}_{(s,\mathbf{T}^{\text{BS}}_{(\ell,s)})}\right]_{:,m_{\text{MS}}}$

      After MS measurements: $\mathbf{y}_{m_{\text{BS}}} = \sqrt{P_s}\left[\mathbf{W}_{(s,\mathbf{T}^{\text{MS}}_{(\ell,s)})}\right]\mathbf{H}\left[\mathbf{F}_{(s,\mathbf{T}^{\text{BS}}_{(\ell,s)})}\right]_{:,m_{\text{BS}}} + \mathbf{n}_{m_{\text{BS}}}$

    $\mathbf{y}_{(s)} = [\mathbf{y}_1^T, \mathbf{y}_2^T, ..., \mathbf{y}_K^T]^T$

    for $p = 1 \leq \ell - 1$ **do** Project out the contributions of the previously estimated paths

      $\mathbf{g} = \mathbf{F}^T_{(s,\mathbf{T}^{\text{BS}}_{(p,s)})}\left[\mathbf{A}_{\text{BS,D}}\right]^*_{:,\mathbf{T}^{\text{BS}}_{(p,s)}(1)} \otimes \mathbf{W}^*_{(s,\mathbf{T}^{\text{MS}}_{(p,s)})}\left[\mathbf{A}_{\text{MS,D}}\right]_{:,\mathbf{T}^{\text{MS}}_{(p,s)}(1)}$

      $\mathbf{y}_{(s)} = \mathbf{y}_{(s)} - \mathbf{y}^*_{(s)}\mathbf{g}\left(\mathbf{g}^*\mathbf{g}\right)\mathbf{g}$

    $\mathbf{Y} = \text{matix}(\mathbf{y}_{(s)})$ Return $\mathbf{y}_{(s)}$ to the matrix form

    $(m^\star_{\text{BS}}, m^\star_{\text{MS}}) = \arg\max_{\forall m_{\text{BS}}, m_{\text{MS}}=1,2,...,K}\left[\mathbf{Y} \odot \mathbf{Y}^*\right]_{m_{\text{MS}},m_{\text{BS}}}$

    $\mathbf{T}^{\text{BS}}_{(\ell,s+1)}(1) = K(m^\star_{\text{BS}} - 1) + 1, \mathbf{T}^{\text{MS}}_{(\ell,s+1)}(1) = K(m^\star_{\text{MS}} - 1) + 1$

    for $p = 1 \leq \ell - 1$ **do**

      $\mathbf{T}^{\text{BS}}_{(\ell,s+1)}(p) = \mathbf{T}^{\text{BS}}_{(p,s+1)}(1), \mathbf{T}^{\text{MS}}_{(\ell,s+1)}(p) = \mathbf{T}^{\text{MS}}_{(p,s+1)}(1)$

  $\hat{\phi}_\ell = \bar{\phi}_{\mathbf{T}^{\text{BS}}_{(\ell,\text{S}+1)}(1)}, \hat{\theta}_\ell = \bar{\theta}_{\mathbf{T}^{\text{MS}}_{(\ell,\text{S}+1)}(1)}$

  $\mathbf{g} = \mathbf{F}^T_{(\text{S},\mathbf{T}^{\text{BS}}_{(\ell,\text{S})})}\left[\mathbf{A}^*_{\text{BS,D}}\right]_{:,\mathbf{T}^{\text{BS}}_{(\ell,\text{S}+1)}(1)} \otimes \mathbf{W}^*_{(\text{S},\mathbf{T}^{\text{MS}}_{(\ell,\text{S})})}\left[\mathbf{A}_{\text{MS,D}}\right]_{:,\mathbf{T}^{\text{MS}}_{(\ell,\text{S}+1)}(1)}$

  $\hat{\alpha}_\ell = \sqrt{\frac{\rho}{P_{(\text{S})}G_{(\text{S})}}}\frac{\mathbf{y}^*_{(\text{S})}\mathbf{g}}{\mathbf{g}^*\mathbf{g}}$

in Section 2.5. As will be explained shortly, the adaptive algorithm in the multi-path case starts by using KL_d precoding and measurement vectors at the BS and MS instead of K. In each stage, L_d of those KL_d partitions are selected for further refinement by dividing each one into K smaller partitions in the next stage. Hence, to take this into account, the first level of the codebook $\mathcal{F}$ in Section 2.5 consists

of one subset with KL_d beamforming vectors that divide the initial AoD range into KL_d ranges. Similarly, in each level $s, s > 1$, the codebook $\mathcal{F}_s$ has $K^{s-1}L_d$ levels, and the ranges $\mathcal{I}_{(s,k)}$, and $\mathcal{I}_{(s,k,m)}$ are consequently defined as $\mathcal{I}_{(s,k)} = \left\{\frac{(k-1)N}{L_d K^{s-1}}, ..., \frac{kN}{L_d K^{s-1}}\right\}$, and $\mathcal{I}_{(k,s,m)} = \left\{\frac{N}{L_d K^s}\left(K(k-1)+m-1\right)+1, ..., \frac{N}{L_d K^s}\left(K(k-1)+m\right)\right\}$. Given these definitions of the quantized AoD ranges associated with each beamforming vector m, of the subset k, of level s, the design of the beamforming vectors proceeds identical to that described in Section 2.5.2.

To estimate the L_d dominant paths of the mmWave channel, Algorithm 2 makes L_d outer iterations. In each one, an algorithm similar to Algorithm 1 is executed to detect one more path after subtracting the contributions of the previously estimated paths. More specifically, Algorithm 2 operates as follows: In the initial stage, both the BS and MS use KL_d beamforming vectors defined by the codebooks in Section 2.5 to divide the AoA, and AoD range into KL_d sub-ranges each. Similar to the single-path case, the algorithm proceeds by selecting the maximum received signal power to determine the L_d most promising sections to carry the dominant paths of the channel. This process is repeated until we reach the required AoD resolution, and only one path is estimated at this iteration. The trajectories used by the BS to detect the first path is stored in the matrix $\mathbf{T}^{\text{BS}}$ to be used in the later iterations. In the next iteration, a similar BS-MS precoding/measurement step is repeated. However, at each stage s, the contribution of the first path that has been already estimated in the previous iteration, which is stored in $\mathbf{T}^{\text{BS}}$, is projected out before determining the new promising AoD ranges. In the next stage $s+1$, two AoD ranges are selected for further refinement, namely, the one selected at stage s of this iteration, and the

one selected by the first path at stage $s+1$ of the first iteration which is stored in $\mathbf{T}^{\mathrm{BS}}$. The selection of those two AoD ranges enables the algorithm to detect different path with AoDs separated by a resolution up to $\frac{2\pi}{N}$. The algorithm proceeds in the same way until the L_d paths are solved. After estimating the AoAs/AoDs with the desired resolution, the algorithm finally calculates the estimated path gains using a linear least squares estimator (LLSE).

Note that one disadvantage of the adaptive beamwidth algorithm in the multi-path case is the possible destructive interference between the path gains when they are summed up in the earlier stages of the algorithm. This disadvantage does not appear in the exhaustive search training algorithms; as only high resolution beams are used in estimating the dominant paths of the channel. The impact of this advantage on the operation of the proposed algorithm, however, is smaller in the case of mmWave channels thanks to the sparse nature of the channel.

The total number of adaptive stages required by Algorithm 2 to estimate the AoAs/AoDs of the L_{d} paths of the channel with a resolution $\frac{2\pi}{N}$ is $\log_K\left(\frac{N}{L_{\mathrm{d}}}\right)$. Since we need KL_{d} precoding vectors, and KL_{d} measurement vectors for each precoding direction in each stage, and since these adaptive stages are repeated for each path, the total number of steps required to estimate L_d paths of the mmWave channel using the proposed algorithm is $K^2L_{\mathrm{d}}^3\log_K\left(\frac{N}{L_{\mathrm{d}}}\right)$. If multiple RF chains are used in the MS to combine the measurements, the required number of training time slots is then reduced to be $KL_d^2\lceil\frac{KL_d}{N_{\mathrm{RF}}}\rceil\log_K\left(\frac{N}{L_{\mathrm{d}}}\right)$.

2.7 Hybrid Precoding Design

We seek now to design the hybrid precoders/combiners, ($\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$, $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{BB}}$), at both the BS and MS to maximize the mutual information achieved with Gaussian signaling over the mmWave link in (2.3) [51] while taking the different RF precoding constraints into consideration. Regardless of whether uplink or downlink transmission is considered, the hybrid precoding problem can be summarized as directly maximizing the rate expression

$$R = \log_2 \left| I_{N_{\mathrm{S}}} + \frac{P}{N_{\mathrm{S}}} \mathbf{R}_{\mathrm{n}}^{-1} \mathbf{W}_{\mathrm{BB}}{}^{*} \mathbf{W}_{\mathrm{RF}}{}^{*} \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \mathbf{F}_{\mathrm{BB}}{}^{*} \mathbf{F}_{\mathrm{RF}}{}^{*} \mathbf{H}^{*} \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{BB}} \right|, \qquad (2.36)$$

over the choice of feasible analog and digital processing matrices ($\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$, $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{BB}}$). Note that in (2.36), $\mathbf{R}_{\mathrm{n}}$ is the post-processing noise covariance matrix, i.e., $\mathbf{R}_{\mathrm{n}} = \mathbf{W}_{\mathrm{BB}}{}^{*} \mathbf{W}_{\mathrm{RF}}{}^{*} \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{BB}}$ in the downlink, and $\mathbf{R}_{\mathrm{n}} = \mathbf{F}_{\mathrm{BB}}{}^{*} \mathbf{F}_{\mathrm{RF}}{}^{*} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}$ in the uplink.

For simplicity of exposition, we summarize the process with which the BS calculates the hybrid precoding matrices, ($\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$), to be used on the downlink. Calculation of the uplink precoders used by the MS follows in an identical manner.

We propose to split the precoding problem into two phases. In the first phase, the BS and MS apply the adaptive channel estimation algorithm of Section 2.4 to estimate the mmWave channel parameters. At the end of the channel training/estimation phase, the BS constructs the downlink channel's matrix leveraging the geometric structure of the channel. If the channel is not reciprocal, the estimation algorithm of Section 2.6 can be used to construct the uplink channel matrix at the MS.

As a result of the downlink channel training/estimation phase in Section 2.6, the BS now has estimated knowledge of its own steering matrix $\hat{\mathbf{A}}_{\mathrm{BS}}$, the MS steering matrix $\hat{\mathbf{A}}_{\mathrm{MS}}$, and the estimated path gain vector $\hat{\boldsymbol{\alpha}}$. Thus, the BS may construct the estimated downlink channel matrix as

$$\hat{\mathbf{H}} = \hat{\mathbf{A}}_{\mathrm{MS}} \mathrm{diag}\left(\hat{\boldsymbol{\alpha}}\right) \hat{\mathbf{A}}_{\mathrm{BS}}^{*}. \tag{2.37}$$

The BS can now build its hybrid data precoders $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ to approximate the dominant singular vectors of the channel, $\hat{\mathbf{H}}$, denoted by the unconstrained precoder $\mathbf{F}_{\mathrm{opt}}$ using a similar procedure to equation (16)-(18) of [20] and using the matching pursuit based algorithm presented in Algorithm 1 in [20].

2.8 Simulation Results

In this section, we present numerical results to evaluate the performance of the proposed training codebook, adaptive channel estimation algorithm, and hybrid precoding algorithm. We firstly consider a single BS-MS link, and then show some results for the mmWave cellular channel model.

2.8.1 Performance Evaluation with Point-to-Point Channels

In these simulations, we consider the case when there is only one BS and one MS, i.e., without any interference. The system model and the simulation scenario are as follows:

System Model We adopt the hybrid analog/digital system architecture presented in Fig. 2.2. The BS has $N_{\mathrm{BS}} = 64$ antennas, and 10 RF chains, the MS has

$N_{\text{MS}} = 32$ antennas and 6 RF chains. The antenna arrays are ULAs, with spacing between antennas equal to $\lambda/2$, and the RF phase shifters are assumed to have only quantized phases. Hence, only a finite set of the RF beamforming vectors is allowed, and assumed to be beamsteering vectors, as discussed in Section 2.5.2, with 7 quantization bits.

Channel Model We consider the channel model described in (2.4), with $\bar{P}_R = 1$, and a number of paths $L = 3$. The AoAs/AoDs are assumed to take continuous values, i.e., not quantized, and are uniformly distributed in the range $[0, 2\pi]$. The system is assumed to operate at 28GHz carrier frequency, has a bandwidth of 100MHz, and the path-loss exponent equals $n_{\text{pl}} = 3$.

Simulation Scenario All the simulations in this section will present spectral efficiency results with different system, and algorithms parameters. To generate these results, the channel parameters are estimated using the algorithms presented in Section 2.6, which in turn use the hierarchical training codebooks designed in Section 2.5. After estimating its parameters, the geometrical channel is reconstructed according to (2.37), and is used in the design of the hybrid precoders and decoders according to Section 2.7. Unless otherwise mentioned, these are the parameters used for both of the two steps:

1. Channel estimation parameters: For the single-path channels, Algorithm 1 is used to estimate the channel parameters with AoA/AoD resolution parameter $N = 64$, and with $K = 2$ beamforming vectors at each stage. For the multi-path case, the parameters N, K, L_d will be defined with each simulation. The

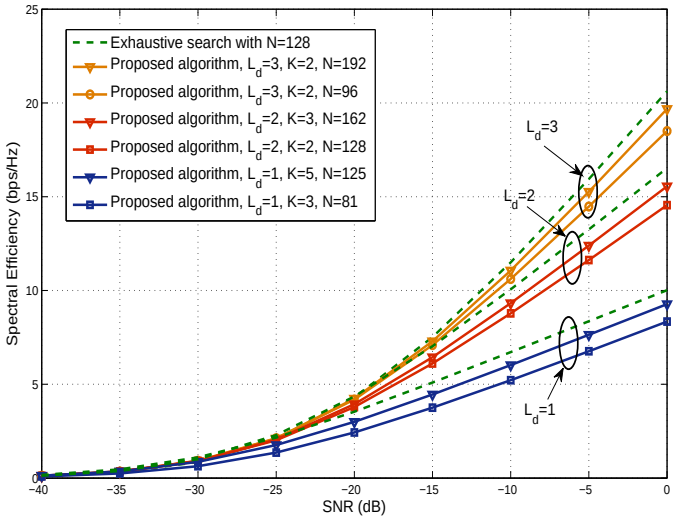


Figure 2.7: Spectral efficiency achieved when the precoding matrices are built using the mmWave channel estimated by the proposed algorithms in a channel with $L = 3$, and $L_{\mathrm{d}} = 1, 2, 3$. The figure compares the performance of the algorithm when different values of the parameter K are chosen. The results indicate that a very close performance to the exhaustive search case can be achieved with $K << N$, which maps to much smaller numbers of iterations.

training power are determined according to Corollary 2, with a desired maximum probability of error $\delta = 0.05$. Hence, the training power changes based on the parameter K, and N. Also, the total training power is distributed over the adaptive estimation stages according to Corollary 2.

2. Hybrid precoding parameters: The hybrid precoding matrices are constructed with the same available system architecture described above, and assuming a number of multiplexed streams $N_{\mathrm{S}} = L_d$.

In Fig. 2.7, the precoding gains given by the proposed mmWave channel estimation algorithms are simulated for the cases when the desired number of estimated

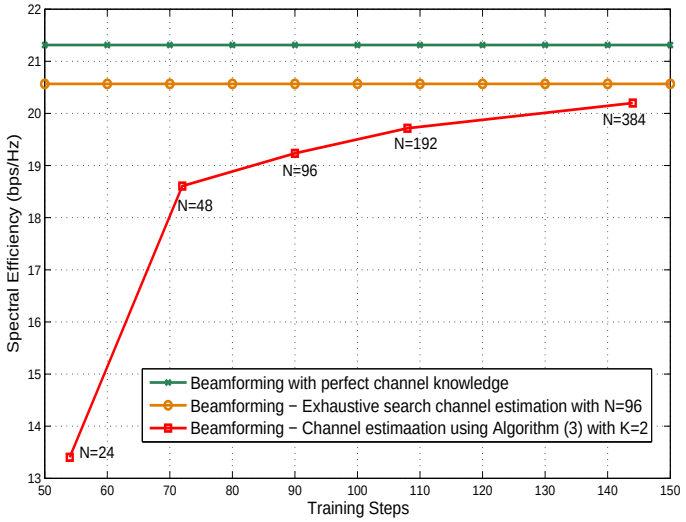


Figure 2.8: The improvement of the spectral efficiency with the development of the adaptive channel estimation algorithm is shown and compared with the exhaustive search and perfect channel knowledge cases. While the exhaustive search in this case needs a large number of iterations, a much smaller number of iterations may be sufficient to approximate its performance using the proposed adaptive algorithms.

paths L_{d} equals $1, 2$, and 3. Algorithm 1, and Algorithm 2 are simulated for different values of K, and compared with the precoding gain of the exhaustive search solution. The results indicate that comparable gains can be achieved using the proposed algorithms despite their low-complexity, and the requirement of a much smaller number of iterations. For example, for $L_{\mathrm{d}} = 3$, and $K = 2$, although only $96 \ll N_{\mathrm{BS}} N_{\mathrm{MS}} = 2048$ training steps are required, the spectral efficiency performance degradation is less than 1 bps/Hz compared with the exhaustive search solution that requires much more iterations.

In Fig. 2.8, the improvement of the precoding gains achieved by the proposed algorithm for $L_{\mathrm{d}} = 3$ with the training iterations is simulated. The results show that

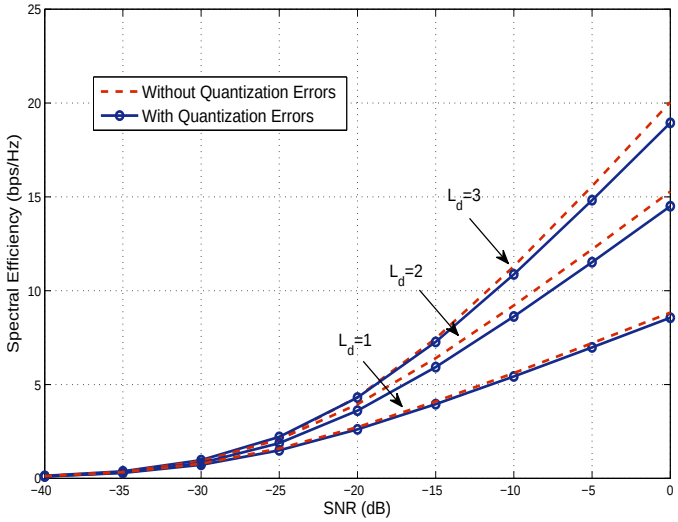


Figure 2.9: The performance error due to the AoAs/AoDs quantization assumption in (2.16) is evaluated. The performance error is the difference between the curve with continuous angles, and the one with quantization, as this continuity of angles' values is not taken into consideration while designing the algorithm.

more than 90% of the exhaustive search gain can be achieved with only 70 iterations with $K = 2$. These results also indicate that a wise choice of the desired resolution parameter N is needed in order to have a good compromise between performance and training overhead. For example, the figure shows that doubling the number of training steps, i.e., from 70 to 140, achieves an improvement of only 1 bps/Hz in the spectral efficiency.

In Fig. 2.9, we evaluate the error in the performance of the proposed channel estimation algorithm caused by the AoAs/AoDs quantization assumption made in (2.16), the proposed algorithms are simulated for the cases when the channel AoAs/AoDs are quantized, i.e., when the used quantization assumption is exact, and

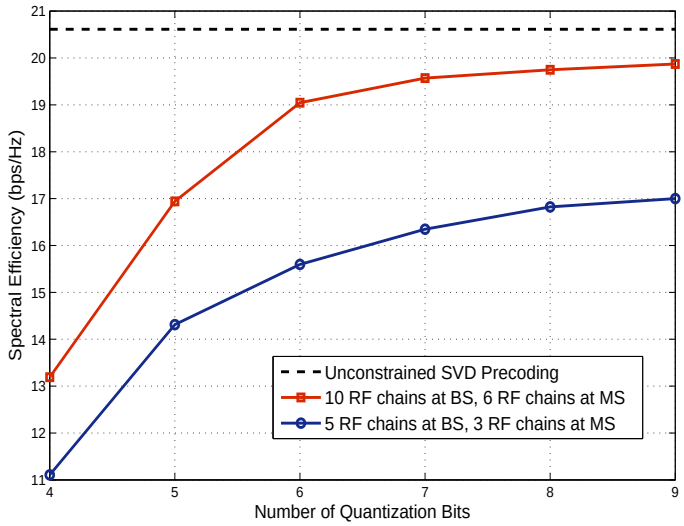


Figure 2.10: Spectral efficiency as a function of phase quantization bits in a hybrid system with only quantized analog phase control. Results compare the performance of the hybrid analog digital channel estimation and precoding algorithms with the unconstrained digital system with perfect channel knowledge at an SNR of 0dB.

when the AoAs/AoDs are continuous, i.e, with quantization error induced in our formulation. The figure plots the performance of the proposed algorithms for the cases $L_d = 1, K = 2, N = 81$,$L_d = 2,\ K = 2, N = 128$, and $L_d = 3, K = 3, N = 96$, and show that the performance loss in our algorithms due to the quantization assumption is very small for large enough resolution parameters N.

In Fig. 2.10, the impact of the RF system limitations on the performance of the proposed channel estimation, and precoding algorithms, is evaluated, and compared with the case of constraints-free system. Two system models are considered in Fig. 2.10, one with 10 RF chains at the BS, 6 RF chains at the MS, and the other with 5 RF chains at the BS, and 3 RF chains at the MS. The other parameters are

the same as the previous simulations with $L_d = 3$. The performance achieved by those two systems is further simulated with different number of quantization bits of the phase shifters. Simulation results show that the proposed hybrid analog-digital precoding algorithm can achieve near-optimal data-rates compared with the unconstrained solutions if a sufficient number of RF chains, and quantization bits exist. Also, the results show that 5 quantization bits may be sufficient to accomplish more that 90% of the maximum gain.

2.8.2 Performance Evaluation with MmWave Cellular System Setup

Now, we consider evaluating the proposed algorithm in a mmWave cellular system setting with out-of-cell interference. To provide a practical evaluation, we adopt the following stochastic geometry model.

Network and System Models The desired BS, in a cell of radius $R_c = 100m$, is assumed to communicate with a MS using the channel estimation, and hybrid precoding algorithms derived. Each MS is assumed to receive its desired signal s_{d} in addition to cellular interference. The interfering BSs follow a Poisson point process (PPP) $\Phi(\lambda)$ with $\lambda = \frac{1}{\pi R_c^2}$ to model the downlink out-of-cell interference [52–54]. To simulate a cellular setting, the nearest BS to the MS is always considered as the desired BS. The received signal at the MS can be then written as

$$\mathbf{y} = \mathbf{W}^* \mathbf{H}_{\mathrm{d}} \mathbf{F}_{\mathrm{d}} \mathbf{s}_{\mathrm{d}} + \sum_{\substack{r_i \in \Phi(\lambda) \\ r_i \geq r_{\mathrm{d}}}} \mathbf{W}^* \mathbf{H}_i \mathbf{F}_i \mathbf{s}_i + \mathbf{n} \tag{2.38}$$

where r_{d}, r_i are the distances from the MS to the desired and the ith interfering BSs, respectively. Each interfering BS is assumed to have the same number of ULA anten-

nas $N_{\mathrm{BS}} = 64$, and to have the same horizontal orientation of the antenna arrays, i.e., all the beamforming is in the azimuth domain. Further, each BS generates a beam-steering beamforming vectors that steers its signal in a uniform random direction, i.e., $\mathbf{F}_i = \mathbf{a}_{\mathrm{BS}}(\phi_i)$, ϕ_i is uniformly chosen in $[0, 2\pi]$. $\mathbf{H}_i$ has the same definition in (2.4) with the path loss calculated for each BS based on its distance r_i. For fairness, all BSs are assumed to transmit with the same average power P. All the other system parameters are similar to the previous section.

In each stage s of the estimation phase, the received signal at the MS is given by (2.38) with $\mathbf{W}$ and $\mathbf{F}$ equal to the BS and MS training precoders and combiners descried in Section 2.6. Hence, the cellular interference affects the maximum power detection problem at every stage of the channel estimation algorithm. After the channel is estimated, the precoders $\mathbf{W}$ and $\mathbf{F}$ are designed as shown in Section 2.7.

To evaluate the performance of the proposed hybrid precoding algorithm, we adopt the coverage probability as a performance metric. As we are interested in multiplexing many streams per user, we define the coverage probability relative to the rate instead of the signal to interference and noise ratio (SINR). Consequently, we use the following definition of the coverage probability

$$P_{(c)}(\eta) = \mathrm{P}(R \geq \eta). \tag{2.39}$$

An outage happens if the user's rate falls below a certain threshold η.

Scenario and Results In Fig. 2.11, the coverage probability is evaluated

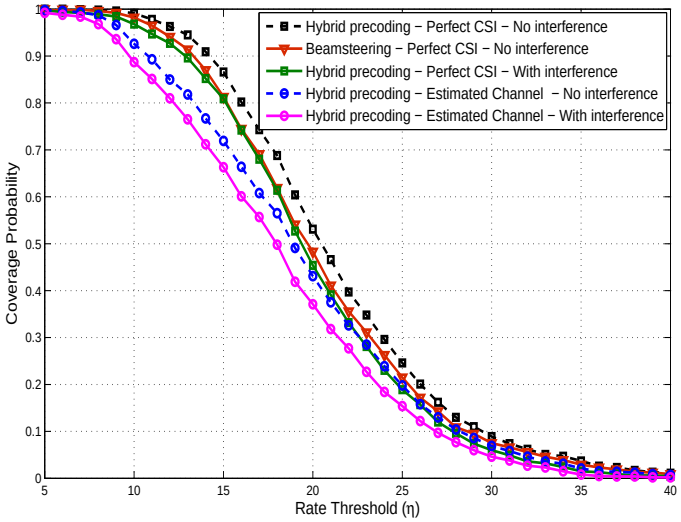


Figure 2.11: Coverage probabilities of the proposed channel estimation and precoding algorithms in a mmWave cellular system setting with PPP interference. The figure compares the different cases when the estimation and/or interference error exist to evaluate the effect of each of them on the proposed algorithms.

as described before. The curves with 'Estimated Channel' label represents the case when Algorithm (3) is used to estimate the channel parameters in the presence of interference. After estimating the channel, this interference is taken into consideration again in calculating the coverage probability in the curve labeled 'With Interference', and omitted for the curve with the label 'No Interference'. Hence, those two curves represent the cases when cellular interference affects both the channel estimation and data transmission phases, or the channel estimation phase only. The presented results compare the performance of the mentioned scenario using the proposed algorithms, with the case when the hybrid precoding algorithm in Section 2.7 is designed based on perfect channel state information (CSI). They are also compared with the case

when only analog beamforming is used to steer the signal towards the dominant channel paths. The results show that a reasonable gain can be achieved with the proposed hybrid precoding algorithm due to its higher capability of managing the inter-stream interference, in addition to overcoming the RF hardware constraints. The simulations also indicate that the effect of the cellular interference of the performance of the channel estimation and precoding algorithms is not critical despite of the low-complexity of the proposed algorithms.

2.9 Comments and Discussion

In this chapter, we clarify important comments on the proposed channel estimation algorithm and discuss its relation to standard compressed sensing framework.

2.9.1 Establishing RIP conditions:

For the adaptive channel estimation measurement matrix, we did not prove that it satisfies RIP conditions because (i) we are not solving the sparse channel estimation problem using a standard (non-adaptive) compressed sensing tools which require these conditions to guarantee support recovery, and (ii) for the adaptive structure of the measurement matrix that we have, we can not prove any RIP conditions as these conditions are by definition contradicting adaptive structures of the measurement matrices. Hereafter, we explain these points in more detail. For sparse reconstruction problems, we can identify two main research directions to solve them (i) standard (non-adaptive) compressed sensing [47, 55, 56], and (ii) adaptive compressive sensing [48, 49, 57, 58].

- In the standard compressed sensing, the measurement matrix is constructed one time a priori of the sensing. To guarantee the recovery of the sparse vector support with high success probability, conditions on the measurement matrix such as RIP and maximum coherence are useful to establish these guarantees. One important point is that these conditions on the sensing matrix should guarantee the recovery of any s-sparse vector as the measurement matrices are designed one time.

- Adaptive compressed sensing that is proposed more recently is based on the point that if the rows of the measurement matrix are designed more smartly in a recursive way given the result of the previous sensing, then we can intuitively get better recovery results with the same number of measurements. As this recursive measurement matrix structure is adaptively designed depending on the sparse vector under sensing, though, the definitions of the recovery conditions (like RIP), that should guarantee recovery of any sparse vector, do not hold. In fact, establishing theoretical support recovery guarantees for the adaptive measurement matrices is more complex than that in the standard way. Instead of the RIP and coherence conditions, adaptive CS solutions usually attempt to derive bounds on the probability of error in recovering the support. In the recent work [48], bounds on the probability of support recover error were derived for a bisection sensing algorithm for the cases when the non-zero elements are real, and was shown to have SNR gain over standard compresses sensing tools.

With this motivation of the work in [48], we designed our hybrid precoders/ combiners to work as a bisection sensing of the sparse vector $\mathbf{z}$ in (2.16). As the

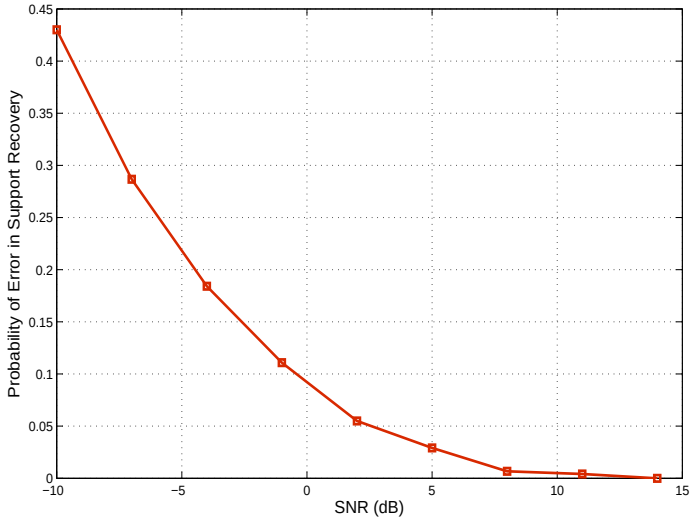


Figure 2.12: Probability of error in estimating the sparse vector support (angles of arrival and departure) with a channel that has 3 paths, and with a desired resolution $\frac{2\pi}{N}$ for $N = 48$.

non-zero elements of $\mathbf{z}$ are complex, though, deriving bounds on the probability of error is non-trivial in general. So, we derived a bound on the error probability for the single-path case, and relied on numerical simulations for evaluating the general performance of the algorithm.

2.9.2 Recovering the support with some error:

In the proposed channel estimation algorithm, we can recover the support but with some probability of error that depends on the SNR, the hardware constraints of the hybrid architecture, and the different algorithm parameters such as the number of measurements per adaptive step k, and the desired resolution N. In Fig. 2.12 and Fig. 2.13, we plot the probability of error in support error, and the normalized

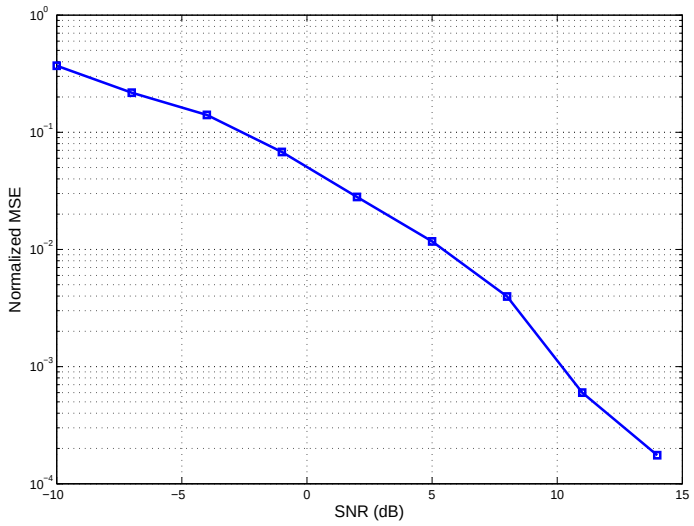


Figure 2.13: Normalized MSE of the estimated channel matrix, defined as NMSE= $\mathbb{E}\left[\frac{\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2}{\|\mathbf{H}\|_F^2}\right]$, with $\mathbf{H}$ and $\hat{\mathbf{H}}$ the original and estimated channel matrices, respectively. The channel is assumed to have 3 paths, and the desired resolution is $\frac{2\pi}{N}$ for $N = 48$.

mean-squared error of the estimated channel matrix, respectively, for different SNR values. The SNR is defined as SNR $= \frac{P}{\sigma^2}$, where P is the average training power. The results in Fig. 2.12-Fig. 2.13 assume a system that employ 64 antennas, 16 RF chains at the transmitter, and 32 antennas with 8 RF chains at the receiver. The channel is assumed to have $L = 3$ paths, and the desired estimate resolution is $N = 48$.

2.9.3 The sparsity level in the proposed sparse formulation of the mmWave channel estimation problem

In our algorithm, we assume exact sparse vectors, i.e., number of zeros equals the number of paths L, which is much less than the size N^2 of the sparse vector $\mathbf{z}$ in (2.16). In the simulation results, we do not neglect the impacts of the grid error

and sidelobe error, and hence we do not have exact sparse vectors. The exact sparse assumption becomes more suitable when the number of antennas and grid size grow large [59], which is usually the case in mmWave massive MIMO systems.

2.10 Conclusions

In this chapter, we considered a single-user mmWave system setting, and investigated the design of suitable mmWave channel estimation and precoding algorithms. First, we formulated, and developed a hierarchical multi-resolution codebook based on hybrid analog/digital precoding. We then proposed mmWave channel estimation algorithms that efficiently detect the different parameters of the mmWave channel with a low training overhead. The proposed algorithms depend on the developed sparse formulation of the poor scattering mmWave channel, and on the designed hierarchical codebooks to adaptively estimate the channel parameters. The performance of the proposed algorithm is analytically evaluated for the single-path channel case, and some insights into efficient training power distributions are obtained. Despite the low-complexity, simulation results showed that the proposed channel estimation algorithm realizes spectral efficiency and precoding gain that are comparable to that obtained by exhaustive search. The mmWave hybrid precoding algorithms are also proved to achieve a near-optimal performance relative to the unconstrained digital solutions, and attain reasonable gains compared with analog-only beamforming.

Chapter 3

Multi-User Hybrid Precoding

3.1 Overview

In this chapter[1], we develop a low-complexity hybrid analog/digital precoding algorithm for downlink multiuser mmWave systems. The proposed algorithm configures hybrid precoders at the transmitter and analog combiners at multiple receivers with a small training and feedback overhead. The performance of the proposed algorithm is analyzed in the large dimensional regime and in single path channels. When the analog and digital precoding vectors are selected from quantized codebooks, the rate loss due to the joint quantization is characterized and insights are given into the performance of hybrid precoding compared with analog-only beamforming solutions. Analytical and simulation results show that the proposed techniques offer higher sum rates compared with analog-only beamforming solutions, and approach the performance of the unconstrained digital beamforming with relatively small codebooks.

[1]This chapter is based on the work published in the journal paper: A. Alkhateeb, G. Leus, and R. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," IEEE Transactions on Wireless Communications, vol. 14, no. 11, pp. 6481-6494, Nov. 2015. This work was supervised by Prof. Robert Heath. Prof. Geert Leus provided important ideas for the multi-user hybrid precoding design that greatly improved the work.

3.2 Introduction

The large bandwidths in the mmWave spectrum make mmWave communication desirable for wireless local area networking and also a promising candidate for future cellular systems [6,7,20,21,27]. Achieving high quality communication links in mmWave systems requires employing large antenna arrays at both the access point or BS and the MSs [17, 20, 60]. For efficient system performance, each BS needs to simultaneously serve a number of MS's. Multiplexing different data streams to different users requires some form of precoding be applied to generate the transmitted signal at the BS. In conventional lower frequency systems, this precoding was commonly done in the baseband to have a better control over the entries of the precoding matrix. Unfortunately, the high cost and power consumption of mixed signal components make fully digital baseband precoding unlikely with current semiconductor technologies [12,20]. Further, the design of the precoding matrices is usually based on complete channel state information, which is difficult to achieve in mmWave systems due to (i) the large number of antennas which would require a huge training overhead and (ii) the small SNR before beamforming. Therefore, new multi-user precoding algorithms that (i) respect the mmWave hardware constraints and (ii) require much less complexity need to be developed for mmWave systems.

In single-user mmWave systems, analog beamforming, which controls the phase of the signal transmitted at each antenna via a network of analog phase shifters and is implemented in the radio frequency (RF) domain, was proposed instead of the baseband solutions [13, 14, 24, 61]. This was also adopted in commercial indoor mmWave communication standards like IEEE 802.11ad [36] and IEEE 802.15.3c [35]. In [13,24],

adaptive beamforming algorithms and multi-resolution codebooks were developed by which the transmitter and receiver jointly design their analog beamforming vectors. In [14], unique signatures are assigned to the different training beamforming vectors and used to minimize the training overhead. In [61], beamspace MIMO was introduced in which DFT beamforming vectors are used to direct the transmitted signals towards the subspaces that asymptotically maximize the received signal power with large numbers of antennas. Analog beamformers as in [13, 14, 24, 61] are subject to additional constraints, for example, the phase shifters might be digitally controlled and have only quantized phase values and adaptive gain control might not be implemented. These constraints limit the potential of analog-only beamforming solutions relative to baseband precoding, as they limit the ability to make sophisticated processing, for example to manage interference between users.

To multiplex several data streams and perform more accurate beamforming, hybrid precoding was proposed [20,27,62], where the processing is divided between the analog and digital domains. In [20], the sparse nature of the mmWave channels was exploited to develop low-complexity hybrid precoding algorithms using the algorithmic concept of basis pursuit assuming the availability of channel knowledge. In [62], low-complexity hybrid beamforming algorithms were proposed for single-user single-stream MIMO-OFDM systems with the objective of maximizing either the received signal strength or the sum-rate over different sub-carriers. In [27], a hybrid precoding algorithm that requires only partial knowledge about the mmWave channels was devised. The hybrid precoding algorithms in [20, 27, 62], though, were designed to obtain either diversity or spatial multiplexing gain from single-user channels, which can

support a limited number of streams [21]. In multi-user systems, the digital precoding layer of hybrid precoding gives more freedom in designing the precoders, compared with analog-only solutions, which can be exploited to reduce the interference between users. Hence, developing low-complexity hybrid precoding algorithms for multi-user mmWave systems is of special interest.

Pre-precoding processing has been investigated for other systems [18, 19, 23]. In [18], the joint analog-digital precoder design problem was studied for both diversity and spatial multiplexing systems. In [19], hybrid analog/digital precoding algorithms were developed to minimize the received signal's mean-squared error in the presence of interference when phase shifters with only quantized phases are available. In [23], two-layer beamforming algorithms were proposed to group the users and reduce the channel feedback overhead in massive MIMO systems. The approaches in [18, 19, 23], however, were not designed specifically for mmWave systems as they did not consider the mmWave-related hardware constraints, and did not leverage mmWave channel characteristics to realize low-complexity solutions.

In this work, we develop a low-complexity yet efficient hybrid analog/digital precoding algorithm for downlink multi-user mmWave systems. The proposed algorithm is general for arbitrary known array geometries, and assumes the availability of only a limited feedback channel between the BS and MS's. The main contributions of the work on this chapter can be summarized as follows:

- Developing a hybrid precoding/combining algorithm for multi-user mmWave systems. Our model assumes that the MS's employ analog-only combining while

the BS performs hybrid analog/digital precoding where the number of RF chains is at least as large as the number of MS's. The proposed algorithm is designed to reduce the training and feedback overhead while achieving performance close to that of unconstrained solutions.

- Analyzing the performance of the proposed algorithm in special cases: (i) when the channels are single-path, and (ii) when the number of transmit and receive antennas are very large, which are relevant for mmWave systems.

- Characterizing the average rate loss due to joint analog and digital codebook quantization, and identifying the cases at which large hybrid precoding gains exist compared with analog-only beamforming solutions.

The proposed algorithm and performance bounds are evaluated by simulations and compared with both analog-only beamforming solutions and digital unconstrained precoding schemes. The results indicate that with a relatively small feedback and training overhead, the proposed hybrid precoding algorithm achieves good performance thanks to the sparse nature of the channel and the large number of antennas used by the BS and MS's.

3.3 System Model

Consider the multi-user mmWave system shown in Fig. 3.1. A base station with N_{BS} antennas and N_{RF} RF chains is assumed to communicate with U mobile stations. Each MS is equipped with N_{MS} antennas as depicted in Fig. 3.2. We focus on the multi-user beamforming case in which the BS communicates with every MS

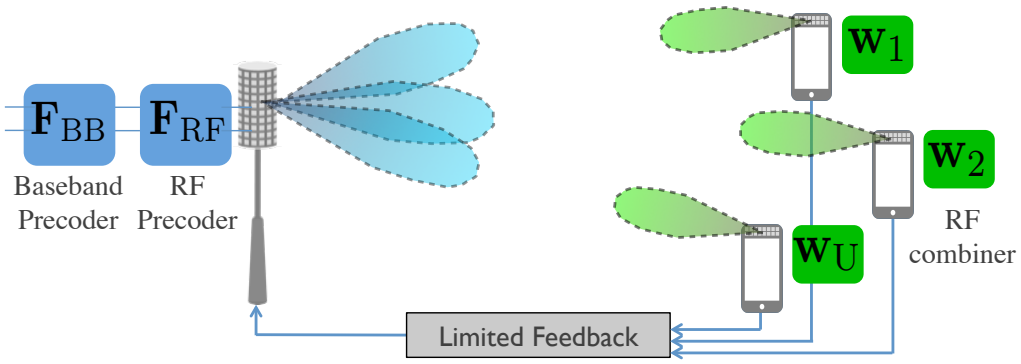


Figure 3.1: A multi-user mmWave downlink system model, in which a BS uses hybrid analog/digital precoding and a large antenna array to serve U MSs. Each MS employs analog-only combining and has a limited feedback channel to the BS.

via *only one stream*. Therefore, the total number of streams $N_{\mathrm{S}} = U$. Further, we assume that the maximum number of users that can be simultaneously served by the BS equals the number of BS RF chains, i.e., $U \leq N_{\mathrm{RF}}$. This is motivated by the spatial multiplexing gain of the described multi-user hybrid precoding system, which is limited by $\min(N_{\mathrm{RF}}, U)$ for $N_{\mathrm{BS}} > N_{\mathrm{RF}}$. For simplicity, we will also assume that the BS will use U out of the N_{RF} available RF chains to serve the U users.

On the downlink, the BS applies a $U \times U$ baseband precoder $\mathbf{F}_{\mathrm{BB}} = \left[\mathbf{f}_1^{\mathrm{BB}}, \mathbf{f}_2^{\mathrm{BB}}, ..., \mathbf{f}_U^{\mathrm{BB}}\right]$ followed by an $N_{\mathrm{BS}} \times U$ RF precoder, $\mathbf{F}_{\mathrm{RF}} = \left[\mathbf{f}_1^{\mathrm{RF}}, \mathbf{f}_2^{\mathrm{RF}}, ..., \mathbf{f}_U^{\mathrm{RF}}\right]$. The sampled transmitted signal is therefore

$$\mathbf{x} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s}, \tag{3.1}$$

where $\mathbf{s} = [s_1, s_2, ..., s_U]^{\mathrm{T}}$ is the $U \times 1$ vector of transmitted symbols, such that $\mathbb{E}[\mathbf{s}\mathbf{s}^*] = \frac{P}{U}\mathbf{I}_U$, and P is the average total transmitted power. We assume equal power allocation among different users' streams. Since $\mathbf{F}_{\mathrm{RF}}$ is implemented using analog phase shifters, its entries are of constant modulus. We normalize these entries to

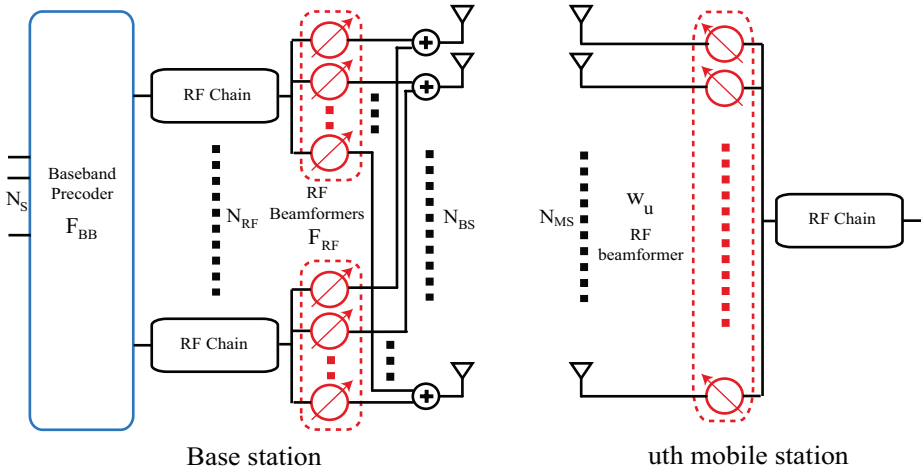


Figure 3.2: A BS with hybrid analog/digital architecture communicating with the uth MS that employs analog-only combining.

satisfy $\left|[\mathbf{F}_{\mathrm{RF}}]_{m,n}\right|^2 = N_{\mathrm{BS}}^{-1}$. Further, we assume that the angles of the analog phase shifters are quantized and have a finite set of possible values. With these assumptions, $[\mathbf{F}_{\mathrm{RF}}]_{m,n} = \frac{1}{\sqrt{N_{\mathrm{BS}}}} e^{\mathrm{j}\phi_{m,n}}$, where $\phi_{m.n}$ is a quantized angle. The angle quantization assumption is discussed in more detail in Section 3.4. The total power constraint is enforced by normalizing $\mathbf{F}_{\mathrm{BB}}$ such that $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = U$.

For simplicity, we adopt a narrowband block-fading channel model as in [20, 27, 63] in which the uth MS observes the received signal as

$$\mathbf{r}_u = \mathbf{H}_u \sum_{n=1}^{U} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_n^{\mathrm{BB}} s_n + \mathbf{n}_u, \tag{3.2}$$

where $\mathbf{H}_u$ is the $N_{\mathrm{MS}} \times N_{\mathrm{BS}}$ matrix that represents the mmWave channel between the BS and the uth MS, and $\mathbf{n}_u \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is the Gaussian noise corrupting the received signal.

At the uth MS, the RF combiner $\mathbf{w}_u$ is used to process the received signal $\mathbf{r}_u$:

$$y_u = \mathbf{w}_u^* \mathbf{H}_u \sum_{n=1}^{U} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_n^{\mathrm{BB}} s_n + \mathbf{w}_u^* \mathbf{n}_u, \tag{3.3}$$

where $\mathbf{w}_u$ has similar constraints as the RF precoders, i.e., the constant modulus and quantized angles constraints. In this work, we assume that only analog (RF) beamforming is used at the MS's as they will likely need cheaper hardware with lower power consumption.

MmWave channels are expected to have limited scattering [38, 40]. To incorporate this effect, we adopt a geometric channel model with L_u scatterers for the channel of user u. Each scatterer is assumed to contribute a single propagation path between the BS and MS [20]. The adopted geometrical channel model can also be transformed into the virtual channel model [64]. The virtual channel model simplifies the generalization for larger angle spreads by incorporating spatial spreading functions as will be briefly discussed in Section 3.6.2. Under this model, the channel $\mathbf{H}_u$ can be expressed as

$$\mathbf{H}_u = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{L_u}} \sum_{\ell=1}^{L_u} \alpha_{u,\ell} \mathbf{a}_{\mathrm{MS}}\left(\theta_{u,\ell}\right) \mathbf{a}_{\mathrm{BS}}^*\left(\phi_{u,\ell}\right), \tag{3.4}$$

where $\alpha_{u,\ell}$ is the complex gain of the ℓ^{th} path, including the path-loss, with $\mathbb{E}\left[|\alpha_{u,\ell}|^2\right] = \bar{\alpha}$. The variables $\theta_{u,\ell}$, and $\phi_{u,\ell} \in [0, 2\pi]$ are the ℓ^{th} path's AoAs/AoDs respectively. Finally, $\mathbf{a}_{\mathrm{BS}}\left(\phi_{u,\ell}\right)$ and $\mathbf{a}_{\mathrm{MS}}\left(\theta_{u,\ell}\right)$ are the antenna array response vectors of the BS and uth MS respectively. The BS and each MS are assumed to know the geometry of their antenna arrays. While the algorithms and results developed in the chapter can be applied to arbitrary antenna arrays, we use ULAs and UPAs in the simulations of

Section 3.8. If a ULA is assumed, $\mathbf{a}_{\mathrm{BS}}(\phi_{u,\ell})$ can be defined as

$$\mathbf{a}_{\mathrm{BS}}(\phi_{u,\ell}) = \frac{1}{\sqrt{N_{\mathrm{BS}}}}\left[1, e^{\mathrm{j}\frac{2\pi}{\lambda}d\sin\left(\phi_{u,\ell}\right)}, ..., e^{\mathrm{j}(N_{\mathrm{BS}}-1)\frac{2\pi}{\lambda}d\sin\left(\phi_{u,\ell}\right)}\right]^{\mathrm{T}}, \tag{3.5}$$

where λ is the signal wavelength, and d is the distance between antenna elements. The array response vectors at the MS, $\mathbf{a}_{\mathrm{MS}}(\theta_{u,\ell})$, can be written in a similar fashion.

3.4 Problem Formulation

Our main objective is to efficiently design the analog (RF) and digital (baseband) precoders at the BS and the analog combiners at the MS's to maximize the sum-rate of the system.

Given the received signal at the uth MS in (3.2) which is then processed using the RF combiner $\mathbf{w}_u$, the achievable rate of user u is

$$R_u = \log_2\left(1 + \frac{\frac{P}{U}\left|\mathbf{w}_u^*\mathbf{H}_u\mathbf{F}_{\mathrm{RF}}\mathbf{f}_u^{\mathrm{BB}}\right|^2}{\frac{P}{U}\sum_{n\neq u}\left|\mathbf{w}_u^*\mathbf{H}_u\mathbf{F}_{\mathrm{RF}}\mathbf{f}_n^{\mathrm{BB}}\right|^2 + \sigma^2}\right). \tag{3.6}$$

The sum-rate of the system is then $R_{\mathrm{sum}} = \sum_{u=1}^{U} R_u$.

Due to the constraints on the RF hardware, such as the availability of only quantized angles for the RF phase shifters, the analog beamforming/combining vectors can take only certain values. Hence, these vectors need to be selected from finite-size codebooks. There are different models for the RF beamforming codebooks, two possible examples are

1. **General quantized beamforming codebooks** Here, the codebooks are designed to satisfy some particular properties, e.g., maximizing the minimum

distance between the codebook vectors as in Grassmannian codebooks. These codebooks are usually designed for rich channels and, therefore, attempt a uniform quantization on the space of beamforming vectors. These codebooks were commonly used in traditional MIMO systems.

2. **Beamsteering codebooks** The beamforming vectors, here, are spatial matched filters for the single-path channels. As a result, they have the same form of the array response vector and can be parameterized by a simple angle. Let $\mathcal{F}$ represents the RF beamforming codebook, with cardinality $|\mathcal{F}| = N_{\mathrm{Q}}$. Then, in the case of beamsteering codebooks, $\mathcal{F}$ consists of the vectors $\mathbf{a}_{\mathrm{BS}}\left(\frac{2\pi k_{\mathrm{Q}}}{N_{\mathrm{Q}}}\right)$, for the variable k_{Q} taking the values $0, 1, 2$, and $N_{\mathrm{Q}} - 1$. The RF combining vectors codebook $\mathcal{W}$ can be similarly defined.

Motivated by the good performance of single-user hybrid precoding algorithms [20, 27] which relied on RF beamsteering vectors, and by the relatively small size of these codebooks which depend on single parameter quantization, we will adopt the beamsteering codebooks for the analog beamforming vectors. While the problem formulation and proposed algorithm in this work are general for any codebook, the performance evaluation of the proposed algorithm done in Sections 3.6-3.7 depends on the selected codebook. For future work, it is of interest to evaluate the performance of the proposed hybrid precoding algorithm with other RF beamforming codebooks.

If the system sum-rate is adopted as a performance metric, the precoding

design problem is then to find $\mathbf{F}_{\mathrm{RF}}^{\star}$, $\left\{\mathbf{f}_u^{\star\mathrm{BB}}\right\}_{u=1}^{U}$ and $\{\mathbf{w}_u^{\star}\}_{u=1}^{U}$ that solve

$$\begin{aligned}\left\{\mathbf{F}_{\mathrm{RF}}^{\star}, \left\{\mathbf{f}_u^{\star\mathrm{BB}}\right\}_{u=1}^{U}, \{\mathbf{w}_u^{\star}\}_{u=1}^{U}\right\} = \arg\max & \sum_{u=1}^{U} \log_2\left(1+\frac{\frac{P}{U}\left|\mathbf{w}_u^{*}\mathbf{H}_u\mathbf{F}_{\mathrm{RF}}\mathbf{f}_u^{\mathrm{BB}}\right|^2}{\frac{P}{U}\sum_{n\neq u}\left|\mathbf{w}_u^{*}\mathbf{H}_u\mathbf{F}_{\mathrm{RF}}\mathbf{f}_n^{\mathrm{BB}}\right|^2+\sigma^2}\right) \\ \text{s.t.} \quad & [\mathbf{F}_{\mathrm{RF}}]_{:,u} \in \mathcal{F}, u=1,2,...,U, \\ & \mathbf{w}_u \in \mathcal{W}, u=1,2,...,U, \\ & \|\mathbf{F}_{\mathrm{RF}}\left[\mathbf{f}_1^{\mathrm{BB}}, \mathbf{f}_2^{\mathrm{BB}}, ..., \mathbf{f}_U^{\mathrm{BB}}\right]\|_F^2 = U.\end{aligned} \tag{3.7}$$

The problem in (3.7) is a mixed integer programming problem. Its solution requires a search over the entire $\mathcal{F}^U \times \mathcal{W}^U$ space of all possible $\mathbf{F}_{\mathrm{RF}}$ and $\{\mathbf{w}_u\}_{u=1}^{U}$ combinations. Further, the digital precoder $\mathbf{F}_{\mathrm{BB}}$ needs to be jointly designed with the analog beamforming/combining vectors. In practice, this may require the feedback of the channel matrices $\mathbf{H}_u, u=1,2,...,U$, or the effective channels, $\mathbf{w}_u^{*}\mathbf{H}_u\mathbf{F}_{\mathrm{RF}}$. Therefore, the solution of (3.7) requires large training and feedback overhead. Moreover, the optimal digital linear precoder is not known in general even without the RF constraints, and only iterative solutions exist [65, 66]. Hence, the direct solution of this sum-rate maximization problem is neither practical nor tractable.

Similar problems to (3.7) have been studied before in literature, but with baseband (not hybrid) precoding and combining [65–71]. The main directions of designing the precoders/combiners in [65–71] can be summarized as follows.

- **Iterative Coordinated Beamforming Designs** The general idea of these algorithms is to iterate between the design of the precoder and combiners in multi-user MIMO downlink systems, with the aim of converging to a good solution [65,66]. These algorithms, however, require either the availability of global

channel knowledge at the transmitter, or the online BS-MS iterations to build the precoders and combiners. In mmWave systems, the application of coordinated beamforming is generally difficult as feeding the large mmWave channel matrix back to the BS requires a huge feedback overhead. Moreover, coordinated beamforming usually depends on using matching vectors at the MS's which can not be perfectly done with hybrid analog/digital architectures due to the hardware limitations on the analog precoders. Further, the convergence of coordinated beamforming has been established only for digital precoders [65,66], and the extension to hybrid precoders has not yet been studied.

- **Non-iterative Designs with Channel State Information at the Transmitter** To avoid the design complexity associated with iterative methods, some non-iterative sub-optimal algorithms, like block diagonalization, were proposed [67, 68]. Block diagonalization, however, requires global channel knowledge at the transmitter, which is difficult to achieve at mmWave systems. Further, the hardware constraints on the analog (or hybrid) precoding make it difficult to exactly design the pre-processing matrix to have no multi-user interference.

- **Non-iterative Designs with Channel State Information at the Receiver** The main idea of these schemes is to first combine the MIMO channel at each receiver according to a certain criterion. Then, each each user quantizes its effective channel based on a pre-defined codebook, and feeds it back to the BS which uses it to construct its multi-user precoding matrix [69–71]. The application of these precoding/combining algorithms in mmWave systems

is generally difficult because of the large dimensions of the mmWave channel matrix which makes the assumption of its availability at the MS's difficult to achieve in practice. Further, the hardware constraints make the direct application of the combining vector design schemes in [69–71] generally infeasible.

Given the practical difficulties associated with applying the prior precoding/combining algorithms in mmWave systems, we propose a new mmWave-suitable multi-user MIMO beamforming algorithm in Section 3.5. Our proposed algorithm is developed to achieve good performance compared with the solution of (3.7), while requiring (i) low training overhead and (ii) small feedback overhead. After explaining the developed algorithm in Section 3.5, its performance is analyzed in Section 3.6 assuming infinite-resolution feedback and neglecting channel estimation errors. The performance degradations due to limited feedback are then analyzed in Section 3.7.

3.5 Two-stage Multi-user Hybrid Precoding Algorithm

The additional challenge in solving (7), beyond the usual coupling between precoders and combiners [65–71], is the splitting of the precoding operation into two different domains, each with different constraints. The main idea of the proposed algorithm is to divide the calculation of the precoders into two stages. In the first stage, the BS RF precoder and the MS RF combiners are jointly designed to maximize the desired signal power of each user, neglecting the resulting interference among users. In the second stage, the BS digital precoder is designed to manage the multi-user interference.

Algorithm 3 Two-Stage Multi-user Hybrid Precoders Algorithm

Input: $\mathcal{F}$ BS RF beamforming codebook of size $|\mathcal{F}| = 2^{B_{\mathrm{RF}}^{\mathrm{BS}}}$

$\mathcal{W}$ MS RF beamforming codebook of size $|\mathcal{W}| = 2^{B_{\mathrm{RF}}^{\mathrm{MS}}}$

First stage: Single-user RF beamforming/combining design

For each MS $u, u = 1, 2, ..., U$

The BS and MS u select $\mathbf{v}_u^\star$ and $\mathbf{g}_u^\star$ that solve

$$\{\mathbf{g}_u^\star, \mathbf{v}_u^\star\} = \underset{\substack{\forall \mathbf{g}_u \in \mathcal{W} \\ \forall \mathbf{v}_u \in \mathcal{F}}}{\arg\max} \|\mathbf{g}_u^* \mathbf{H}_u \mathbf{v}_u\|$$

MS u sets $\mathbf{w}_u = \mathbf{g}_u^\star$

BS sets $\mathbf{F}_{\mathrm{RF}} = [\mathbf{v}_1^\star, \mathbf{v}_2^\star, ..., \mathbf{v}_U^\star]$

Second stage: Multi-user digital precoding design

For each MS $u, u = 1, 2, ..., U$

MS u estimates its effective channel $\overline{\mathbf{h}}_u = \mathbf{w}_u^* \mathbf{H}_u \mathbf{F}_{\mathrm{RF}}$

MS u quantizes $\overline{\mathbf{h}}_u$ using a codebook $\mathcal{H}$ of size $2^{B_{\mathrm{BB}}}$ and feeds back $\widehat{\mathbf{h}}_u$ where

$$\widehat{\mathbf{h}}_u = \underset{\widehat{\mathbf{h}}_u \in \mathcal{H}}{\arg\max} \|\overline{\mathbf{h}}_u^* \widehat{\mathbf{h}}_u\|$$

BS designs $\mathbf{F}_{\mathrm{BB}} = \widehat{\mathbf{H}}^* \left(\widehat{\mathbf{H}} \widehat{\mathbf{H}}^*\right)^{-1}$ with $\widehat{\mathbf{H}} = \left[\widehat{\mathbf{h}}_1^{\mathrm{T}}, ..., \widehat{\mathbf{h}}_U^{\mathrm{T}}\right]^{\mathrm{T}}$

$\mathbf{f}_u^{\mathrm{BB}} = \frac{\mathbf{f}_u^{\mathrm{BB}}}{\|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_u^{\mathrm{BB}}\|_F}, u = 1, 2, ..., U$

Algorithm 3 can be summarized as follows. In the first stage, the BS and each MS u design the RF beamforming and combining vectors, $\mathbf{f}_u^{\mathrm{RF}}$ and $\mathbf{w}_u$, to maximize the desired signal power for user u, and neglecting the other users' interference. As this is the typical single-user RF beamforming design problem, efficient beam training algorithms developed for single-user systems such as [13, 72], which do not require explicit channel estimation and have a low training overhead, can be used to design the RF beamforming/combining vectors.

In the second stage, the BS trains the effective channels, $\overline{\mathbf{h}}_u = \mathbf{w}_u^* \mathbf{H}_u \mathbf{F}_{\mathrm{RF}}, u = 1, 2, ..., U$, with the MS's. Note that the dimension of each effective channel vector is $U \times 1$ which is much less than the original channel matrix. This is not the case for the

algorithms developed in [69, 71] in which the effective channels have larger $N_{\mathrm{BS}} \times 1$ dimensions. Then, each MS u quantizes its effective channel using a codebook $\mathcal{H}$, and feeds the index of the quantized channel vector back to the BS with B_{BB} bits. Finally, the BS designs its zero-forcing digital precoder based on the quantized channels. Thanks to the narrow beamforming and the sparse mmWave channels, the effective MIMO channel is expected to be well-conditioned [73, 74], which makes adopting a simple multi-user digital beamforming strategy like zero-forcing capable of achieving near-optimal performance [75], as will be shown in Sections 3.6-3.7.

Both the separate and joint designs of the analog and digital precoders were investigated before for single-user mmWave systems. For example, the work in [62] considered a single-user single-stream MIMO-OFDM system, where the analog and digital precoders were sequentially designed to maximize either the received signal strength or the sum-rate over different frequency sub-carriers. Alternatively, the analog and digital precoders were jointly designed in [20, 27, 62] to maximize the rate of single-user systems. In this work, we consider a different setup which is multi-user downlink transmission. Therefore, the objective of the hybrid analog/digital beamforming in our work is different than that in [20, 27, 62] as we need to manage the multi-user interference as well. This leads to a completely different analysis.

In the next two sections, we analyze the performance of the proposed multi-user hybrid precoding algorithm in different settings. For this analysis, we adopt the beamsteering codebook for the design of the analog beamforming/combining vectors. We also assume that the effective channels in the second stage of Algorithm 3 are quantized using a RVQ codebook. RVQ simplifies the analytical performance analysis

of the proposed algorithm and allows leveraging some results from the limited feedback MIMO literature [66, 69, 70].

3.6 Performance Analysis with Infinite Codebooks

The analysis of hybrid precoding is non-trivial due to the coupling between analog and digital precoders. Therefore, we will study the performance of the proposed algorithm in two cases: With single-path channels and with large numbers of antennas. These cases are of special interest as mmWave channels are likely to be sparse, i.e., only a few paths exist [21], and both the BS and MS need to employ large antenna arrays to have sufficient received power [20]. Further, the analysis of these special cases will give useful insights into the performance of the proposed algorithms in more general settings which will also be confirmed by the simulations in Section 3.8.

In this section, we analyze the achievable rates of the proposed algorithm assuming perfect effective channel knowledge and supposing that the angles of the RF beamsteering vectors can take continuous values, i.e., we assume that both the RF codebooks ($\mathcal{F}$ and $\mathcal{W}$) and the RVQ codebook $\mathcal{H}$ are of infinite size. In Section 3.7, we will study how limited feedback and finite codebooks affect the rates achieved by the developed hybrid precoding algorithm.

3.6.1 Single-Path Channels

In this section, we consider the case when $L_u = 1, u = 1, 2, ..., U$. For ease of exposition, we will omit the subscript ℓ in the definition of the channel parameters in

(3.4). The following theorem characterizes a lower bound on the achievable rate by each MS when Algorithm 3 is used to design the hybrid precoders at the BS and RF combiners at the MS's.

Theorem 4 *Let Algorithm 3 be used to design the hybrid precoders and RF combiners described in Section 3.3 under the following assumptions*

1. *All channels are single-path, i.e.,* $L_u = 1, u = 1, 2, ..., U$.
2. *The RF precoding vectors* $\mathbf{f}_u^{\mathrm{RF}}, u = 1, 2, ..., U$, *and the RF combining vectors* $\mathbf{w}_u, u = 1, 2, ..., U$ *are beamsteering vectors with continuous angles.*
3. *Each MS* u *perfectly knows its channel* $\mathbf{H}_u$, $u = 1, 2, ..., U$.
4. *The BS perfectly knows the effective channels* $\overline{\mathbf{h}}_u, u = 1, 2, ..., U$.

and define the $N_{\mathrm{BS}} \times U$ *matrix* $\mathbf{A}_{\mathrm{BS}}$ *to gather the BS array response vectors associated with the* U *AoDs, i.e.,* $\mathbf{A}_{\mathrm{BS}} = [\mathbf{a}_{\mathrm{BS}}(\phi_1), \mathbf{a}_{\mathrm{BS}}(\phi_2), ..., \mathbf{a}_{\mathrm{BS}}(\phi_U)]$, *with maximum and minimum singular values* $\sigma_{\max}(\mathbf{A}_{\mathrm{BS}})$ *and* $\sigma_{\min}(\mathbf{A}_{\mathrm{BS}})$, *respectively. Then, the achievable rate of user* u *is lower bounded by*

$$R_u \geq \log_2 \left(1 + \frac{\mathsf{SNR}}{U} N_{\mathrm{BS}} N_{\mathrm{MS}} |\alpha_u|^2 G\left(\{\phi_u\}_{u=1}^U\right)\right), \tag{3.8}$$

where $G\left(\{\phi_u\}_{u=1}^U\right) = 4\left(\frac{\sigma_{\max}^2(\mathbf{A}_{\mathrm{BS}})}{\sigma_{\min}^2(\mathbf{A}_{\mathrm{BS}})} + \frac{\sigma_{\min}^2(\mathbf{A}_{\mathrm{BS}})}{\sigma_{\max}^2(\mathbf{A}_{\mathrm{BS}})} + 2\right)^{-1}$, $\mathsf{SNR} = \frac{P}{\sigma^2}$.

Proof: Consider the BS and MS's with the system and channels described in Section 3.3. Then, in the first stage of Algorithm 3, the BS and each MS u find $\mathbf{v}_u^\star$

and $\mathbf{g}_u^\star$ that solve

$$\{\mathbf{g}_u^\star, \mathbf{v}_u^\star\} = \underset{\substack{\forall \mathbf{g}_u \in \mathcal{W} \\ \forall \mathbf{v}_u \in \mathcal{F}}}{\arg\max} \|\mathbf{g}_u^* \mathbf{H}_u \mathbf{v}_u\|. \tag{3.9}$$

As the channel $\mathbf{H}_u$ has only one path, and given the continuous beamsteering capability assumption, the optimal RF precoding and combining vectors will be $\mathbf{g}_u^\star = \mathbf{a}_{\mathrm{MS}}(\theta_u)$, and $\mathbf{v}_u^\star = \mathbf{a}_{\mathrm{BS}}(\phi_u)$. Consequently, the MS sets $\mathbf{w}_u = \mathbf{a}_{\mathrm{MS}}(\theta_u)$ and the BS takes $\mathbf{f}_u^{\mathrm{RF}} = \mathbf{a}_{\mathrm{BS}}(\phi_u)$. Gathering the beamforming vectors for the U users, the BS RF beamforming matrix is then $\mathbf{F}_{\mathrm{RF}} = \mathbf{A}_{\mathrm{BS}} = [\mathbf{a}_{\mathrm{BS}}(\phi_1), \mathbf{a}_{\mathrm{BS}}(\phi_2), ..., \mathbf{a}_{\mathrm{BS}}(\phi_U)]$.

The effective channel for user u after designing the RF precoders and combiners is

$$\begin{aligned} \overline{\mathbf{h}}_u &= \mathbf{w}_u \mathbf{H}_u \mathbf{F}_{\mathrm{RF}} \\ &= \sqrt{N_{\mathrm{BS}} N_{\mathrm{MS}}} \alpha_u \mathbf{a}_{\mathrm{BS}}^*(\phi_u) \mathbf{F}_{\mathrm{RF}}. \end{aligned} \tag{3.10}$$

Now, defining $\overline{\mathbf{H}} = [\overline{\mathbf{h}}_1^{\mathrm{T}}, \overline{\mathbf{h}}_2^{\mathrm{T}}, ..., \overline{\mathbf{h}}_U^{\mathrm{T}}]^{\mathrm{T}}$, and given the design of $\mathbf{F}_{\mathrm{RF}}$, we can write the effective channel matrix $\overline{\mathbf{H}}$ as

$$\overline{\mathbf{H}} = \mathbf{D} \mathbf{A}_{\mathrm{BS}}^* \mathbf{A}_{\mathrm{BS}}, \tag{3.11}$$

where $\mathbf{D}$ is a $U \times U$ diagonal matrix with $[\mathbf{D}]_{u,u} = \sqrt{N_{\mathrm{BS}} N_{\mathrm{MS}}} \alpha_u$.

Based on this effective channel, the BS zero-forcing digital precoder is defined as

$$\mathbf{F}_{\mathrm{BB}} = \overline{\mathbf{H}}^* \left(\overline{\mathbf{H}} \overline{\mathbf{H}}^* \right)^{-1} \mathbf{\Lambda}, \tag{3.12}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with the diagonal elements adjusted to satisfy the precoding power constraints $\left\| \mathbf{F}_{\mathrm{RF}} \mathbf{f}_u^{\mathrm{BB}} \right\|^2 = 1, u = 1, 2, ..., U$. The diagonal elements

of $\mathbf{\Lambda}$ are then equal to

$$\mathbf{\Lambda}_{u,u} = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{\left(\mathbf{A}_{\mathrm{BS}}^{*} \mathbf{A}_{\mathrm{BS}}\right)_{u,u}^{-1}}} \left|\alpha_u\right|, u = 1, 2, ..., U. \tag{3.13}$$

Note that this $\mathbf{\Lambda}$ is different than the traditional digital zero-forcing precoder due to the different power constraints in the hybrid analog/digital architecture. [See Appendix A in [30] for a derivation]

The achievable rate for user u is then

$$\begin{aligned} R_u &= \log_2\left(1 + \frac{\mathsf{SNR}}{U}\left|\overline{\mathbf{h}}_u^{*} \mathbf{f}_u^{\mathrm{BB}}\right|^2\right), \\ &= \log_2\left(1 + \frac{\mathsf{SNR}}{U} \frac{N_{\mathrm{BS}} N_{\mathrm{MS}} \left|\alpha_u\right|^2}{\left(\mathbf{A}_{\mathrm{BS}}^{*} \mathbf{A}_{\mathrm{BS}}\right)_{u,u}^{-1}}\right). \end{aligned} \tag{3.14}$$

To bound this rate, the following lemma which characterizes a useful property of the matrix $\mathbf{A}_{\mathrm{BS}}^{*} \mathbf{A}_{\mathrm{BS}}$ can be used.

Lemma 5 *Assume* $\mathbf{A}_{\mathrm{BS}} = \left[\mathbf{a}_{\mathrm{BS}}\left(\phi_1\right), ..., \mathbf{a}_{\mathrm{BS}}\left(\phi_U\right)\right]$*, with the angles* $\phi_u, u = 1, 2, ..., U$ *taking continuous values in* $[0, 2\pi]$*, then the matrix* $\mathbf{P} = \mathbf{A}_{\mathrm{BS}}^{*} \mathbf{A}_{\mathrm{BS}}$ *is positive definite almost surely.*

Proof: Let the matrix $\mathbf{P} = \mathbf{A}_{\mathrm{BS}}^{*} \mathbf{A}_{\mathrm{BS}}$, then for any non-zero complex vector $\mathbf{z} \in \mathcal{C}^U$, it follows that $\mathbf{z}^{*} \mathbf{P} \mathbf{z} = \left\|\mathbf{A}_{\mathrm{BS}} \mathbf{z}\right\|_2^2 \geq 0$. Hence, the matrix $\mathbf{P}$ is positive semi-definite. Further, if the vectors $\mathbf{a}_{\mathrm{BS}}\left(\phi_1\right), \mathbf{a}_{\mathrm{BS}}\left(\phi_2\right)$, ..., $\mathbf{a}_{\mathrm{BS}}\left(\phi_U\right)$ are linearly independent, then for any non-zero complex vector $\mathbf{z}$, $\mathbf{A}_{\mathrm{BS}} \mathbf{z} \neq 0$, and the matrix $\mathbf{P}$ is positive definite. To show that, consider any two vectors $\mathbf{a}_{\mathrm{BS}}\left(\phi_u\right), \mathbf{a}_{\mathrm{BS}}\left(\phi_n\right)$. These vectors are linearly dependent if and only if $\phi_u = \phi_n$. As the probability of this event equals zero when

the AoDs ϕ_u and ϕ_n are selected independently from a continuous distribution, then the matrix $\mathbf{P}$ is positive definite with probability one. $\square$

Now, using the Kantorovich inequality [76], we can bound the diagonal entries of the matrix $\left(\mathbf{A}_{\mathrm{BS}}^{*}\mathbf{A}_{\mathrm{BS}}\right)^{-1}$ using the following lemma from [77].

Lemma 6 *For any $n \times n$ Hermitian and positive definite matrix $\mathbf{P}$ with the ordered eigenvalues satisfying $0 < \lambda_{\min} \leq \lambda_2 \leq ... \leq \lambda_{\max}$, the element $(\mathbf{P})_{u,u}^{-1}, u = 1, 2, ..., n$ satisfies*

$$(\mathbf{P})_{u,u}^{-1} \leq \frac{1}{4[\mathbf{P}]_{u,u}} \left(\frac{\lambda_{\max}(\mathbf{P})}{\lambda_{\min}(\mathbf{P})} + \frac{\lambda_{\min}(\mathbf{P})}{\lambda_{\max}(\mathbf{P})} + 2 \right). \tag{3.15}$$

Finally, noting that $\left(\mathbf{A}_{\mathrm{BS}}^{*}\mathbf{A}_{\mathrm{BS}}\right)_{u,u} = 1$, $\lambda_{\min}\left(\mathbf{A}_{\mathrm{BS}}^{*}\mathbf{A}_{\mathrm{BS}}\right) = \sigma_{\min}^{2}\left(\mathbf{A}_{\mathrm{BS}}\right)$, and $\lambda_{\max}\left(\mathbf{A}_{\mathrm{BS}}^{*}\mathbf{A}_{\mathrm{BS}}\right) = \sigma_{\max}^{2}\left(\mathbf{A}_{\mathrm{BS}}\right)$ and using lemma 6, we get the lower bound on the achievable rate in (3.8). $\square$

In addition to characterizing a lower bound on the rates achieved by the proposed hybrid analog/digital precoding algorithm, the bound in (3.8) separates the dependence on the channel gains α_u, and the AoDs $\phi_u, u = 1, 2, ..., U$ which can be used to claim the optimality of the proposed algorithm in some cases and to give useful insights into the gain of the proposed algorithm over analog-only beamsteering solutions. This is illustrated in the following results.

Proposition 7 *Denote the single-user rate as $\mathring{R}_u = \log_2\left(1 + \frac{\mathsf{SNR}}{U} N_{\mathrm{BS}} N_{\mathrm{MS}} |\alpha_u|^2\right)$. When Algorithm 3 is used to design the hybrid precoders and RF combiners described*

in Section 3.3, and given the assumptions stated in Theorem 4, the relation between the achievable rate by any user u, and the single-user rate, $\mathring{R}_u$ satisfies

1. $\mathbb{E}\left[\mathring{R}_u - R_u\right] \leq K\left(N_{\mathrm{BS}}, U\right)$.
2. $\lim_{N_{\mathrm{BS}}\to\infty} R_u = \mathring{R}_u$ *almost surely.*

where $K\left(N_{\mathrm{BS}}, U\right)$ is a constant whose value depends only on N_{BS} and U.

Proof: See Appendix B in [30]. □

Proposition 7 indicates that the average achievable rate of any user u using the proposed low-complexity precoding/combining algorithm grows with the same slope of the single-user rate at high SNR, and stays within a constant gap from it. This gap, $K\left(N_{\mathrm{BS}}, U\right)$, depends only on the number of users and the number of BS antennas. As the number of BS antennas increases, the matrix $\mathbf{A}_{\mathrm{BS}}$ becomes more well-conditioned, and the ratio between its maximum and minimum singular values will approach one. Hence, the value of $G\left(\{\phi_u\}_{u=1}^{U}\right)$ in (3.8) will be closer to one, and the gap between the achievable rate using Algorithm 3 and the single-user rate will decrease. This will also be shown by numerical simulations in Section 3.8. One important note here is that this gap does not depend on the number of MS antennas, which is contrary to the analog-only beamsteering, given by the first stage only of Algorithm 3. This leads to the following corollary.

Corollary 8 *Let R_{BS} denote the rate achieved by user u when the BS employs analog-only beamsteering designed according to Step* 1 *of Algorithm 3. Then, the relation*

between the average achievable rate using Algorithm 3 R_u and the average rate of analog-only beamsteering solution when the number of MS antennas goes to infinity satisfies: $\lim_{N_{\mathrm{MS}}\to\infty} \mathbb{E}\left[R_u - R_{\mathrm{BS}}|\right] = \infty$.

Proof: See Appendix C in [30]. □

This corollary implies that multi-user interference management is still important at mmWave systems even when very large numbers of antennas are used at the BS and MS's. Note also that this is not the case when the number of BS antennas goes to infinity as it can be easily shown that the performance of RF beamsteering alone becomes optimal in this case.

3.6.2 Large-dimensional Regime

Under the assumption of large numbers of transmit antennas, a different approximation of the achievable rate can be derived. We approach this problem using the virtual channel model framework and its simplifications in large MIMO systems [64] [78]. The results of this section are, therefore, valid only for uniform arrays, e.g., ULAs and UPAs [64,79]. The virtual channel model characterizes physical channels via joint spatial beams in *fixed* virtual transmit and receive directions exploiting the finite dimensionality of the MIMO system, i.e., the finite number of transmit and receive antennas. The virtual transmit and receive directions are fixed because they depend only on the number of BS and MS antennas. Hence, they are common for the different users with the same number of antennas. Using this channel model, the

uth user channel $\mathbf{H}_u$ can be written as [64]

$$\mathbf{H}_u = \overline{\mathbf{A}}_{\mathrm{MS}} \mathbf{H}_u^{\mathrm{v}} \overline{\mathbf{A}}_{\mathrm{BS}}^{*}, \tag{3.16}$$

where $\overline{\mathbf{A}}_{\mathrm{BS}} = [\mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_1\right), \mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_2\right), ..., \mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_{N_{\mathrm{BS}}}\right)]$ is an $N_{\mathrm{BS}} \times N_{\mathrm{BS}}$ matrix carrying the BS array response vectors in the virtual directions $\bar{\phi}_p, p = 1, 2, ..., N_{\mathrm{BS}}$ that satisfy $\frac{2\pi d}{\lambda} \sin\left(\bar{\phi}_p\right) = \frac{2\pi p}{N_{\mathrm{BS}}}$. Similarly, $\overline{\mathbf{A}}_{\mathrm{MS}} = [\mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_1\right), \mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_2\right), ..., \mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_{N_{\mathrm{MS}}}\right)]$ carries the MS array response vectors in the virtual directions $\bar{\theta}_q, q = 1, 2, ..., N_{\mathrm{MS}}$ that satisfy $\frac{2\pi d}{\lambda} \sin\left(\bar{\theta}_q\right) = \frac{2\pi q}{N_{\mathrm{MS}}}$. Thanks to these special virtual channel angles, the matrices $\overline{\mathbf{A}}_{\mathrm{BS}}$ and $\overline{\mathbf{A}}_{\mathrm{MS}}$ are DFT matrices [64]. Finally, $\mathbf{H}_u^{\mathrm{v}}$ is the uth MS virtual channel matrix with each element $[\mathbf{H}_u^{\mathrm{v}}]_{q,p}$ representing a group of physical spatial paths, and approximately equal to the sum of the gains of those paths [64].

One advantage of using the virtual channel model in analyzing our proposed multi-user precoding algorithm lies in the fact that it provides a common space of the transmit eigenvectors of the different users. This means that the BS eigenvectors for each MS form a subset of the columns of the DFT matrix $\overline{\mathbf{A}}_{\mathrm{BS}}$. The virtual channel model also provides a simple way to incorporate the angle spread associated with mmWave channel scatterers by defining each element of the virtual channel matrix as the sum of the channel gains associated with the scatterers located in a certain direction multiplied by the integration of the spatial spreading functions of these scatterers [64].

Before leveraging this channel model in analyzing the proposed hybrid precoding algorithm, we rewrite the channel in (3.16) as

$$\mathbf{H}_u = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{L_u}} \sum_{m=1}^{N_{\mathrm{BS}} N_{\mathrm{MS}}} \gamma_{u,m} \mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_{u,m}\right) \mathbf{a}_{\mathrm{BS}}^{*}\left(\bar{\phi}_{u,m}\right), \tag{3.17}$$

where $|\gamma_{u,1}| \geq |\gamma_{u,2}| \geq ... \geq |\gamma_{u,N_{\mathrm{BS}}N_{\mathrm{MS}}}|$. $\gamma_{u,m}$ equals the element in $\mathbf{H}_u^{\mathrm{v}}$ with the mth largest virtual channel element magnitude, and $\bar{\phi}_{u,m}, \bar{\theta}_{u,m}$ are the corresponding transmit and receive virtual directions, respectively.

In the following proposition, we use this channel model to characterize a simple lower bound on the achievable rate of Algorithm 3 for arbitrary numbers of channel paths assuming for simplicity that $L_u = L, u = 1, 2, ..., U$. The derived results give useful analytical insights into the asymptotic performance of the proposed algorithm in the multi-path case.

Proposition 9 *Define the single-user rate as* $\mathring{R}_u = \log 2(1 + \frac{\mathsf{SNR}}{UL} N_{\mathrm{BS}} N_{\mathrm{NS}} |\gamma_{u,1}|^2)$. *Then, when Algorithm 3 is used to design the hybrid analog/digital precoders at the BS and RF combiners at the MSs, with the assumptions in Theorem 4, and adopting the virtual channel model in* (3.17)*, the average achievable rate of user* u *is lower bounded by*

$$\mathbb{E}\left[R_u\right] \geq \tag{3.18}$$

$$\mathbb{E}\left[\mathring{R}_u\right] \left(\prod_{i=1}^{U-1} \left(1 - \frac{i}{N_{\mathrm{BS}}}\right) \left(1 - \frac{L-1}{N_{\mathrm{MS}}}\right)^U + \mathbb{1}_{(L>1)} \prod_{i=1}^{U-1} \left(1 - \frac{iL}{N_{\mathrm{BS}}}\right) \left(\frac{1}{N_{\mathrm{MS}}}\right)^{(L-1)U} \right). \tag{3.19}$$

Proof: Consider the BS and MS's with the system model described in Section 3.3, and the approximated channel model in (3.17). In the first stage of Algorithm 3, the BS and each MS u find $\mathbf{v}_u^\star$ and $\mathbf{g}_u^\star$ that solve (3.9). Given the virtual channel model in (3.17), we get $\mathbf{w}_u = \mathbf{g}_u^\star = \mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_{u,1}\right)$ and $\mathbf{v}_u^\star = \mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_{u,1}\right)$. Consequently, the RF precoder at the BS becomes $\mathbf{F}_{\mathrm{RF}} = \left[\mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_{1,1}\right), \mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_{2,1}\right), ..., \mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_{U,1}\right)\right]$. Now, we

can write the uth MS effective channel as

$$\overline{\mathbf{h}}_u = \mathbf{w}_u^* \mathbf{H}_u \mathbf{F}_{\mathrm{RF}} = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{L}} \gamma_{u,1} \left[\zeta_{u,1}, \zeta_{u,2}, ..., \zeta_{u,U}\right], \tag{3.20}$$

where the values of the $\zeta_{u,n}$ elements are

- $\zeta_{u,u} = 1$,
- $\zeta_{u,n} = \mathbb{1}_{\left(\bar{\phi_{u,1}} = \bar{\phi}_{n,1}\right)} + \sum_{m=2}^{L} \frac{\gamma_{u,2}}{\gamma_{u,1}} \mathbb{1}_{\left(\bar{\phi}_{u,m} = \bar{\phi}_{n,1}\right)} \mathbb{1}_{\left(\bar{\theta}_{u,1} = \bar{\theta}_{u,m}\right)}, \quad \forall n \neq u,$

where the summation in $\zeta_{u,n}$ is over the first L elements only due to the sparse channel. Note that the characterization of $\zeta_{u,n}$ is due to the DFT structure of the matrices $\overline{\mathbf{A}}_{\mathrm{BS}}$ and $\overline{\mathbf{A}}_{\mathrm{MS}}$.

The overall effective channel, $\overline{\overline{\mathbf{H}}}$, can be then written as

$$\overline{\overline{\mathbf{H}}} = \mathbf{D}_{\mathrm{v}} \mathbf{P}_{\mathrm{v}}, \tag{3.21}$$

where $\mathbf{D}_{\mathrm{v}}$ is a diagonal matrix with the diagonal elements $[\mathbf{D}_{\mathrm{v}}]_{u,u} = \sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{L}} \gamma_{u,1}, u = 1, 2, ..., U$, and the $U \times U$ matrix $\mathbf{P}_{\mathrm{v}}$ has $[\mathbf{P}_{\mathrm{v}}]_{u,n} = \zeta_{u,n}, \forall u, n$.

The digital zero-forcing precoder is therefore $\mathbf{F}_{\mathrm{BB}} = \overline{\overline{\mathbf{H}}}^* \left(\overline{\overline{\mathbf{H}}} \overline{\overline{\mathbf{H}}}^*\right)^{-1} \mathbf{\Lambda}$, and the diagonal elements of $\mathbf{\Lambda}$ are chosen to satisfy the precoding power constraint $\left\|\mathbf{F}_{\mathrm{RF}} \mathbf{f}_u^{\mathrm{BB}}\right\|^2 = 1$. Using a similar derivation to that in [30, Appendix A], we get

$$[\mathbf{\Lambda}]_{u,u} = \frac{\sqrt{\frac{N_{\mathrm{BS}} N_{\mathrm{MS}}}{L}} |\gamma_{u,1}|}{\sqrt{\left(\left(\mathbf{P}_{\mathrm{v}} \mathbf{P}_{\mathrm{v}}^*\right)^{-1} \mathbf{P}_{\mathrm{v}} \mathbf{F}_{\mathrm{RF}}^* \mathbf{F}_{\mathrm{RF}} \mathbf{P}_{\mathrm{v}}^* \left(\mathbf{P}_{\mathrm{v}} \mathbf{P}_{\mathrm{v}}^*\right)^{-1}\right)_{u,u}}}. \tag{3.22}$$

Using the designed digital and analog precoders, the rate of user u can be written as

$$\mathbb{E}\left[R_u\right] = \mathbb{E}\left[\log_2\left(1 + \frac{\mathsf{SNR}}{U} \left|\overline{\mathbf{h}}_u^* \mathbf{f}_u^{\mathrm{BB}}\right|^2\right)\right], \tag{3.23}$$

$$= \mathbb{E}\left[\log_2\left(1+\frac{\mathsf{SNR}}{UL}\frac{N_{\mathrm{BS}}N_{\mathrm{MS}}\left|\gamma_{u,1}\right|^2}{\left(\left(\mathbf{P}_{\mathrm{v}}\mathbf{P}_{\mathrm{v}}^*\right)^{-1}\mathbf{P}_{\mathrm{v}}\mathbf{F}_{\mathrm{RF}}^*\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{v}}^*\left(\mathbf{P}_{\mathrm{v}}\mathbf{P}_{\mathrm{v}}^*\right)^{-1}\right)_{u,u}}\right)\right]. \tag{3.24}$$

Now, we note that the term $\left(\left(\mathbf{P}_{\mathrm{v}}\mathbf{P}_{\mathrm{v}}^*\right)^{-1}\mathbf{P}_{\mathrm{v}}\mathbf{F}_{\mathrm{RF}}^*\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{v}}^*\left(\mathbf{P}_{\mathrm{v}}\mathbf{P}_{\mathrm{v}}^*\right)^{-1}\right)_{u,u} = 1$ if $\mathbf{P}_{\mathrm{v}} = \mathbf{I}$. Then, considering only the case when $\mathbf{P}_{\mathrm{v}} = \mathbf{I}$ gives a simple lower bound on the achievable rate

$$\mathbb{E}\left[R_u\right] \geq \mathbb{E}\left[\log_2\left(1+\frac{\mathsf{SNR}}{UL}N_{\mathrm{BS}}N_{\mathrm{MS}}\left|\gamma_{u,1}\right|^2\right)\mathbb{1}\left(\mathbf{P}_{\mathrm{v}} = \mathbf{I}_U\right)\right], \tag{3.25}$$

$$\overset{(a)}{=} \mathbb{E}\left[\log_2\left(1+\frac{\mathsf{SNR}}{UL}N_{\mathrm{BS}}N_{\mathrm{MS}}\left|\gamma_{u,1}\right|^2\right)\right]\mathrm{P}\left(\mathbf{P}_{\mathrm{v}} = \mathbf{I}_U\right), \tag{3.26}$$

where (a) is by leveraging the independence between $\gamma_{u,1}$ and the virtual transmit angles of the different users. Thanks to the sparse nature of mmWave channels, this simple bound in (3.26) can be a tight bound on the achievable rate. Finally, the probability of the event $\mathbf{P}_{\mathrm{v}} = \mathbf{I}$ can be bounded as follows by considering only the cases when all the AoAs are equal or all of them are different

$$\mathrm{P}\left(\mathbf{P}_{\mathrm{v}} = \mathbf{I}_U\right) \geq \mathbb{P}\left(\mathbf{P}_{\mathrm{v}} = \mathbf{I}_U \middle| \bigcap_{u=1}^{U}\left(\bar{\theta}_{u,1} \neq \bar{\theta}_{u,m}, \forall m \neq 1\right)\right)\mathrm{P}\left(\bigcap_{u=1}^{U}\left(\bar{\theta}_{u,1} \neq \bar{\theta}_{u,m}, \forall m \neq 1\right)\right)$$
$$+ \mathrm{P}\left(\mathbf{P}_{\mathrm{v}} = \mathbf{I}_U \middle| \bigcap_{u=1}^{U}\left(\bar{\theta}_{u,1} = \bar{\theta}_{u,m}, \forall m \neq 1\right)\right)\mathrm{P}\left(\bigcap_{u=1}^{U}\left(\bar{\theta}_{u,1} = \bar{\theta}_{u,m}, \forall m \neq 1\right)\right), \tag{3.27}$$

$$\geq \prod_{i=1}^{U-1}\left(1-\frac{i}{N_{\mathrm{BS}}}\right)\left(1-\frac{L-1}{N_{\mathrm{MS}}}\right)^U + \mathbb{1}_{(L>1)}\prod_{i=1}^{U-1}\left(1-\frac{iL}{N_{\mathrm{BS}}}\right)\left(\frac{1}{N_{\mathrm{MS}}}\right)^{(L-1)U}, \tag{3.28}$$

where all these probabilities are calculated from the expression of $\zeta_{u,n}, n \neq u$ (the off-diagonal entries of $\mathbf{P}_{\mathrm{v}}$). □

This bound shows the asymptotic optimality of the sum-rate achieved by the proposed hybrid precoding algorithm in the large-dimensional regime, as it approaches

1 with large numbers of antennas. Hence, the average achievable rate by the proposed algorithm in (3.26) will be very close to the single-user rate. Indeed, this simple bound can be shown to be tight when the number of paths is very small compared with the number of antennas which is the case in mmWave systems. Also, this bound shows the relatively small importance of the other paths, rather than the strongest path, on the performance as $\frac{L}{N_{\mathrm{BS}}} \ll 1$ and $\frac{L-1}{N_{\mathrm{MS}}} \ll 1$. Finally, note that the bound in (3.19) is an approximated bound, as it depends on the asymptotic properties of the virtual channel model in (3.17), which becomes a good approximation when the number of antennas is very large.

3.7 Rate Loss with Limited Feedback

In this section, we consider RF and digital codebooks with *finite* sizes, and analyze the rate loss due to the joint RF/baseband quantization. Although the analysis will consider the special cases of single-path mmWave channels, and large-dimensional regimes, it helps making important conclusions about the performance of the hybrid precoding over finite-rate feedback channels.

3.7.1 Single-Path Channels

Considering single-path mmWave channels, the following theorem characterizes the average rate loss when the hybrid analog/digital precoders and RF combiners are designed according to Algorithm 3 with the quantized beamsteering RF precoders $\mathcal{F}, \mathcal{W}$, and the effective channel RVQ codebook $\mathcal{H}$.

Theorem 10 *Let R_u^{Q} denote the rate achieved by user u when Algorithm 3 is used to*

design the hybrid precoders and RF combiners described in Section 3.3 while assuming that

1. *All channels are single-path, i.e.,* $L_u = 1, u = 1, 2, ..., U$.
2. *The RF precoding and combining vectors,* $\mathbf{f}_u^{\mathrm{RF}}, u = 1, 2, ..., U$ *and* $\mathbf{w}_u, u = 1, 2, ..., U$, *are beamsteering vectors selected from the quantized codebooks* $\mathcal{F}$ *and* $\mathcal{W}$.
3. *Each MS* u *perfectly knows its channel* $\mathbf{H}_u$, $u = 1, 2, ..., U$.
4. *Each MS* u *quantizes its effective channel* $\overline{\mathbf{h}}_u$ *using a RVQ codebook* $\mathcal{H}$ *of size* $|\mathcal{H}| = 2^{B_{\mathrm{BB}}}$.

Recall that R_u *is the rate achieved by user* u *with the assumptions in Theorem 4. Then the average rate loss per user,* $\overline{\Delta R}_u = \mathbb{E}\left[R_u - R_u^{\mathrm{Q}}\right]$, *is upper bounded by*

$$\overline{\Delta R}_u \leq \log_2 \left(\frac{1 + \frac{\mathsf{SNR}}{U} N_{\mathrm{BS}} N_{\mathrm{MS}} \bar{\alpha} \left(1 + \frac{U-1}{N_{\mathrm{BS}}}\right) 2^{-\frac{B_{\mathrm{BB}}}{U-1}}}{\left|\overline{\mu}_{\mathrm{BS}}\right|^2 \left|\overline{\mu}_{\mathrm{MS}}\right|^2} \right), \tag{3.29}$$

where $|\overline{\mu}_{\mathrm{BS}}| = \min\limits_{\mathbf{f}_u \in \mathcal{F}} \max\limits_{\mathbf{f}_n \in \mathcal{F}} |\mathbf{f}_u^* \mathbf{f}_n|$, *and* $|\overline{\mu}_{\mathrm{MS}}| = \min\limits_{\mathbf{w}_u \in \mathcal{W}} \max\limits_{\mathbf{w}_n \in \mathcal{W}} |\mathbf{w}_u^* \mathbf{w}_n|$.

Proof: See [30, Appendix D]. □

Theorem 10 characterizes an upper bound on the rate loss due to quantization. It can be used to determine how the number of baseband and RF quantization bits should scale with the different system and channel parameters to be within a constant gap of the optimal rate. This is captured in the following corollary.

Corollary 11 *To maintain a rate loss of* $\log_2(b)$ *bps/Hz per user, the number of baseband quantization bits should satisfy*

$$B_{\mathrm{BB}} \geq \frac{U-1}{3}\mathsf{SNR}_{\mathrm{dB}} + (U-1)\log_2\left(\frac{N_{\mathrm{BS}}N_{\mathrm{MS}}}{U}\bar{\alpha}\left(1-\frac{U-1}{N_{\mathrm{BS}}}\right)\right) - (U-1)\log_2\left(\left|\overline{\mu}_{\mathrm{BS}}\right|^2\left|\overline{\mu}_{\mathrm{MS}}\right|^2 b - 1\right), \tag{3.30}$$

$$\left|\overline{\mu}_{\mathrm{BS}}\right|^2\left|\overline{\mu}_{\mathrm{MS}}\right|^2 > \frac{1}{b}. \tag{3.31}$$

This corollary shows that the number of bits used to quantize the effective channels should increase linearly with the SNR in dB for any given number of users and logarithmically with the number of antennas. It also illustrates that more baseband quantization bits will be needed if the RF beamsteering vectors are poorly quantized, i.e., if $|\overline{\mu}_{\mathrm{BS}}|$ and $|\overline{\mu}_{\mathrm{MS}}|$ are small.

The relation between the RF and baseband quantization bits is important to understand the behavior of hybrid precoding algorithms. Indeed, in some cases, e.g., when the effective channel is poorly quantized, the performance of analog-only beamforming can exceed that of the hybrid precoding. In Section 3.8, the hybrid precoding and beamsteering algorithms are compared for different quantization settings, and some insights are given to highlight the cases in which using a digital layer to manage the multi-user interference is useful.

3.7.2 Large-dimensional Regime

When large antenna arrays are used at both the BS and MS's, using the virtual channel model in Section 3.6.2, we can bound the average rate loss using the proposed hybrid precoding algorithm with finite size codebooks.

Proposition 12 *Using Algorithm 3 to design the hybrid precoders at the BS and RF combiners at the MSs, with the assumptions in Theorem 10, and adopting the virtual channel model in* (3.17)*, the average rate loss per user due to quantization,* $\overline{\Delta R}_u = \mathbb{E}\left[R_u - R_u^{\mathrm{Q}}\right]$*, is upper bounded by*

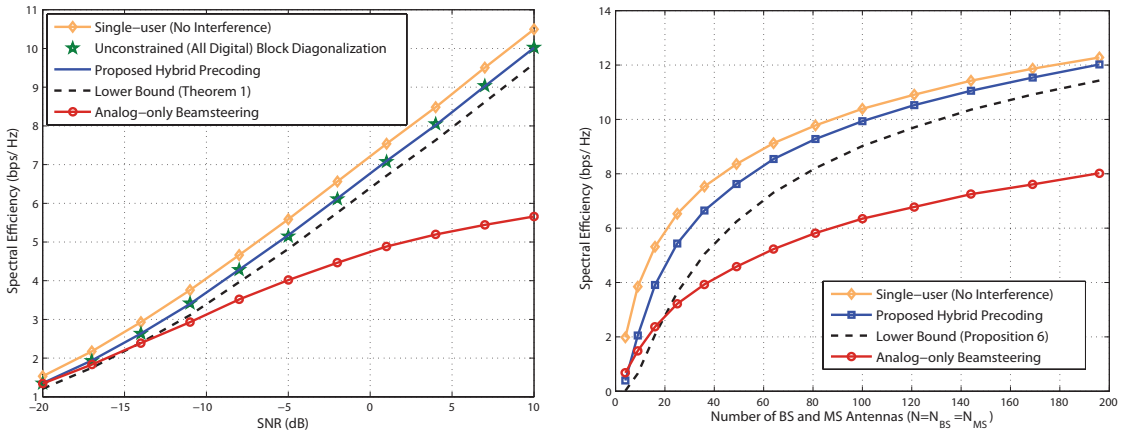
$$\overline{\Delta R}_u \leq \log_2\left(1 + \frac{\mathsf{SNR}}{U}\bar{\alpha} N_{\mathrm{BS}} N_{\mathrm{MS}} \left(1 + \frac{U-1}{N_{\mathrm{BS}}}\left(1 + \frac{L-1}{N_{\mathrm{BS}} N_{\mathrm{MS}}}\right)\right) 2^{-\frac{B_{\mathrm{BB}}}{U-1}}\right). \quad (3.32)$$

The proof is similar to Theorem 10, but leverages the definition of the effective channel in (3.20). In addition to characterizing the rate loss due to quantization for more general settings with multi-path mmWave channels, this result illustrates the marginal impact of the other paths on the performance of mmWave systems as $\frac{L-1}{N_{\mathrm{BS}} N_{\mathrm{MS}}} \ll 1$. In other words, this indicates that considering only the path with the maximum gain gives a very good performance.

3.8 Simulation Results

In this section, we evaluate the performance of the proposed hybrid analog/digital precoding algorithm and derived bounds using numerical simulations. All the plotted rates in Fig. 3.3-Fig. 3.7 are the averaged achievable rates per user; $\mathbb{E}\left[\frac{1}{U}\sum_{u=1}^{U} R_u\right]$ with R_u in equation 3.6.

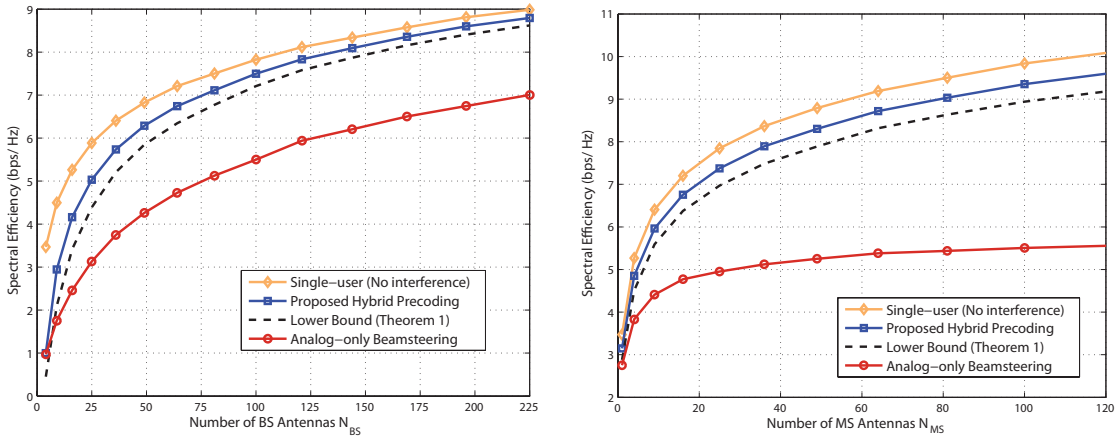
First, we compare the achievable rates without quantization loss and with perfect effective channel knowledge in Fig. 3.3(a) and Fig. 3.3(b). In Fig. 3.3(a), we consider the system model in Section 3.3 with a BS employing an 8×8 UPA with 4 MS's, each having a 4×4 UPA. The channels are single-path, the azimuth AoAs/AoDs are assumed to be uniformly distributed in $[0, 2\pi]$, and the elevation



(a) (b)

Figure 3.3: Achievable rates using the hybrid precoding and beamsteering algorithms with perfect channel knowledge. Single-path channels are assumed in (a), while channels with $L = 3$ paths are examined in (b).

AoAs/AoDs are uniformly distributed in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. The SNR in the plots is defined as $\text{SNR} = \frac{P\bar{\alpha}}{\sigma^2 U}$. The rate achieved by the proposed hybrid precoding/combining algorithm is compared with the single-user rate and the rate obtained by beamsteering. These rates are also compared with the performance of a particular unconstrained (all digital) block diagonalization algorithm in [67, Section III], where the beamforming and combining vectors of user u are selected to be $\mathbf{f}_u = \tilde{\mathbf{V}}_u^{(0)} \mathbf{v}_u$ and $\mathbf{w}_u = \mathbf{u}_u$, where $\mathbf{v}_u$ and $\mathbf{u}_u$ are the dominant right and left singular vectors of the effective channel matrix $\mathbf{H}_u \tilde{\mathbf{V}}_u^{(0)}$, with $\tilde{\mathbf{V}}_u^{(0)}$ an orthogonal basis for the null space of the matrix $\left[\mathbf{H}_1^T \ ... \ \mathbf{H}_{u-1}^T \mathbf{H}_{u+1}^T \ ... \ \mathbf{H}_U^T\right]^T$. This block diagonalization algorithm requires $N_{\text{BS}} - \text{rank}\left(\left[\mathbf{H}_1^T \ ... \ \mathbf{H}_{u-1}^T \mathbf{H}_{u+1}^T \ ... \ \mathbf{H}_U^T\right]^T\right) > 0, \ \forall u$ which is expected to be satisfied with high probability in mmWave systems with large arrays and sparse (low-rank) channels. Note that other block diagonalization algorithms, like coordinated Tx-Rx



(a) (b)

Figure 3.4: Achievable rates using the hybrid precoding and beamsteering algorithms with perfect channel knowledge. In (a), the performance of hybrid precoding is shown to approach the single-user rate with large numbers of BS antennas. In (b), the performance gap between hybrid precoding and beamsteering increases with more MS antennas.

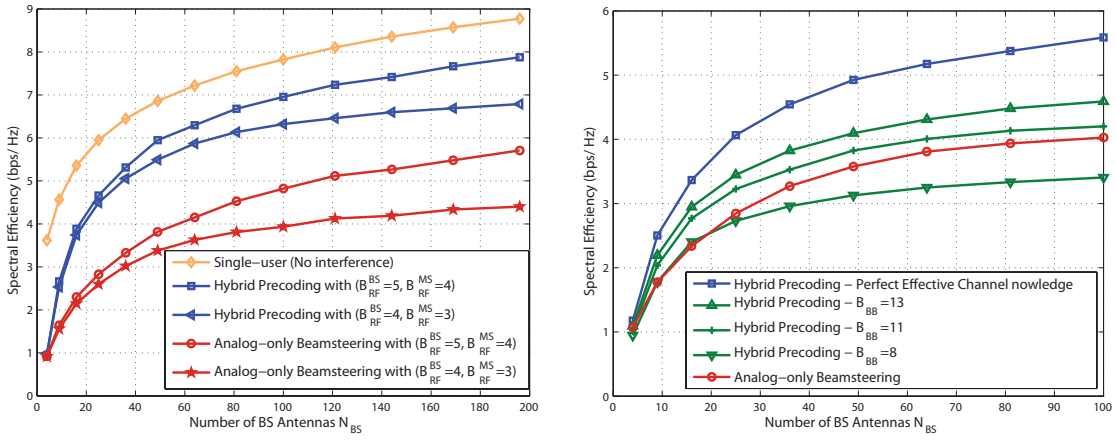
block diagonalization [67], may have more relaxed dimension constraints. The figure indicates that the performance of hybrid precoding is very close to the single-user rate thanks to canceling the residual multi-user interference, and is almost similar to the performance of the unconstrained block diagonalization. Note also that the gain of any other unconstrained precoding solution over the proposed hybrid precoding is expected to be small given the small gap between the hybrid precoding solution and the single user upper bound, which is also a bound for any other unconstrained precoding solution. The figure also illustrates the gain of hybrid precoding over analog-only beamsteering solution which increases with SNR as the beamsteering rate starts to be interference limited. The tightness of the derived lower bound in Theorem 4 is also shown.

In Fig. 3.3(b), we consider the same setup, but when each channel has $L = 3$ paths. The rates of the single-user, hybrid precoding, and beamsteering are simulated with different numbers of BS and MS antennas, assuming that $N_{\mathrm{BS}} = N_{\mathrm{MS}}$. The bound derived in Proposition 9 was also plotted where it is shown to be tight at large number of antennas as discussed in Section 3.6.2.

In Fig. 3.4(a), the same setup in Fig. 3.3(a) is considered at SNR= 0 dB, but with different values of BS antennas. The figure shows that even at very large numbers of antennas, there is still a considerable gain of hybrid precoding over beamsteering. This figure also shows that the difference between hybrid precoding and the single-user rate decreases at a large number of BS antennas which validates the second part of Proposition 7.

In Fig. 3.4(b), the same setup is considered with an 8×8 BS UPA and with different numbers of MS antennas. The figure illustrates how the performance gap between hybrid precoding and beamsteering increases with increasing the number of MS antennas which coincides with Corollary 8. This means that hybrid precoding has a higher gain over analog-only beamforming solutions in mmWave systems when large antenna arrays are employed at the MS's.

To illustrate the impact of RF quantization, the performance of hybrid precoding and analog-only beamsteering are evaluated in Fig. 3.5(a) with different numbers of quantization bits at the BS and MS. We consider the same setup of Fig. 3.3(a) with 4×4 MS UPAs and when each channel has $L = 3$ paths. As shown in the figure, the performance of the beamforming strategies degrades with decreasing the number of quantization bits. The gain, however, stays almost constant for the same number of



(a) (b)

Figure 3.5: Achievable rates using the hybrid precoding and beamsteering algorithms are plotted for different numbers of RF beamforming quantization bits in (a), and for different numbers of effective channel quantization bits in (b).

antennas. The figure also shows that the number of quantization bits should increase with the antenna numbers to avoid significant performance degradations.

In Fig. 3.5(b), the case when both RF and baseband quantized codebooks exist is illustrated. For this figure, the same system setup of Fig. 3.5(a) is adopted again, and the spectral efficiency achieved by hybrid precoding is shown for different sizes of the RVQ codebook used in quantizing the effective channels. The RF codebooks are also quantized with $B_{\mathrm{RF}}^{\mathrm{BS}} = 3$ bits and $B_{\mathrm{RF}}^{\mathrm{MS}} = 2$ bits. These results show that when the effective channel is poorly quantized, the loss of multi-user interference management is larger than its gain, and using analog-only beamsteering achieves better rates. For reasonable numbers of effective channel quantization bits, however, the performance of hybrid precoding maintains its gain over the described analog-only solutions.

In Fig. 3.6, we evaluate the performance of the proposed two-stage hybrid

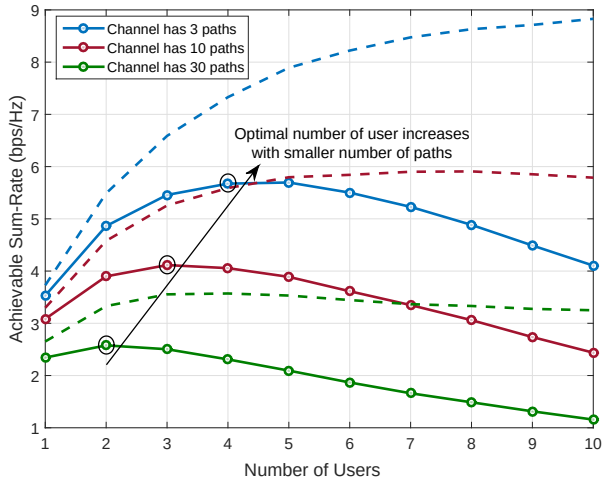


Figure 3.6: Performance of the multi-user hybrid precoding versus the number of users, for different numbers of channel paths. The users are assumed to be moving with speed 30m/s.

precoding algorithm in the presence of user mobility. In this simulation, we consider the system model in Section 3.3 with $N_{\mathrm{BS}} = 64$ antennas and $N_{\mathrm{MS}} = 8$ antennas. The users' channels follow the geometric channel model in (3.4), but with each path having an additional exponential term that captures the mobility(Doppler) effect, as shown in [80]. The users are assumed to be moving with speed 30 m/s on a distance 50m away from the serving BS. In the beginning of every beam coherence time, the BS and MSs perform beam training to find the best beams, then use these beams for the rest of the beam coherence time. In Fig. 3.6, the beam coherence time is assumed to be .1 s, based on the results in [80]. This figure shows that the optimal number of users that can be simultaneously served depends on the number of channel paths. As the channel becomes more sparse, i.e., with less number of paths, more users can be served to optimize the system sum-rate.

Finally, Fig. 3.7 evaluates the performance of the proposed hybrid precoding algorithm in a mmWave cellular setup including inter-cell interference, which is not explicitly incorporated into our designs. In this setup, BS's and MS's are assumed to be spatially distributed according to a Poisson point process with MS's densities 30 times the BS densities. The channels between the BS's and MS's are single-path and each link is determined to be line-of-sight or non-line-of-sight based on the blockage model in [7]. Each MS is associated to the BS with less path-loss and the BS randomly selects $n = 2, .., 5$ users of those associated to it to be simultaneously served. BS's are assumed to have 8×8 UPAs and MS's are equipped with 4×4 UPAs. All UPA's are vertical, elevation angles are assumed to be fixed at $\pi/2$, and azimuth angles are uniformly distributed in $[0, 2\pi]$. Fig. 3.7 shows the per-user coverage probability, defined as $\mathcal{P}\left(\mathrm{R_u} \geq \eta\right)$, where η is an arbitrary threshold. This figure illustrates that hybrid precoding has a reasonable coverage gain over analog-only beamsteering thanks to its interference management capability.

3.9 Conclusions

In this chapter, we proposed a low-complexity hybrid analog/digital precoding algorithm for downlink multi-user mmWave systems leveraging the sparse nature of the channel and the large number of deployed antennas. The performance of the proposed algorithm was analyzed when the channels are single-path and when the system dimensions are very large. In these cases, the asymptotic optimality of the proposed algorithm, and the gain over beamsteering solutions were illustrated. The results indicate that interference management in multi-user mmWave systems is

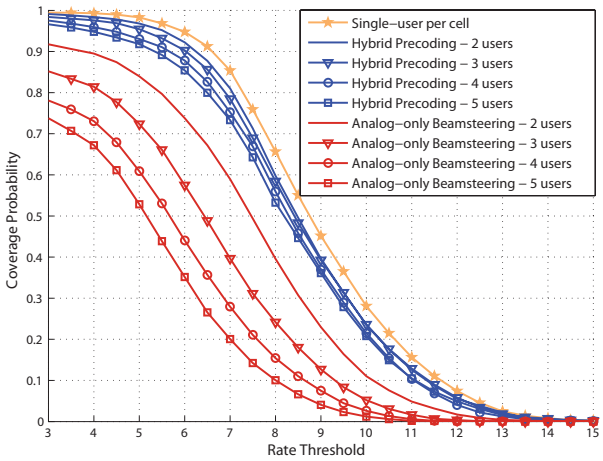


Figure 3.7: Coverage probability of the proposed hybrid precoding algorithm compared with single-user per cell and analog-only beamsteering solutions. The figure shows the per-user performance with different numbers of users per cell.

required even when the number of antennas is very large. When the feedback channels are limited, the average rate loss due to joint analog/digital codebook quantization was analyzed and numerically simulated. These simulations show that the hybrid precoding gain is not very sensitive to RF angles quantization. It is important, however, to have a good quantization for the digital precoding layer to maintain a reasonable precoding gain over analog only solutions.

Chapter 4

Multi-Layer Precoding for Massive MIMO Systems

4.1 Overview

In this chapter[1], we propose a general precoding framework, called multi-layer precoding, to enable efficient and low complexity full-dimensional massive MIMO operation. Multi-layer precoding (i) leverages the directional characteristics of large-scale MIMO channels to manage inter-cell interference with low channel knowledge requirements, and (ii) allows for an efficient implementation using low-complexity hybrid analog/digital architectures. We present a specific multi-layer precoding design for full-dimensional massive MIMO systems. The performance of this precoding design is analyzed and the per-user achievable rate is characterized for general channel models. The asymptotic optimality of the proposed multi-layer precoding design is then proved for some special yet important channel models. Numerical simulations verify the analytical results and illustrate the potential gains of multi-layer precoding compared to traditional pilot-contaminated massive MIMO solutions.

[1]This chapter is based on the work that is to be published in the journal paper: A. Alkhateeb, G. Leus, and R. Heath, "Multi-layer Precoding: A Potential Solution for Full-Dimensional Massive MIMO Systems,". This work was supervised by Prof. Robert Heath. Prof. Geert Leus provided important ideas for the multi-layer precoding design that greatly improved the work.

4.2 Introduction

Massive MIMO promises significant spectral efficiency gains for cellular systems. Scaling up the number of antennas, however, faces a number of challenges that prevent the corresponding scaling of the gains [2, 4, 22, 81]. The large-dimensional channels have high feedback overhead in frequency division duplexing (FDD) systems. To overcome that, channel reciprocity in conjunction with time division duplexing (TDD) systems is used alternatively [1,82]. Reusing the uplink training pilots among cells, however, causes channel estimation errors which in turn lead to downlink inter-cell interference, especially for cell-edge users [1]. Handling inter-cell interference using traditional network MIMO techniques requires high coordination overhead, which limits the overall system performance [25]. Another challenge with the large number of antennas lies in the hardware implementation [4, 12]. Traditional MIMO precoding techniques normally assumes entire baseband processing, which requires dedicating an RF chain per antenna. This may lead to high cost and power consumption in massive MIMO systems [4]. Therefore, developing precoding schemes that can overcome the challenges of inter-cell interference and complete baseband processing is of great interest.

Inter-cell interference is a critical problem for cellular MIMO systems. Typical network MIMO solutions for managing this interference require some sort of collaboration between the base stations (BSs) [26]. The overhead of this cooperation, though, fundamentally limits the system performance [25]. When the number of antennas grows to infinity, the performance of the network becomes limited by pilot contamination [1], which is one form of inter-cell interference. Pilot contamination

happens because of the channel estimation errors that result from reusing the uplink training pilots among users in TDD massive MIMO systems. Several solutions have been proposed to manage inter-cell interference in massive MIMO systems [83–86]. In [83,84], multi-cell zero-forcing and MMSE MIMO precoding strategies were developed to cancel or reduce the inter-cell interference. The solutions in [83,84], however, require global channel knowledge at every BS, which makes them feasible only for finite numbers of antennas [87]. To reduce the channel requirement overhead, pilot contamination precoding was proposed in [85] to overcome the pilot contamination problem, relying on the large channel statistics. The technique in [85], though, requires sharing the transmitted messages between all BSs, which is difficult to achieve in practice. In [86], the directional characteristics of large-dimensional channels were leveraged to improve the uplink channel training in TDD systems. This solution, however, requires fully-digital hardware and does not leverage the higher degrees of freedom provided in full-dimensional massive MIMO systems.

On a relevant research direction, precoding solutions that divide the processing between two stages have been developed in [20,23,30,88,89] for mmWave and massive MIMO systems. Motivated by the high cost and power consumption of RF chains, [20] developed hybrid analog/digital precoding algorithm for mmWave systems. Hybrid precoding divides the precoding between RF and baseband domains, and requires a much smaller number of RF chains compared to the number of antennas. For multi-user systems [30] proposed a two-stage hybrid precoding design where the first precoding matrix is designed to maximize the signal power for each user and the second matrix is designed to manage the multi-user interference. Similar solutions

were also developed for massive MIMO systems [88, 89], with the general objective of maximizing the system sum-rate. In [23], a two-stage joint spatial division and multiplexing (JSDM) precoding scheme was developed to reduce the channel training overhead in FDD massive MIMO systems. In JSDM, the base station (BS) divides the mobile stations (MSs) into groups of approximately similar covariance eigenspaces, and designs a pre-beamforming matrix based on the large channel statistics. The interference between the users of each group is then managed using another precoding matrix given the effective reduced-dimension channels. The work in [20, 23, 30, 88, 89], however, did not consider out-of-cell interference, which ultimately limits the performance of massive MIMO systems.

In this work, we introduce a general framework, called multi-layer precoding, that (i) coordinates inter-cell interference in massive MIMO systems leveraging large channel statistics and (ii) allows for efficient implementations using hybrid analog/digital architectures. The main contributions of our work are summarized as follows.

- Designing a specific multi-layer precoding solution for full-dimensional massive MIMO systems. The proposed precoding strategy writes the precoding matrix of each BS as a multiplication of three precoding matrices, called layers. The three precoding layers are designed to avoid inter-cell interference, maximize effective signal power, and manage intra-cell multi-user interference, with low channel training overhead.

- Analyzing the performance of the proposed multi-layer precoding design. First,

the per-user achievable rate using multi-layer precoding is derived for a general channel model. Then, asymptotic optimality results for the achievable rates with multi-layer precoding are derived for two special channel models: the one-ring and the single-path models. Lower bounds on the achievable rates for the cell-edge users are also characterized under the one-ring channel model.

The developed multi-layer precoding solution and other proposed extensions are also evaluated by numerical simulations. Results show the multi-layer precoding can approach the single-user rate, which is free of inter-cell and intra-cell interference, in some special cases. Further, results illustrate that significant rate and coverage gains can be obtained by multi-layer precoding compared to conventional conjugate beamforming and zero-forcing massive MIMO solutions.

4.3 System and Channel Models

In this section, we present the full-dimensional massive MIMO system and channel models adopted in the paper.

4.3.1 System Model

Consider a cellular system model consisting of B cells with one BS and K MS's in each cell, as shown in Fig. 4.1. Each BS is equipped with a two-dimensional (2D) antenna array of N elements, $N = N_{\mathrm{V}}$ (vertical antennas) $\times N_{\mathrm{H}}$ (horizontal antennas), and each MS has a single antenna. We assume that all BSs and MSs are synchronized and operate a TDD protocol with universal frequency reuse. In the downlink, each BS b, $b = 1, 2, ..., B$, applies an $N \times K$ precoder $\mathbf{F}_b$ to transmit a

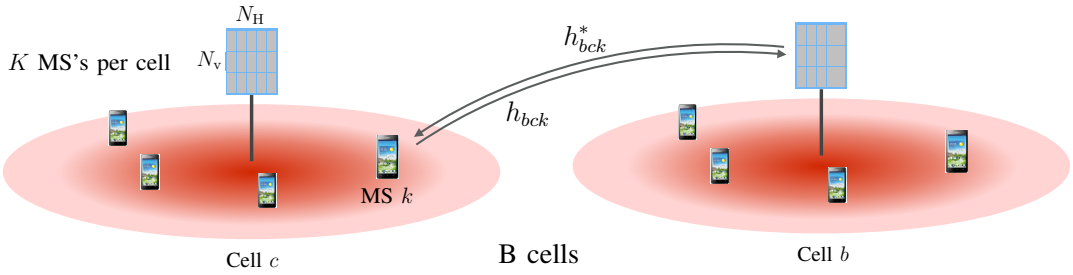


Figure 4.1: A full-dimensional MIMO cellular model where each BS has a 2D antenna array and serves K users.

symbol for each user, with a power constraint $\| [\mathbf{F}_b]_{:,k} \|^2 = 1$, $k = 1, 2, ..., K$. Uplink and downlink channels are assumed to be reciprocal. If $\mathbf{h}_{bck}$ denotes the $N \times 1$ uplink channel from user k in cell c to BS b, then the received signal by this user in the downlink can be written as

$$y_{ck} = \sum_{b=1}^{B} \mathbf{h}^*_{bck} \mathbf{F}_b \mathbf{s}_b + n_{ck}, \tag{4.1}$$

where $\mathbf{s}_b = [s_{b,1}, ..., s_{b,K}]^T$ is the $K \times 1$ vector of transmitted symbols from BS b, such that $\mathbb{E}[\mathbf{s}_b \mathbf{s}_b^*] = \frac{P}{K}\mathbf{I}$, with P representing the average total transmitted power, and $n_{ck} \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian noise at user k in cell c. It is useful to expand (4.1) as

$$y_{ck} = \underbrace{\mathbf{h}^*_{cck} [\mathbf{F}_c]_{:,k} s_{c,k}}_{\text{Desired signal}} + \underbrace{\sum_{m \neq k} \mathbf{h}^*_{cck} [\mathbf{F}_c]_{:,m} s_{c,m}}_{\text{Intra-cell interference}} + \underbrace{\sum_{b \neq c} \mathbf{h}^*_{bck} \mathbf{F}_b \mathbf{s}_b}_{\text{Inter-cell interference}} + n_{ck}, \tag{4.2}$$

to illustrate the different components of the received signal.

4.3.2 Channel Model

We consider a full-dimensional MIMO configuration where 2D antenna arrays are deployed at the BS's. Consequently, the channels from the BS's to each

user have a 3D structure. Extensive efforts are currently given to 3D channel measurements, modeling, and standardization [90, 91]. One candidate is the Kronecker product correlation model, which provides a reasonable approximation to 3D covariance matrices [92]. In this model, the covariance of the 3D channel $\mathbf{h}_{bck}$, defined as $\mathbf{R}_{bck} = \mathbb{E}\left[\mathbf{h}_{bck}\mathbf{h}_{bck}^{*}\right]$, is approximated by

$$\mathbf{R}_{bck} = \mathbf{R}_{bck}^{\mathrm{A}} \otimes \mathbf{R}_{bck}^{\mathrm{E}}, \tag{4.3}$$

where $\mathbf{R}_{bck}^{\mathrm{A}}$ and $\mathbf{R}_{bck}^{\mathrm{E}}$ represent the covariance matrices in the azimuth and elevation directions. If $\mathbf{R}_{bck}^{\mathrm{A}} = \mathbf{U}_{bck}^{\mathrm{A}}\mathbf{\Lambda}_{bck}^{\mathrm{A}}{\mathbf{U}_{bck}^{\mathrm{A}}}^{*}$ and $\mathbf{R}_{bck}^{\mathrm{E}} = \mathbf{U}_{bck}^{\mathrm{E}}\mathbf{\Lambda}_{bck}^{\mathrm{E}}{\mathbf{U}_{bck}^{\mathrm{E}}}^{*}$ are the eigenvalue decompositions of $\mathbf{R}_{bck}^{\mathrm{A}}$ and $\mathbf{R}_{bck}^{\mathrm{E}}$, then using Karhunen-Loeve representation, the channel $\mathbf{h}_{bck}$ can be expressed as

$$\mathbf{h}_{bck} = \left[\mathbf{U}_{bck}^{\mathrm{A}}{\mathbf{\Lambda}_{bck}^{\mathrm{A}}}^{\frac{1}{2}} \otimes \mathbf{U}_{bck}^{\mathrm{E}}{\mathbf{\Lambda}_{bck}^{\mathrm{E}}}^{\frac{1}{2}}\right]\mathbf{w}_{bck}, \tag{4.4}$$

where $\mathbf{w}_{bck} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a $\mathrm{rank}\left(\mathbf{R}_{bck}^{\mathrm{A}}\right)\mathrm{rank}\left(\mathbf{R}_{bck}^{\mathrm{E}}\right) \times 1$ vector, with $\mathrm{rank}(\mathbf{A})$ representing the rank of the matrix $\mathbf{A}$. Without loss of generality, and to simplify the notations, we assume that all the users have the same ranks for the azimuth and elevation covariance matrices, which are denoted as r_{A} and r_{E}.

4.4 Multi-Layer Precoding: The General Concept

In this section, we briefly introduce the motivation and general concept of multi-layer precoding. Given the system model in Section 4.3, the signal-to-interference-

plus-noise ratio (SINR) at user k in cell c is

$$\text{SINR}_{ck} = \frac{\frac{P}{K}\left|\mathbf{h}_{cck}^{*}\left[\mathbf{F}_{c}\right]_{:,k}\right|^{2}}{\frac{P}{K}\sum_{m\neq k}|\mathbf{h}_{cck}^{*}\left[\mathbf{F}_{c}\right]_{:,m}|^{2}+\frac{P}{K}\sum_{b\neq c}\|\mathbf{h}_{bck}^{*}\mathbf{F}_{b}\|^{2}+\sigma^{2}}, \tag{4.5}$$

where the terms $\left|\mathbf{h}_{cck}^{*}\left[\mathbf{F}_{c}\right]_{:,k}\right|^{2}$, $\sum_{m\neq k}|\mathbf{h}_{cck}^{*}\left[\mathbf{F}_{c}\right]_{:,m}|^{2}$, and $\sum_{b\neq c}\|\mathbf{h}_{bck}^{*}\mathbf{F}_{b}\|^{2}$ represent the desired signal power, intra-cell multi-user interference, and inter-cell interference, respectively. Designing one precoding matrix per BS, $\mathbf{F}_c$, to manage all these kinds of signals by, for example, maximizing the system sum-rate is non-trivial. This normally leads to a non-convex problem whose closed-form solution is unknown [26]. Also, coordinating inter-cell interference between BS's typically results in high cooperation overhead that makes the value of this cooperation limited [25]. Another challenge lies in the entire baseband implementation of these precoding matrices, which may yield high cost and power consumption in massive MIMO systems [4].

Our objective is to design the precoding matrices, $\mathbf{F}_b$, $b = 1, 2, ..., B$, such that (i) they manage the inter-cell and intra-cell interference with low requirements on the channel knowledge, and (ii) they can be implemented using low-complexity hybrid analog/digital architectures [30], i.e., with a small number of RF chains. Next, we present the main idea of multi-layer precoding, a potential solution to achieve these objectives.

Inspired by prior work on multi-user hybrid precoding [30] and joint spatial division multiplexing [23], and leveraging the directional characteristics of large-scale MIMO channels [86], we propose to design the precoding matrix $\mathbf{F}_c$ as a product of a number of precoding matrices (layers). In this paper, we will consider a 3-layer

precoding matrix

$$\mathbf{F}_c = \mathbf{F}_c^{(1)}\mathbf{F}_c^{(2)}\mathbf{F}_c^{(3)}, \tag{4.6}$$

where these precoding layers are designed according to the following criteria.

- **One precoding objective per layer:** Each layer is designed to achieve only one precoding objective, e.g., maximizing desired signal power, minimizing inter-cell interference, or minimizing multi-user interference. This simplifies the precoding design problem and divides it into easier and/or convex sub-problems. Further, this decouples the required channel knowledge for each layer.

- **Successive dimensionality reduction:** Each layer should be properly designed such that the effective channel, including this layer, has smaller dimensions compared to the original channel. This reduces the channel training overhead of every precoding layer compared to the previous one. Further, this makes a successive reduction in the dimensions of the precoding matrices, which eases implementing them using hybrid analog/digital architectures [4, 20, 30, 93] with small number of RF chains.

- **Different channel statistics:** These precoding objectives are distributed over the precoding layers such that $\mathbf{F}_c^{(1)}$ requires slower time-varying channel state information compared with $\mathbf{F}_c^{(2)}$, which in turn requires slower channel state information compared with $\mathbf{F}_c^{(3)}$. Given the successive dimensionality reduction criteria, this means that the first precoding layer, which needs to be designed based on the large channel matrix, requires very large-scale channel statistics and needs to be updated every very long period of time. Similarly, the second

and third precoding layers, which are designed based on the effective channels that have less dimensions, need to be updated more frequently.

In the next sections, we will present a specific multi-layer precoding design for full-dimensional massive MIMO systems, and show how it enables leveraging the large-scale MIMO channel characteristics to manage different kinds of interference with limited channel knowledge. We will also show how the multiplicative and successive reduced dimension structure of multi-layer precoding allows for efficient implementations using hybrid analog/digital architectures.

4.5 Proposed Multi-Layer Precoding Design

In this section, we present a multi-layer precoding algorithm for the full-dimensional massive MIMO system and channel models described in Section 4.3. Following the proposed multi-layer precoding criteria explained in Section 4.4, we propose to design the $N_{\mathrm{v}}N_{\mathrm{H}} \times K$ precoding matrix $\mathbf{F}_b$ of cell b, $b = 1, ..., B$ as

$$\mathbf{F}_b = \mathbf{F}_b^{(1)}\mathbf{F}_b^{(2)}\mathbf{F}_b^{(3)}, \tag{4.7}$$

where the first precoding layer $\mathbf{F}_b^{(1)}$ is dedicated to avoid the out-of-cell interference, the second precoding layer $\mathbf{F}_b^{(2)}$ is designed to maximize the effective signal power, and the third layer $\mathbf{F}_b^{(3)}$ is responsible of canceling the intra-cell multi-user interference. Writing the received signal at user k in cell c in terms of the multi-layer precoding in (4.7), we get

$$y_{ck} = \underbrace{\mathbf{h}_{cck}^*\mathbf{F}_c^{(1)}\mathbf{F}_c^{(2)}\mathbf{F}_c^{(3)}\mathbf{s}_c}_{\text{received signal from serving BS}} + \underbrace{\sum_{b\neq c}\mathbf{h}_{bck}^*\mathbf{F}_b^{(1)}\mathbf{F}_b^{(2)}\mathbf{F}_b^{(3)}\mathbf{s}_b}_{\text{received signal from other BSs}} + n_{ck}. \tag{4.8}$$

Next, we explain in detail the proposed design of each precoding layer as well as the required channel knowledge.

4.5.1 First Layer: Inter-Cell Interference Management

We will design the first precoding layer $\mathbf{F}_b{}^{(1)}$ to avoid the inter-cell interference, i.e., to cancel the second term of (4.8). Exploiting the Kronecker structure of the channel model in (4.4), we propose to construct $\mathbf{F}_b^{(1)}$ as

$$\mathbf{F}_b^{(1)} = \mathbf{F}_b^{\mathrm{A}(1)} \otimes \mathbf{F}_b^{\mathrm{E}(1)}. \tag{4.9}$$

Adopting the channel model in (4.4) with $\overline{\mathbf{w}}_{bck} = \left(\mathbf{\Lambda}_{bck}^{\mathrm{A}\ \frac{1}{2}} \otimes \mathbf{\Lambda}_{bck}^{\mathrm{E}\ \frac{1}{2}}\right) \mathbf{w}_{bck}$ and employing the Kronecker precoding structure in (4.9), the second term of the received signal y_{ck} in (4.8) can be expanded as

$$\sum_{b\neq c} \mathbf{h}_{bck}^* \mathbf{F}_b^{(1)} \mathbf{F}_b^{(2)} \mathbf{F}_b^{(3)} \mathbf{s}_b = \sum_{b\neq c} \overline{\mathbf{w}}_{bck}^* \left(\mathbf{U}_{bck}^* \mathbf{F}_b^{\mathrm{A}(1)} \otimes \mathbf{U}_{bck}^{\mathrm{E}^*} \mathbf{F}_b^{\mathrm{A}(1)}\right) \mathbf{F}_b^{(2)} \mathbf{F}_b^{(3)} \mathbf{s}_b. \tag{4.10}$$

Avoiding the inter-cell interference for the users at cell c can then be satisfied if $\mathbf{F}_b^{(1)}, b \neq c$ is designed such that $\mathbf{U}_{bck}^{\mathrm{E}^*} \mathbf{F}_b^{\mathrm{E}(1)} = \mathbf{0}, \forall k$. Equivalently, for any cell c to avoid making interference on the other cell users, it designs its precoder $\mathbf{F}_c^{\mathrm{E}(1)}$ to be in the null-space of the elevation covariance matrices of all the channels connecting BS c and the other cell users, i.e., to be in Null $\left(\sum_{b\neq c} \sum_{k\in\mathcal{K}_b} \mathbf{R}_{cbk}^{\mathrm{E}}\right)$ with $\mathcal{K}_b$ denoting the subset of K scheduled users in cell b.

Thanks to the directional structure of large-scale MIMO channels, we note that with a large number of vertical antennas, N_{V}, the null-space Null $\left(\sum_{b\neq c} \sum_{k\in\mathcal{K}_b} \mathbf{R}_{cbk}^{\mathrm{E}}\right)$ of different scheduled users $\mathcal{K}_b$ will have a large overlap. This means that designing

$\mathbf{F}_c^{\mathrm{E}(1)}$ based on the interference covariance subspace averaged over different scheduled users may be sufficient. Leveraging this intuition relaxes the required channel knowledge to design the first precoding layer. Hence, we define the average interference covariance matrix for BS c as

$$\mathbf{R}_c^{\mathrm{I}} = \sum_{b \neq c} \mathbb{E}_{\mathcal{K}_b} \left[\mathbf{R}_{cbk}^{E} \right]. \tag{4.11}$$

Let $\left[\mathbf{U}_c^{\mathrm{I}}\ \mathbf{U}_c^{\mathrm{NI}}\right] \mathbf{\Lambda}_c \left[\mathbf{U}_c^{\mathrm{I}} \mathbf{U}_c^{\mathrm{NI}}\right]^*$ represent the eigen-decomposition of $\mathbf{R}_c^{\mathrm{I}}$ with the $N_{\mathrm{v}} \times r_{\mathrm{I}}$ matrix $\mathbf{U}_c^{\mathrm{I}}$ and $N_{\mathrm{v}} \times r_{\mathrm{NI}}$ matrix $\mathbf{U}_c^{\mathrm{NI}}$ corresponding to the non-zero and zero eigenvalues, respectively. Then, we design the first precoding layer $\mathbf{F}_c^{(1)}$ to be in the null-space of the average interference covariance matrix by setting

$$\mathbf{F}_c^{(1)} = \mathbf{I}_{N_{\mathrm{H}}} \otimes \mathbf{U}_c^{\mathrm{NI}}, \tag{4.12}$$

which is an $N_{\mathrm{V}} N_{\mathrm{H}} \times r_{\mathrm{NI}} N_{\mathrm{H}}$ matrix.

Given the design of the first precoding layer in (4.12), and defining the $r_{\mathrm{NI}} \times r_{\mathrm{E}}$ effective elevation eigen matrix $\overline{\mathbf{U}}_{cck}^{\mathrm{E}} = \mathbf{U}_c^{\mathrm{NI}*} \mathbf{U}_{cck}^{\mathrm{E}}$, the received signal at user k of cell c in (4.8) becomes

$$y_{ck} = \overline{\mathbf{w}}_{cck}^{*} \left(\mathbf{U}_{cck}^{\mathrm{A}^*} \otimes \overline{\mathbf{U}}_{cck}^{\mathrm{E}^*} \right) \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} \mathbf{s}_c + n_{ck}. \tag{4.13}$$

Note that the first precoding layer in (4.9) acts as a spatial filter that entirely eliminates the inter-cell interference in the elevation domain. This filter, however, may have a negative impact on the desired signal power for the served users at cell c if they share the same elevation subspace with the out-of-cell users. Therefore, this first layer precoding design is particularly useful for systems with low-rank elevation subspaces. It is worth mentioning here that recent measurements of 3D channels show

that elevation eigenspaces may have low ranks at both low-frequency and millimeter wave systems [9, 38, 94]. Relaxations of the precoding design in (4.9) are proposed in Section 4.7 to compromise between inter-cell interference avoidance and desired signal power degradation.

Required channel knowledge: The design of the first precoding layer in (4.9) requires only the knowledge of the interference covariance matrix averaged over different scheduled users. It depends therefore on very large time-scale channel statistics, which means that this precoding layer needs to be updated every very long period of time. This makes its acquisition overhead relatively negligible from an overall system perspective. In fact, this is a key advantage of the decoupled multi-layer precoding structure that allows dedicating one layer for canceling the out-of-cell interference based on very large time-scale channel statistics while leaving the other layers to do other functions based on different time scales. This can not be done by typical precoding schemes that relies on one precoding matrix to manage different precoding objectives, as this precoding matrix will likely need to be updated based on the fastest channel statistics.

4.5.2 Second Layer: Desired Signal Beamforming

The second precoding layer $\mathbf{F}_c{}^{(2)}$ is designed to focus the transmitted power on the served users' effective subspaces, i.e., on the user channels' subspaces including the effect of the first precoding layer. If we define the matrix consisting of the effective eigenvectors of user k in cell c as $\overline{\mathbf{U}}_{cck} = \left(\mathbf{U}^{\mathrm{A}}_{cck} \otimes \overline{\mathbf{U}}^{\mathrm{E}}_{cck}\right)$, then we design the second

precoding layer $\mathbf{F}_c{}^{(2)}$ as a large-scale conjugate beamforming matrix, i.e., we set

$$\mathbf{F}_c{}^{(2)} = \left[\overline{\mathbf{U}}_{cc1}, ..., \overline{\mathbf{U}}_{ccK}\right], \tag{4.14}$$

which has $N_{\mathrm{H}} r_{\mathrm{NI}} \times K r_{\mathrm{A}} r_{\mathrm{E}}$ dimensions.

Given the second precoding layer design, and defining $\mathbf{G}_{c,(k,r)} = \overline{\mathbf{U}}_{cck}^{*} \overline{\mathbf{U}}_{ccr}$, the received signal by user k in cell c can be written as

$$y_{ck} = \overline{\mathbf{w}}_{cck}^{*} \left[\mathbf{G}_{c,(k,1)}, ..., \mathbf{G}_{c,(k,K)}\right] \mathbf{F}_c^{(3)} \mathbf{s}_c + n_{ck}. \tag{4.15}$$

The main objectives of this precoding layer can be summarized as follows. First, the effective channel vectors, including the first and second precoding layers, will have reduced dimensions compared to the original channels, especially when very large numbers of antennas are employed. This reduces the overhead associated with training the effective channels, which is particularly important for FDD systems [23, 30]. Second, this precoding layer supports the multiplicative structure of multi-layer precoding with successive dimensional reduction, which simplifies its implementation using hybrid analog/digital architectures, as will be briefly discussed in Section 4.7.

Required channel knowledge: The design of the second precoding layer requires only the knowledge of the effective eigenvector matrices $\overline{\mathbf{U}}_{cck}, k = 1, ..., K$, which depends on large-scale channel statistics. Further, it is worth noting that during the uplink training of the matrices $\overline{\mathbf{U}}_{cck}$, the first precoding layer works as spatial filtering for the other cell interference. Hence, this reduces (and ideally eliminates) the channel estimation error due to pilot reuse among cells, and consequently leads to a pilot decontamination effect.

4.5.3 Third Layer: Multi-User Interference Management

The third precoding layer $\mathbf{F}_c{}^{(3)}$ is designed to manage the multi-user interference based on the effective channels, i.e., including the effect of the first and second precoding layers. If we define the effective channel of user k in cell c as $\overline{h}_{ck} = \left[\mathbf{G}_{c,(k,1)}, ..., \mathbf{G}_{c,(k,K)}\right]^* \overline{\mathbf{w}}_{cck}$, and let $\overline{\mathbf{H}}_c = [\overline{h}_{c1}, ..., \overline{h}_{cK}]$, then we construct the third precoding layer $\mathbf{F}_c^{(3)}$ as a zero-forcing matrix

$$\mathbf{F}_c^{(3)} = \overline{\mathbf{H}}_c \left(\overline{\mathbf{H}}_c^* \overline{\mathbf{H}}_c\right)^{-1} \boldsymbol{\Upsilon}_c, \tag{4.16}$$

where $\boldsymbol{\Upsilon}_c$ is a diagonal power normalization matrix that ensures satisfying the precoding power constraint $\| [\mathbf{F}_b]_{:,k} \|^2 = 1$. Note that this zero-forcing design requires $N_{\mathrm{H}} r_{\mathrm{NI}} \geq K r_{\mathrm{A}} r_{\mathrm{E}}$, which is satisfied with high probability in massive MIMO systems, especially with sparse and low-rank channels. Given the design of the precoding matrix $\mathbf{F}_c^{(3)}$, the received signal at user k in cell c can be expressed as

$$y_{ck} = [\boldsymbol{\Upsilon}_c]_{k,k} \, s_{c,k} + n_{ck}. \tag{4.17}$$

Required channel knowledge: The design of the third precoding layer relies on the instantaneous effective channel knowledge. Thanks to the first and second precoding layers, these effective channels may have much smaller dimensions compared to the original channels in massive MIMO systems, which reduces the required training overhead.

4.6 Performance Analysis

The proposed multi-layer precoding design in Section 4.5 eliminates inter-cell interference as well as multi-user intra-cell interference. This interference cancellation,

however, may have a penalty on the desired signal power which is implicitly captured by the power normalization factor $[\boldsymbol{\Upsilon}_c]_{k,k}$ in (4.17). In this section, we will first characterize the achievable rate by the proposed multi-layer precoding design for a general channel model in Lemma 13. Then, we will show that this precoding design can achieve optimal performance for some special yet important channel models in Section 4.6.1 and Section 4.6.2.

Lemma 13 *Consider the system and channel models in Section 4.3 and the multi-layer precoding design in Section 4.5. The achievable rate by user k in cell c is given by*

$$R_{ck} = \log_2 \left(1 + \frac{\mathsf{SNR}}{\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1}_{k,k}} \right), \tag{4.18}$$

where $\mathbf{W}_c = \mathbf{I}_K \circ [\overline{\mathbf{w}}_{cc1}, ..., \overline{\mathbf{w}}_{ccK}]$ *and* $\mathsf{SNR} = \frac{P}{K\sigma^2}$.

Proof: See Appendix A □

Note that the achievable rate in (4.18) is upper bounded by the single-user rate—the rate when the user is solely served in the network—which is given by $\overline{R}_{ck} = \log_2 \left(1 + \mathsf{SNR} \left\| \overline{\mathbf{w}}_{cck} \right\|^2 \right)$. Therefore, Lemma 13 indicates that the proposed multi-layer precoding can achieve an optimal performance if $\mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} = \mathbf{I}$. To achieve that, it is sufficient to satisfy the following two conditions.

(i) $\mathbf{G}_{c,(k,m)} = \mathbf{0}, \forall m \neq k$, a condition that captures the impact of multi-user interference cancellation on the desired signal power.

(ii) $\mathbf{G}_{c,(k,k)} = \left(\mathbf{U}_{cck}^{\mathrm{A}\ *} \otimes \mathbf{U}_{cck}^{\mathrm{E}\ *}\right) \mathbf{F}_c^{(1)} \mathbf{F}_c^{(1)*} \left(\mathbf{U}_{cck}^{\mathrm{A}} \otimes \mathbf{U}_{cck}^{\mathrm{E}}\right) = \mathbf{I}, \forall k$, a condition that captures the possible impact of the inter-cell interference avoidance on the desired signal power.

Next, we characterize the performance of multi-layer precoding for two special yet important channel models, namely, the one-ring and single-path channel models.

4.6.1 Performance with One-Ring Channel Models

Motivated by its analytical tractability and meaningful geometrical interpretation, we will consider the one-ring channel model in this subsection [95–98]. This will enable us to draw useful insights into the performance of multi-layer precoding, which can then be extended to more general channel models. Note that due to its tractability, one-ring channel models have also been adopted in prior massive MIMO work [23, 86, 99, 100].

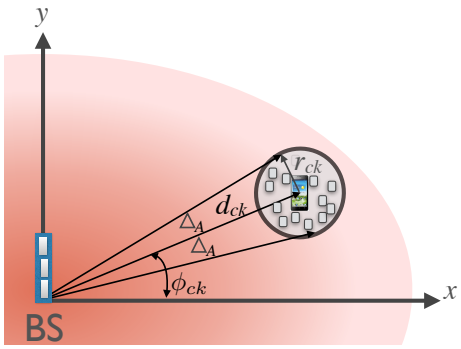


Figure 4.2: An illustration of the one-ring channel model in the azimuth direction. The BS that, has a UPA in the y-z plane, serves a mobile user in the x-y plane at distance d_{ck}. The user is surrounded by scatterers on a ring of radius r_{ck}, and its channel experience an azimuth angular spread Δ_{A}.

The one-ring channel model describes the case when a BS is elevated away from scatterers and is communicating with a mobile user that is surrounded by a ring of scatterers. Consider a BS at height $H_{\rm BS}$ employing an $N_V \times N_H$ UPA, and serving a mobile user at a distance d_{ck} with azimuth and elevation angles ϕ_{ck}, θ_{ck}, as depicted in Fig. 4.2. If the mobile is surrounded by scatterers on a ring of radius r_{ck} in the azimuth dimension, then the azimuth angular spread $\Delta_{\rm A}$ can be approximated as $\Delta_{\rm A} = \arctan\left(\frac{r_{ck}}{d_{ck}}\right)$. Further, assuming for simplicity that the received power is uniformly distributed over the ring, then the correlation between any two antenna elements with orders n_1, n_2 in the horizontal direction is given by

$$\left[\mathbf{R}^{A}_{cck}\right]_{n_1,n_2} = \frac{1}{2\Delta_{\rm A}} \int_{-\Delta_{\rm A}}^{\Delta_{\rm A}} e^{-j\frac{2\pi}{\lambda} d(n_2-n_1)\sin(\phi_{ck}+\alpha)\sin(\theta_{ck})} d\alpha. \tag{4.19}$$

The elevation correlation matrix can be similarly defined for the user k, in terms of its elevation angular spread $\Delta_{\rm E}$.

In the next theorem, we characterize the achievable rate for an arbitrary user k in cell c under the one-ring channel model.

Theorem 14 *Consider the full-dimensional cellular system model in Section 4.3.1 with cells of radius $r_{\rm cell}$, and the channel model in Section 4.3.2 with the one-ring correlation matrices in* (4.19)*. Let ϕ_{ck}, θ_{ck} denote the azimuth and elevation angles of user k at cell c, and let $\Delta_{\rm A}, \Delta_{\rm E}$ represent the azimuth and elevation angular spread. Define the maximum distance with no blockage on the desired signal power as $d_{\max} = H_{\rm BS} \tan\left(\arctan\left(\frac{r_{\rm cell}}{H_{\rm BS}}\right) - 2\Delta_{\rm E}\right)$. If $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_{\rm A}$ or $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_{\rm E}$, $\forall m \neq k$,*

and $d_{ck} \leq d_{\max}$, then the achievable rate of user k at cell c, when applying the multi-layer precoding algorithm in Section 4.5, satisfies

$$\lim_{N_{\mathrm{V}},N_{\mathrm{H}}\to\infty} R_{ck} = \overline{R}_{ck} = \log_2\left(1+\mathsf{SNR}\left\|\overline{\mathbf{w}}_{cck}\right\|^2\right). \tag{4.20}$$

Proof: See Appendix B □

Theorem 14 indicates that achievable rate with multi-layer precoding converges to the optimal single-user rate for the users that are not at the cell edge ($r_{\mathrm{cell}} - d_{\max}$ away from cell edge), provided that they maintain either an azimuth or elevation separation by double the angular spread. For example, consider a cellular system with cell radius 100m and BS antenna height 50m, if the elevation angular spread equals $\Delta_{\mathrm{E}} = 3^\circ$, then all the users within $\sim$ 80m distance from the BS achieves optimal rate. It is worth noting here that these rates do not experience any pilot contamination or multi-user interference impact and can, therefore, grow with the antenna numbers or transmit power without any bound on the maximum values that they can reach.

The angular separation between the users in Theorem 14 can be achieved via user scheduling techniques or other network optimization tools. In fact, even without user scheduling, this angular separation is achieved with high probability as will be illustrated by simulations in Section 2.8 under reasonable system and channel assumptions. Further, for sparse channels with finite number of paths, it can be shown that this angular separation is not required to achieve the optimal rate. Studying these topics are interesting future extensions.

In the following theorem, we derive a lower bound on the achievable rate with multi-layer precoding for the cell-edge users.

Theorem 15 *Consider the system and channel models described in Theorem 14. If* $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_{\mathrm{A}}$ *or* $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_{\mathrm{E}}$, $\forall m \neq k$, *and* $d_{\max} \leq d_{ck} \leq r_{\mathrm{cell}}$, *then the achievable rate of user* k *at cell* c, *when applying the multi-layer precoding algorithm in Section 4.5, satisfies*

$$\lim_{N_{\mathrm{V}},N_{\mathrm{H}}\to\infty} R_{ck} \geq \log_2\left(1 + \mathsf{SNR}\,\|\overline{\mathbf{w}}_{cck}\|^2\,\sigma^2_{\min}\left(\overline{\mathbf{U}}^{\mathrm{E}}_{cck}\right)\right). \tag{4.21}$$

Proof: Similar to the proof of Theorem 14, if $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_{\mathrm{A}}$ or $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_{\mathrm{E}}$, $\forall m \neq k$, then $\lim_{N_{\mathrm{V}},N_{\mathrm{H}}\to\infty} \mathbf{G}_{c,(k,m)} = \mathbf{0}$. Using the matrix inversion lemma and leveraging the block diagonal structure of $\mathbf{W}_c$, we get $\left(\mathbf{W}^*_c\mathbf{F}^{(2)^*}_c\mathbf{F}^{(2)}_c\mathbf{F}^{(2)^*}_c\mathbf{F}^{(2)}_c\mathbf{W}_c\right)^{-1}_{k,k} = \left(\overline{\mathbf{w}}^*_{cck}\mathbf{G}_{c(k,k)}\overline{\mathbf{w}}_{cck}\right)^{-1}$. Note that since $d > d_{\max}$, $\mathbf{U}_{cck}$ is not guaranteed to be in Range $\left(\mathbf{U}^{\mathrm{NI}}_c\right)$, and $\overline{\mathbf{U}}^{\mathrm{E}^*}_{cck}\overline{\mathbf{U}}^{\mathrm{E}}_{cck} \neq \mathbf{I}$ in general. The achievable rate of user k at cell c can therefore be written as

$$\begin{aligned}
\lim_{N_{\mathrm{V}},N_{\mathrm{H}}\to\infty} R_{ck} &= \log_2\left(1 + \mathsf{SNR}\;\overline{\mathbf{w}}^*_{cck}\mathbf{G}_{c(k,k)}\overline{\mathbf{w}}_{cck}\right) && (4.22)\\
&\overset{(a)}{\geq} \log_2\left(1 + \mathsf{SNR}\;\;\|\overline{\mathbf{w}}_{cck}\|^2\,\sigma^2_{\min}\left(\mathbf{I}\otimes\overline{\mathbf{U}}^{\mathrm{E}}_{cck}\right)\right) && (4.23)\\
&\overset{(b)}{=} \log_2\left(1 + \mathsf{SNR}\;\;\|\overline{\mathbf{w}}_{cck}\|^2\,\sigma^2_{\min}\left(\overline{\mathbf{U}}^{\mathrm{E}}_{cck}\right)\right), && (4.24)
\end{aligned}$$

where (a) follows by applying the Rayleigh-Ritz theorem [101], and (b) results from the properties of the Kronecker product.

□

Theorem 15 indicates that cell edge users experience some degradation in their SNRs as a cost for the perfect inter-cell interference avoidance. In Section 4.7, we will discuss some solutions that make compromises between the degradation of the desired signal power and the management of the inter-cell interference for cell-edge users, under the multi-layer precoding framework.

4.6.2 Performance with Single-Path Channel Models

Rank-1 channel models describe the cases where the signal propagation through the channel is dominated by one line-of-sight (LOS) or non-LOS (NLOS) path. This is particularly relevant to systems with sparse channel, such as mmWave systems [7, 21, 38]. A special case of rank-1 channel models is the single-path channels. Consider a user k at cell c with a single path channel, defined by its azimuth and elevation angles ϕ_{ck}, θ_{ck}, the channel vector can be expressed as

$$\mathbf{h}_{cck} = \rho_{cck}^{\frac{1}{2}} \, \beta_{ck} \, \mathbf{a}_{\mathrm{A}} \left(\phi_{ck}, \theta_{ck} \right) \otimes \mathbf{a}_{\mathrm{E}} \left(\phi_{ck}, \theta_{ck} \right), \tag{4.25}$$

where $\mathbf{a}_{\mathrm{A}} \left(\phi_{ck}, \theta_{ck} \right)$ and $\mathbf{a}_{\mathrm{E}} \left(\phi_{ck}, \theta_{ck} \right)$ are the azimuth and elevation array response vectors, β_{ck} is the complex path gain, and ρ_{cck} is its path loss.

In the next corollary, we characterize the achievable rate of the proposed multi-layer precoding design for single-path channels.

Corollary 16 *Consider the full-dimensional cellular system model in Section 4.3.1, and the single-path channel model in* (4.25)*. When applying the multi-layer precoding algorithm in Section 4.5, the achievable rate of user k at cell c satisfies*

$$\lim_{N_{\mathrm{V}}, N_{\mathrm{H}} \to \infty} R_{ck} = \overline{R}_{ck} = \log_2 \left(1 + \mathsf{SNR} \left\| \mathbf{h}_{cck} \right\|^2 \right). \tag{4.26}$$

Proof: The proof is similar to that of Theorem 14, and is omitted due to space limitations. □

Corollary 16 indicates that the proposed multi-layer precoding design can achieve an optimal performance for single-path channels, making it a promising solution for mmWave and low channel rank massive MIMO systems. This will also be verified by numerical simulations in Section 2.8.

4.7 Discussion and Extensions

While we proposed and analyzed a specific multi-layer precoding design in this paper, there are many possible extensions as well as important topics that need further investigations. In this section, we briefly discuss some of these points, leaving their extensive study for future work.

4.7.1 Multi-Layer Precoding with Augmented Vertical Dimensions

As explained in Section 4.5, the proposed multi-layer precoding design attempts to perfectly avoid the inter-cell interference by forcing its transmission to be in the elevation null-space of the interference. While this guarantees optimal performance for cell-interior users and decontaminates the pilots for all the cell users, it may also block some of the desired signal power at the cell-edge. In this section, we propose a modified design for the first precoding layer $\mathbf{F}_c^{(1)}$ that compromises between the inter-cell interference avoidance and the desired signal degradation. The main idea of the proposed design, that we call multi-layer precoding with augmented vertical dimensions, is to simply extend the null-space of the inter-cell interference via

exploiting large channel characteristics. This is summarized as follows. Leveraging Lemma 2 in [86], the rank of the one-ring correlation matrix can be related to its angular range $[\theta_{\min}, \theta_{\max}]$ as

$$\operatorname{rank}(\mathbf{R}) = \frac{ND}{\lambda}\left(\cos(\theta_{\min}) - \cos(\theta_{\max})\right) \text{ as } N \to \infty. \tag{4.27}$$

Applying this lemma to the elevation inter-cell interference subspace, setting $\theta_{\min} = \pi/2$, BS c can estimate its maximum interference elevation angle, denoted θ_{I}, as

$$\theta_{\mathrm{I}} = \arccos\left(-\frac{\operatorname{rank}(\mathbf{R}_{\mathrm{I}})\,\lambda}{N_{\mathrm{V}} D}\right). \tag{4.28}$$

Extending the null space of the interference can then be done by virtually reducing the inter-cell interference subspace. Let δ_{E} denote the angular range of the extended subspace, the modified inter-cell interference covariance can then be calculated as

$$\left[\overline{\mathbf{R}}_{\mathrm{I}}\right]_{n_1,n_2} = \frac{1}{\theta_{\mathrm{I}} - \delta_{\mathrm{E}} - \pi/2}\int_{\frac{\pi}{2}}^{\theta_{\mathrm{I}}-\delta_{\mathrm{E}}} e^{jkD(n_2-n_1)\cos(\alpha)} d\alpha. \tag{4.29}$$

Finally, if $\left[\overline{\mathbf{U}}_c^{\mathrm{I}}\ \overline{\mathbf{U}}_c^{\mathrm{NI}}\right]\overline{\mathbf{\Lambda}}_c\left[\overline{\mathbf{U}}_c^{\mathrm{I}}\overline{\mathbf{U}}_c^{\mathrm{NI}}\right]^*$ represents the eigen-decomposition of $\overline{\mathbf{R}}_{\mathrm{I}}$, with $\overline{\mathbf{U}}_c^{\mathrm{I}}$ and $\overline{\mathbf{U}}_c^{\mathrm{NI}}$ correspond to the non-zero and zero eigenvalues, then the modified first precoding layer can be constructed as

$$\mathbf{F}_c^{(1)} = \mathbf{I} \otimes \overline{\mathbf{U}}_c^{\mathrm{NI}}. \tag{4.30}$$

Note that under this multi-layer precoding design, only cell edge users will experience inter-cell interference and pilot contamination while optimal performance is still guaranteed for cell-interior users. This yields an advantage for multi-layer precoding over conventional massive MIMO precoding schemes, which will also be illustrated by numerical simulations in Section 2.8.

4.7.2 TDD and FDD Operation with Multi-Layer Precoding

While we focused on TDD systems in this paper, the fact that multi-layer precoding relies on large-scale channel statistics makes it attractive for FDD operation as well. In FDD systems, the adjacent cells will cooperate to construct the elevation inter-cell interference subspace, which is needed to build the first precoding layer. Since this channel knowledge is of very large-scale statistics and this precoding layer needs to be updated every long time period, this cooperation overhead can be reasonably low. Given the first layer spatial filtering, every BS can estimate its users covariance knowledge free of inter-cell interference. Thanks to the multiplicative structure of the multi-layer precoding and its successive dimensional reduction, only the third precoding layer requires the instantaneous knowledge of the effective channel, which has much smaller dimensions. It is worth noting here that other FDD massive MIMO precoding schemes, such as JSDM [23] with its user grouping functions, can be easily integrated into the proposed multi-layer precoding framework for full-dimensional massive MIMO cellular systems.

In TDD systems, the required channel knowledge for the three stages can be done through uplink training with different time scales. One important note is that the second precoding layer (and its channel training) may not be needed in TDD systems with fully-digital transceivers, as the instantaneous channels can be easily trained in the uplink with a small number of pilots. This precoding layer, however, is important if multi-layer precoding is implemented using hybrid architectures, as will be shown in the following subsection.

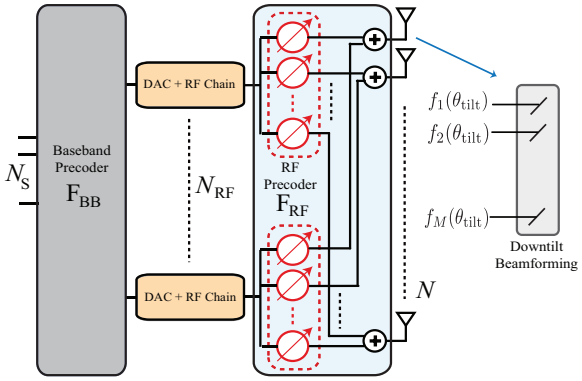


Figure 4.3: The figure shows a hybrid analog/digital architecture, at which baseband precoding, RF precoding, and antenna downtilt beamforming can be utilized to implement the multi-layer precoding algorithm.

4.7.3 Multi-Layer Precoding using Hybrid Architectures

Thanks to the multiplicative structure and the specific multi-layer precoding design in Section 4.5, we note that each precoding layer has less dimensions compared to the prior layers. This allows the multi-layer precoding matrices to be implemented using hybrid analog/digital architectures [4, 20, 30, 93], which reduces the required number of RF chains. In this section, we briefly highlight one possible idea for the hybrid analog/digital implementation, leaving its optimization and extensive investigation for future work.

Considering the three-stage multi-layer precoding design in Section 4.5, we propose to implement the first and second layers in the RF domain and perform the third layer precoding at baseband. Given the successive dimensional reductions, the required number of RF chains is expected to be much less than the number of antennas, especially in sparse and low-rank channels. As the first precoding layer focuses on avoiding the inter-cell interference in the elevation direction, we can implement it

using downtilt directional antenna patterns. We assume that each antenna port has a directional pattern and electrically adjusted downtilt angle [90, 102]. For example, The 3GPP antenna port elevation gain $G^{\mathrm{E}}(\theta)$ is defined as [90]

$$G^{\mathrm{E}}(\theta) = G^{\mathrm{E}}_{\max} - \min\left\{12\left(\frac{\theta-\theta_{\mathrm{tilt}}}{\theta_{\mathrm{3dB}}}\right)^2, \mathrm{SL}\right\}, \tag{4.31}$$

where θ_{tilt} is the downtilt angle, and SL is the sidelobe level. Therefore, one way to approximate $\mathbf{F}_c^{\mathrm{E}(1)}$ is to adjust the downtilt angle θ_{tilt} to minimize the leakage transmission outside the interference null-space $\mathbf{U}_c^{\mathrm{NI}}$.

Once $\mathbf{F}_c{}^{(1)}$ is implemented, the second precoding layer $\mathbf{F}_c{}^{(2)}$ can be designed similar to [30], i.e., each column of $\mathbf{F}_c{}^{(2)}$ can be approximated by a beamsteering vector taken from a codebook that captures the analog hardware constraints. Finally, the third precoding layer $\mathbf{F}_c^{(3)}$ is implemented in the baseband to manage the multi-user interference based on the effective channels that include the effect of the first and second precoding layers.

4.8 Simulation Results

In this section, we evaluate the performance of the proposed multi-layer precoding algorithm using numerical simulations. We also draw insights into the impact of the different system and channel parameters.

We consider a single-tier 7-cell cellular system model as depicted in Fig. 4.4(a), and calculate the performance for the cell in the center. Unless otherwise mentioned, every BS is assumed to a have a UPA, oriented in the y-z plane, at a height $H_{\mathrm{BS}} =$ 35m, and serving users at cell radius $r_{\mathrm{cell}} = 100$m. Users are randomly and uniformly

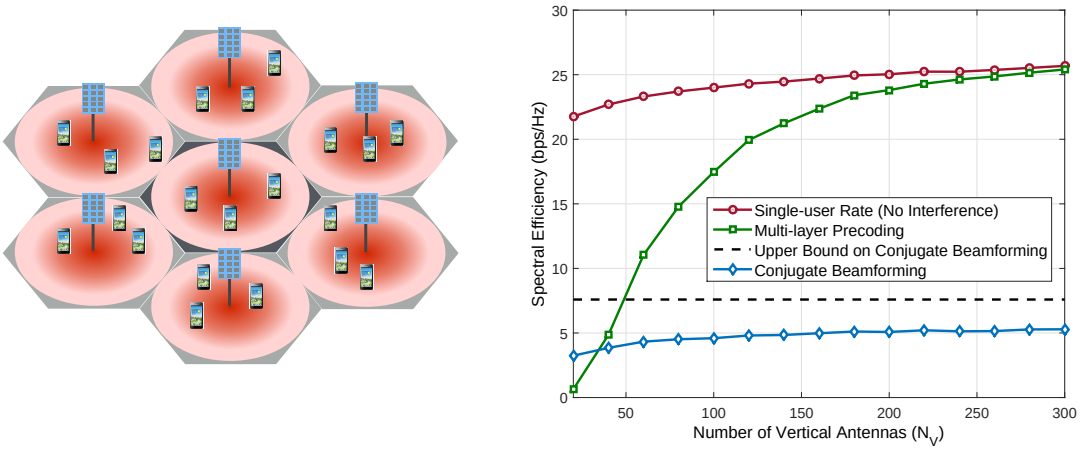
dropped in the cells, and every cell randomly schedules $K = 20$ users to be served at the same time and frequency slot. The BS transmit power is $P = 5$ dB and the receiver noise figure is 7 dB. The system operates at a carrier frequency 4 GHz, with a bandwidth 10 MHz, and a path loss exponent 3.5. Two channel models are assumed, namely, the single-path and the one-ring channel models.

The BSs in the adopted system apply the multi-layer precoding algorithm in Section 4.5. The required channel knowledge is obtained through three-stage uplink channel training, and every stage includes the effect of the previously designed precoding layers. Next, we present the simulation results for the two adopted channel models.

4.8.1 Results with Single-Path Channels

In this section, we adopt a single-path model for the user channels as described in (4.25). The azimuth and elevation angles are geometrically determined based on users locations relative to the BSs, and the complex path gains $\beta_{ck} \sim \mathcal{CN}(0, 1)$.

Optimality with large antennas: In Fig. 4.4(b), we compare the per-user achievable rate of multi-layer precoding with the single-user rate and the rate with conventional conjugate beamforming. The BSs are assumed to employ UPAs that have $N_{\mathrm{H}} = 30$ horizontal antennas and different numbers of vertical antennas. First, we note that the per-user achievable rate with multi-layer precoding approaches the optimal single-user rate as the number of antennas grow large. This verifies the asymptotic optimality result of multi-layer precoding given in Corollary 16. Note that the single-user rate is the rate if only this user is served in the network, i.e.,



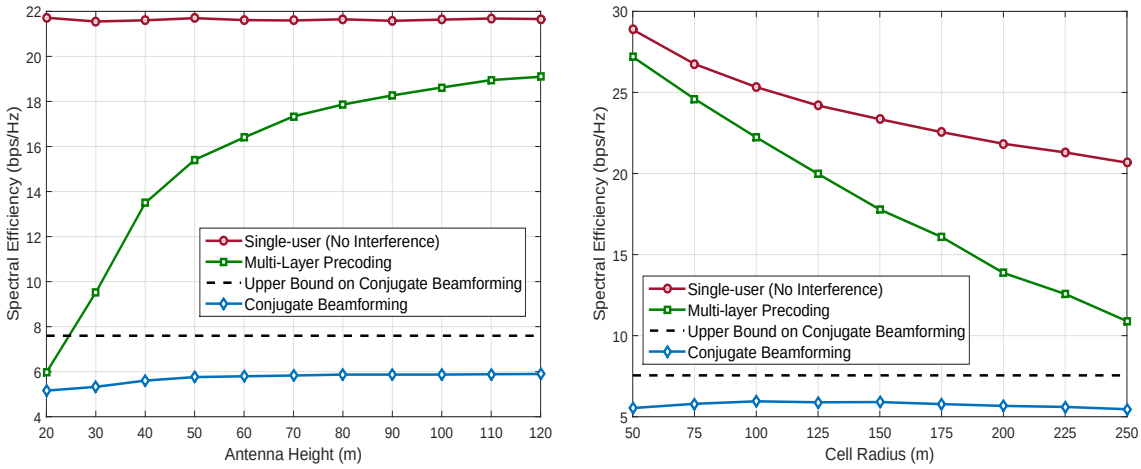
(a) FD massive MIMO cellular model (b) $N_{\mathrm{H}} = 30$

Figure 4.4: The adopted single-tier (7-cells) cellular model with FD massive MIMO antennas at the BSs is illustrated in (a). In (b), the achievable rate of the proposed multi-layer precoding is compared to the single-user rate and the rate with conventional conjugate beamforming, for different numbers of vertical antennas. The number of BS horizontal antennas is $N_{\mathrm{H}} = 30$, and the users are assumed to have single-path channels.

with no inter-cell or multi-user intra-cell interference. In the figure, we also plot the achievable rate with conventional conjugate beamforming. This assumes that channels are estimated using uplink training and then conjugate beamforming is applied in the downlink data transmission [1]. As a function of the path-loss ρ_{bck} in (4.25), the conjugate beamforming rate is theoretically bounded from above by [1]

$$\overline{R}_{ck}^{CB} = \log_2\left(1 + \mathsf{SNR}\frac{\rho_{cck}^2}{\sum_{b\neq c}\rho_{bck}^2}\right), \tag{4.32}$$

which limits its rate from growing with the number of antenna beyond this value. Interestingly, the multi-layer precoding rate does not have a limit on its rate and can grow with the number of antennas and transmit power without a theoretical limit.



(a) $r_{\text{cell}} = 200\text{m}$ (b) $H_{\text{BS}} = 35\text{m}$

Figure 4.5: The achievable rate of the proposed multi-layer precoding is compared to the single-user rate and the rate with conventional conjugate beamforming for different cell radii. The BSs are assumed to employ 120×30 UPAs, and the users have single-path channels.

The intuition behind that lies in the inter-cell interference avoidance using large channel statistics in multi-layer precoding. This works as a spatial filtering that avoids uplink channel estimation errors due to pilot reuse among cells and cancels inter-cell interference in the downlink data transmission. Therefore, the multi-layer precoding rate is free of the pilot-contamination impact. Note that while the asymptotic optimality of multi-layer precoding is realized at large antennas numbers, Fig. 4.4(b) shows it can still achieve gain over conventional massive MIMO beamforming schemes at much lower number of antennas.

Impact of antenna heights and cell radii: In Fig. 4.5, we evaluate the impact of the BS antenna height and cell radius on the achievable rates. This figure adopts the same system and channel assumptions in Fig. 4.4(b). In Fig. 4.5(a), the

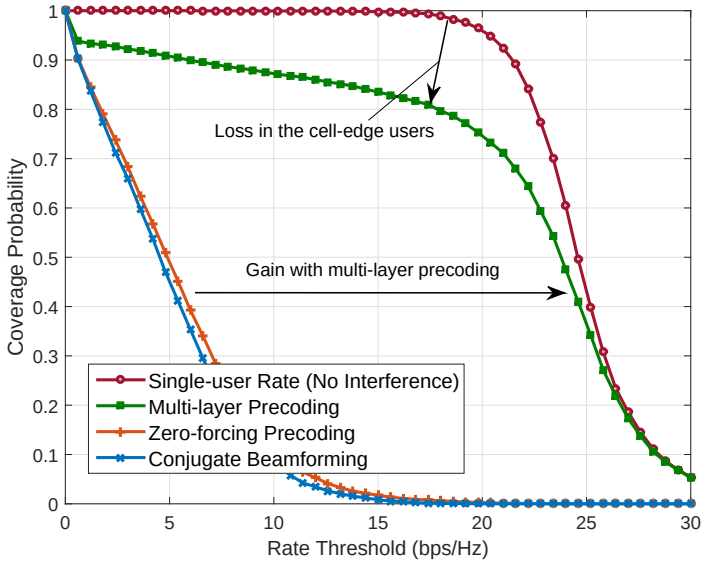


Figure 4.6: The rate coverage gain of the proposed multi-layer precoding algorithm over conventional conjugate beamforming and zero-forcing is illustrated. This rate coverage is also shown to be close to the single-user case. The BSs are assumed to employ 120×30 UPAs at heights $H_{\text{BS}} = 35$m, the cell radius is $r_{\text{cell}} = 100$m, and the users have single-path channels.

achievable rates for multi-layer precoding, single-user, and conjugate beamforming are compared for different antenna heights, assuming cells of radius 200m. The figure shows that multi-layer precoding approaches single-user rates at higher antenna heights. This is intuitive because forcing the transmission to become in the elevation null-space of the interference may have less impact on the desired signal blockage if higher antennas are employed. Note that the convergence to the single-user rate is expected to happen at lower antenna heights when large arrays are deployed. These achievable rates are again compared in Fig. 4.5(b), but for different cell radii. This figures illustrates that higher cell radius generally leads to less rate because of the

higher path loss. Further, the difference between single-user and multi-layer precoding rates increases at higher cell radii. In fact, this is similar to the degradation with less antenna heights, i.e., due to the impact of the inter-cell interference avoidance on the desired signal power. For reasonable antenna heights and cell radii, however, the multi-layer precoding still achieves good gain over conventional conjugate beamforming.

Rate coverage: To evaluate the rate coverage of multi-layer precoding, we plot Fig. 4.6. The same setup of Fig. 4.4(b) is adopted again with cells of radius 100m, and BSs with 120×30 UPAs at heights 35m. First, the figure shows that multi-layer precoding achieves very close coverage to the single-user case, especially for users not at the cell edge. For example, $\sim 60\%$ of the multi-layer precoding users get the same rate of the single-user case. At the cell edge, some degradation is experienced due to the first precoding layer that filters out-of-cell interference and affects the desired signal power. This loss, though, is expected to decrease as more antennas are employed. The figure also shows significant rate coverage gain over conventional conjugate beamforming and zero-forcing precoding solutions.

4.8.2 Results with One-Ring Channels

In this section, we adopt a one-ring model for the user channels as described in (4.19). The azimuth and elevation angles are geometrically determined based on users locations relative to the BSs, and the angular spread is set to $\Delta_{\mathrm{A}} = 5^\circ, \Delta_{\mathrm{E}} = 3^\circ$. Every BS randomly selects $K = 20$ users to be served, i.e., no scheduling is done to guarantee the angular separation condition in Theorem 14 and Theorem 15.

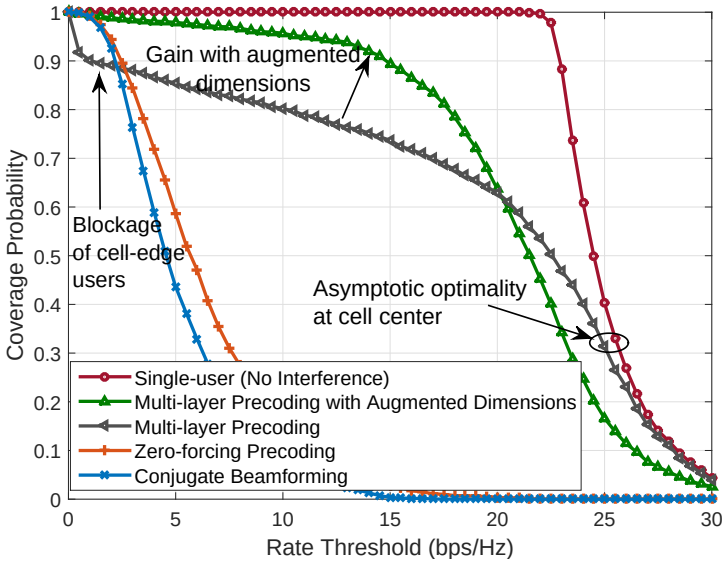


Figure 4.7: The rate coverage gain of the proposed multi-layer precoding algorithms over conventional conjugate beamforming and zero-forcing is illustrated. This rate coverage is also shown to be close to the single-user case. Further, the modified algorithm with augmented vertical dimensions can overcome the cell-edge blockage. The BSs are assumed to employ 100×40 UPAs at heights $H_{\text{BS}} = 35$m, the cell radius is $r_{\text{cell}} = 100$m. The users have one-ring channel models of azimuth and elevation angular spread $\Delta_{\text{A}} = 5°, \Delta_{\text{E}} = 3°$

Rate coverage: In Fig. 4.7-Fig. 4.9, we compare the rate coverage of multi-layer precoding, single-user, and conventional conjugate beamforming, for different antenna sizes. We also plot the rate coverage of the multi-layer precoding with augmented vertical dimensions described in Section 4.7.1, assuming an extended angle $\delta_{\text{E}} = 2\Delta_{\text{E}}$. This choice makes the maximum no-blockage distance $d_{\max}$, defined in Theorem 14, to be equal to the cell radius. Optimization of this parameter deserves more study in future extensions. Fig. 4.7 considers the system model in Section 4.3 with 100×40 BS UPAs and one-ring channel model. First, the figure shows that multi-layer precoding achieves close coverage to single-user case at cell center. For

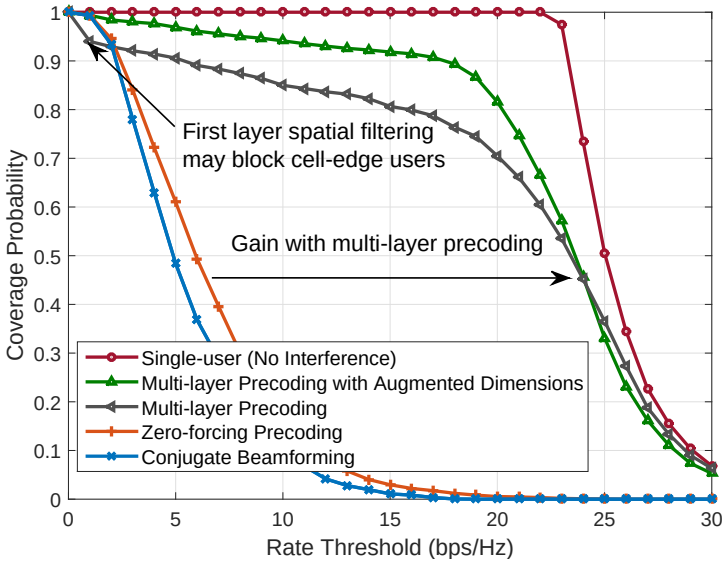


Figure 4.8: The rate coverage gain of the proposed multi-layer precoding algorithms over conventional conjugate beamforming and zero-forcing is illustrated. This rate coverage is also shown to be close to the single-user case. Further, the modified algorithm with augmented vertical dimensions can overcome the cell-edge blockage. The BSs are assumed to employ 140×40 UPAs at heights $H_{\text{BS}} = 35$m, the cell radius is $r_{\text{cell}} = 100$m. The users have one-ring channel models of azimuth and elevation angular spread $\Delta_{\text{A}} = 5°, \Delta_{\text{E}} = 3°$.

cell edge, though, multi-layer precoding users experience high blockage, which results from the elevation inter-cell interference avoidance. This can be improved when augmenting vertical subspaces as described in Section 4.7.1. Note that, different than the multi-layer precoding case, the small degradation at the cell-edge users is due to inter-cell interference, not signal blockage. Further, it is important to note that cell-center with the modified algorithm in Section 4.7.1 still achieve asymptotic optimal rate, i.e., no inter-cell interference or pilot contamination impact exist. The same behavior is shown again in Fig. 4.8, when larger array sizes are employed. In this case, though, the cell-edge blockage with multi-layer precoding is less at better

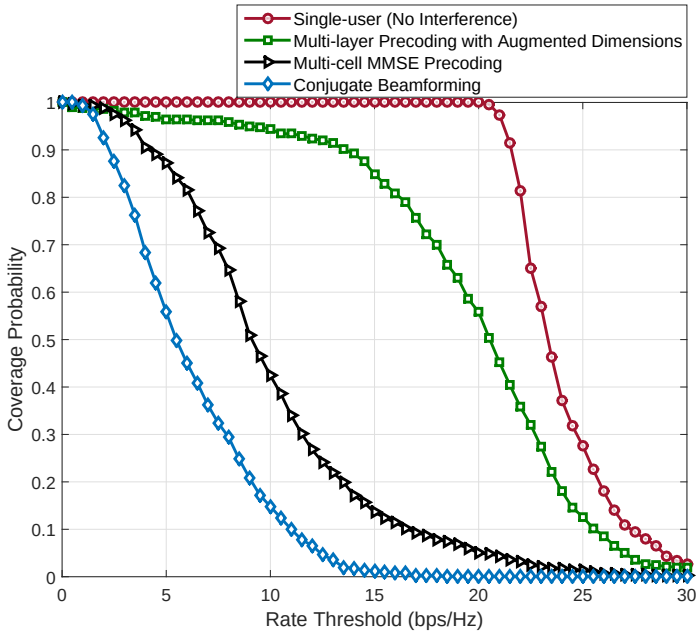


Figure 4.9: The rate coverage gain of the proposed multi-layer precoding algorithms over conventional single-cell conjugate beamforming and multi-cell MMSE precoding. This rate coverage is also shown to be close to the single-user case. The BSs are assumed to be at heights $H_{\text{BS}} = 35$m, the cell radius is $r_{\text{cell}} = 100$m. The users have one-ring channel models of azimuth and elevation angular spread $\Delta_{\text{A}} = 5°, \Delta_{\text{E}} = 3°$.

separation between the desired cell and the other cells' users can be achieved. In the two figures, multi-layer precoding with augmented vertical subspaces is shown to have good coverage gain over conventional massive MIMO precoding solutions.

In Fig. 4.9, we consider the same system and channel models in Fig. 4.7, but with 80×20 UPAs and $K = 5$ users to reduce the computational complexity. Fig. 4.9 compares the rate coverage of the proposed augmented dimension based multi-layer precoding with the single-user rate and the single-cell conjugate beamforming. The figure also plots the rate coverage of the multi-cell MMSE precoding in [84] that

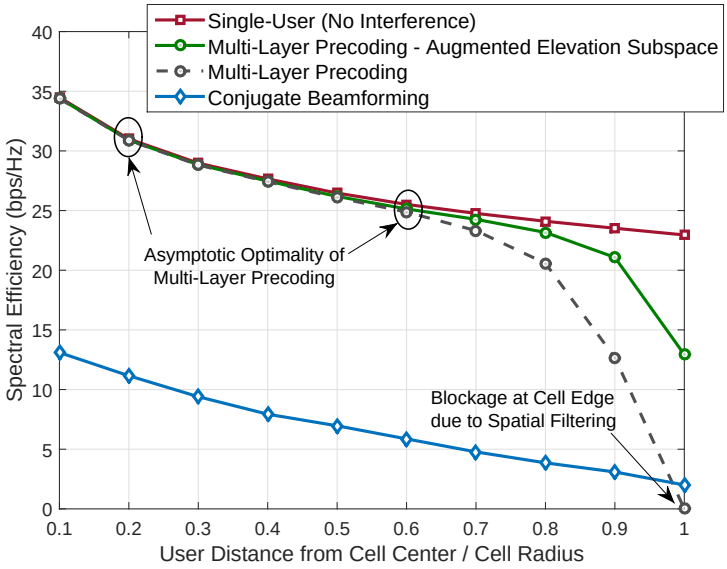


Figure 4.10: The achievable rates of the proposed multi-layer precoding algorithms are compared to the single-user rate and the rate with conventional conjugate beamforming, for different distances from cell center. The BSs are assumed to employ 120×40 UPAs at heights $H_{\text{BS}} = 35$m and the cell radius is $r_{\text{cell}} = 100$m. The users have one-ring channel models of azimuth and elevation angular spread $\Delta_{\text{A}} = 5°, \Delta_{\text{E}} = 3°$.

manages the inter-cell interference. As shown in the figure, multi-layer precoding achieves a close performance to single-user rate and good gain over single-cell precoding. Fig. 4.9 also illustrates that multi-layer precoding achieves a reasonable gain over multi-cell MMSE precoding despite the requirement of less channel knowledge.

Rates at cell-interior and cell-edge: To illustrate the achievable rates for cell-interior and cell-edge users, we plot the achievable rates of multi-layer precoding, single-user, and conventional conjugate beamforming in Fig. 4.10. The rates are plotted versus the user distance to the BS, normalized by the cell radius $r_{\text{cell}} = 100$m. The figure confirms the asymptotic optimal performance of multi-layer precoding at

cell-interior, given in Theorem 14. At cell edge, users experience some blockage that can be fixed with the augmented vertical dimension modification in Section 4.7.1. Compared to conventional conjugate beamforming performance, the multi-layer precoding with augmented vertical dimensions exhibits very good gains, even at cell edge.

4.9 Conclusion

In this paper, we proposed a general precoding framework for full-dimensional massive MIMO systems, called multi-layer precoding. We developed a specific design for multi-layer precoding that efficiently manages different kinds of interference, leveraging the large channel characteristics. Using analytical derivations and numerical simulations, we showed that multi-layer precoding can guarantee asymptotically optimal performance for the cell-interior users under the one-ring channel models and for all the users under single-path channels. For the cell-edge users, we proposed a modified multi-layer precoding design that compromises between desired signal power maximization and inter-cell interference avoidance. Results indicated that multi-layer precoding can achieve close performance, in terms of rate and coverage, to the single-user case. Further, results showed that multi-layer precoding achieves clear gains over conventional massive MIMO precoding techniques. For future work, it would be interesting to investigate and optimize the implementation of multi-layer precoding using hybrid analog/digital architectures. It is also important to develop techniques for the channel training and estimation under hybrid architecture hardware constraints.

Chapter 5

Frequency Selective Hybrid Precoding

5.1 Overview

Most prior work has focused on hybrid precoding for narrow-band mmWave systems, with perfect or estimated channel knowledge at the transmitter. MmWave systems, however, will likely operate on wideband channels with frequency selectivity. In this chapter[1], therefore, we consider wideband mmWave systems with a limited feedback channel between the transmitter and receiver. First, the optimal hybrid precoding design for a given RF codebook is derived. This provides a benchmark for any other heuristic algorithm and gives useful insights into codebook designs. Second, efficient hybrid analog/digital codebooks are developed for spatial multiplexing in wideband mmWave systems. Finally, a low-complexity yet near-optimal greedy frequency selective hybrid precoding algorithm is proposed based on Gram-Schmidt orthogonalization. Simulation results show that the developed hybrid codebooks and precoder designs achieve very good performance compared with the unconstrained solutions while requiring much less complexity.

[1]This chapter is based on the work published in the journal paper: A. Alkhateeb and R. W. Heath, "Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems," in IEEE Transactions on Communications, vol. 64, no. 5, pp. 1801-1818, May 2016. This work was supervised by Prof. Robert Heath.

5.2 Introduction

MmWave communication can leverage the large bandwidths potentially available at millimeter wave carrier frequencies to provide high data rates [103]. This makes mmWave a promising carrier frequency for 5G cellular systems [3,6–8,104,105]. Recent channel measurements have confirmed the feasibility of using mmWave not only for backhaul [24, 105, 106], but also for the access link [21]. Further, system level evaluation of mmWave network performance indicate that mmWave cellular systems can achieve a similar spectral efficiency to that obtained at lower-frequency while providing orders of magnitudes more data rate thanks to the larger bandwidth [9, 107–109]. Though mmWave cellular is recently of interest for 5G, it was proposed as early as thirty years ago [110]. MmWave wireless communication has been considered for many other applications beyond cellular systems including wireless local area networks [36], personal area networks [111], wearable device communications [112, 113], joint vehicular communication and radar systems [114–116], and simultaneous energy/data transfer [117–119].

To guarantee sufficient received signal power at mmWave frequencies, large antenna arrays are beneficial at both the transmitter and receiver [6,11,20,21,93,103]. Fortunately, large antenna arrays can be packed into small form factors due to the small mmWave antenna size [17, 120]. Exploiting large arrays using MIMO signal processing techniques like precoding and combining, however, is different at mmWave compared with sub-6 GHz solutions. This is mainly due to the different hardware constraints on the mixed signal components because of their high cost and power consumption [29]. Further, the best precoders are designed based on instantaneous

channel state information, which is difficult to acquire at the transmitter in large mmWave systems [29] due to the high channel dimensionality. Therefore, developing precoding algorithms and codebooks for limited feedback wideband mmWave systems is important for building these systems.

Hybrid analog/digital precoding, which divides the precoding between analog and digital domains, was proposed to handle the trade-off between the low-complexity limited-performance analog-only solutions and the high-complexity good-performance fully digital precoding [18–20, 27, 39, 62, 121–123]. The main advantage of hybrid precoding over conventional precoding is that it can deal with having fewer RF chains than antennas. For general MIMO systems, hybrid precoding for diversity and multiplexing gain were investigated in [18], and for interference management in [19]. These solutions, however, did not make use of the special mmWave channel characteristics in the design as they were not specifically developed for mmWave systems. In [20], the sparse nature of mmWave channels was exploited; low-complexity iterative algorithms based on orthogonal matching pursuit were devised, assuming perfect channel knowledge at the transmitter. Extensions to the case when only partial channel knowledge is required and when the channel and hybrid precoders are jointly designed were considered in [27, 39]. Algorithms that do not rely on orthogonal matching pursuit were proposed in [121–123] for the hybrid precoding design with perfect channel knowledge at the transmitter. The main objective of these algorithms was to achieve an achievable rate that approaches the rate achieved by fully-digital solutions. The work in [20, 27, 39, 121, 122], though, assumed a narrow-band mmWave channel, with perfect or partial channel knowledge at the transmitter. In [62], hybrid beamforming

with only a single-stream transmission over MIMO-OFDM system was considered. The solution in [62] though relied on the joint exhaustive search over both RF and baseband codebooks without giving specific criteria for the design of these codebooks. As mmWave communication is expected to employ broadband channels, developing spatial multiplexing hybrid precoding algorithms for wideband mmWave systems is important. Further, acquiring the large mmWave MIMO channels at the transmitter is difficult, which highlights the need to devise limited feedback hybrid precoding solutions.

In this work, we develop hybrid precoding solutions and codebooks for limited feedback wideband mmWave systems. In our proposed system, the digital precoding is done in the frequency domain and can be different for each subcarrier, while the RF precoder is frequency flat. The contributions of the work in this chapter are summarized as follows.

- First, we consider a frequency-selective hybrid precoding system with the RF precoders taken from a quantized codebook. For this system, we derive the optimal hybrid precoding design that maximizes the achievable mutual information under total power and unitary power constraints. Even though an exhaustive search over the RF codebook will be still required, the derived solution provides insights into hybrid analog/digital codebooks and greedy hybrid precoding design problems. Further, this solution gives a benchmark for the other heuristic algorithms that can be useful for evaluating their performance.

- Second, we consider a limited feedback frequency-selective hybrid precoding

system where both the baseband and RF precoders are taken from quantized codebooks. For this system, we develop efficient hybrid analog and digital precoding codebooks that attempt to minimize a distortion function defined by the average mutual information loss due to the quantized hybrid precoders when compared with the unconstrained digital solution.

- Finally, we design a greedy hybrid precoding algorithm based on Gram-Schmidt orthogonalization for limited feedback frequency selective mmWave systems. Despite its low-complexity, the proposed algorithm is illustrated to achieve a similar performance compared with the optimal hybrid precoding design that requires an exhaustive search over the RF and baseband codebooks.

The performance of the proposed codebooks and precoding algorithms is evaluated by numerical simulations in wideband mmWave setups, and compared with digital only precoding schemes in Section 5.8.

5.3 System and Channel Models

In this section, we describe the adopted frequency selective hybrid precoding system model and the wideband mmWave channel model. Key assumptions made for each model are also highlighted.

Consider the OFDM based system model in Fig. 5.1 where a BS with N_{BS} antennas and N_{RF} RF chains is assumed to communicate with a single MS with N_{MS} antennas and N_{RF} RF chains. The BS and MS communicate via N_{S} length-K data symbol blocks, such that $N_{\mathrm{S}} \leq N_{\mathrm{RF}} \leq N_{\mathrm{BS}}$ and $N_{\mathrm{S}} \leq N_{\mathrm{RF}} \leq N_{\mathrm{MS}}$. In practice, the

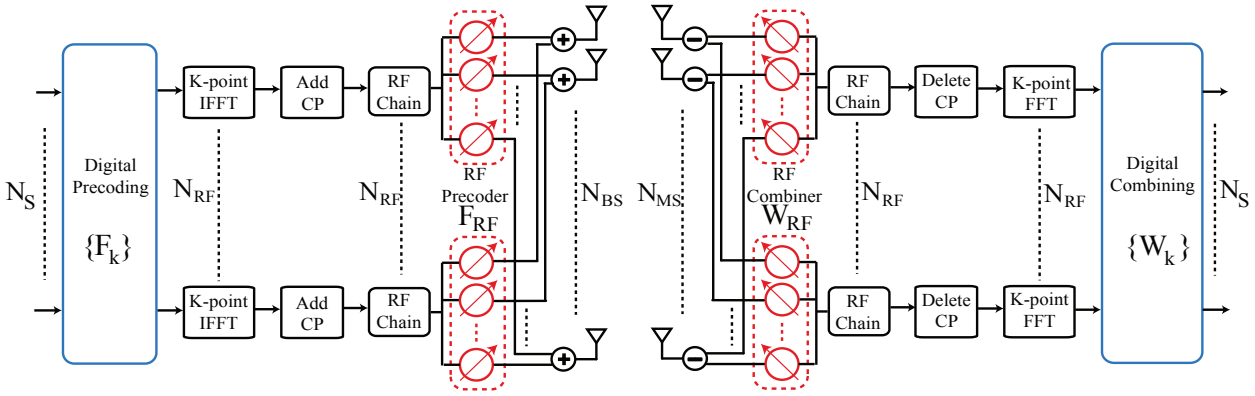


Figure 5.1: A block diagram of the OFDM based BS-MS transceiver that employs hybrid analog/digital precoding.

number of RF chains at the MS's is usually less than that of the BS's, but we do not exploit this fact in our model for simplicity of exposition.

At the transmitter, the N_{S} data symbols $\mathbf{s}_k$ at each subcarrier $k = 1, ..., K$ are first precoded using an $N_{\mathrm{RF}} \times N_{\mathrm{S}}$ digital precoding matrix $\mathbf{F}[k]$, and the symbol blocks are transformed to the time-domain using N_{RF} K-point IFFT's. Note that our model assumes that all subcarriers are used and, therefore, the data block length is equal to the number of subcarriers. A cyclic prefix of length D is then added to the symbol blocks before applying the $N_{\mathrm{BS}} \times N_{\mathrm{RF}}$ RF precoding $\mathbf{F}_{\mathrm{RF}}$. It is important to emphasize here that the RF precoding matrix $\mathbf{F}_{\mathrm{RF}}$ is the same for *all* subcarriers. This means that the RF precoder is assumed to be frequency flat while the baseband precoders can be different for each subcarrier. This is an important feature of the frequency selective hybrid precoding architecture in Fig. 5.1 that differentiates it from the conventional OFDM-based unconstrained digital scheme where only frequency-selective digital precoders exist. The discrete-time transmitted complex baseband

signal at subcarrier k can therefore be written as

$$\mathbf{x}[k] = \mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\mathbf{s}[k], \tag{5.1}$$

where $\mathbf{s}[k]$ is the $N_{\mathrm{S}} \times 1$ transmitted vector at subcarrier k, such that $\mathbb{E}\left[\mathbf{s}[k]\mathbf{s}^*[k]\right] = \frac{P}{KN_{\mathrm{S}}}\mathbf{I}_{N_{\mathrm{S}}}$, and P is the average total transmit power. Since $\mathbf{F}_{\mathrm{RF}}$ is implemented using analog phase shifters, its entries are of constant modulus. To reflect that, we normalize the entries $\left|[\mathbf{F}_{\mathrm{RF}}]_{m,n}\right|^2 = 1$. Further, we assume that the angles of the analog phase shifters are quantized and have a finite set of possible values. With these assumptions, $[\mathbf{F}_{\mathrm{RF}}]_{m,n} = e^{\mathrm{j}\phi_{m,n}}$, where $\phi_{m.n}$ is a quantized angle. The angle quantization assumption is discussed in more detail in Section 5.6. Note that the RF beamforming can also be designed as a frequency selective filter [124], with additional hardware complexity. Two precoding power constraints are considered in this work: (i) a total power constraint, where the hybrid precoders satisfy $\sum_{k=1}^{K} \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\|_F^2 = KN_{\mathrm{S}}$, and (ii) a unitary power constraint, where the hybrid precoders meet $\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k] \in \mathcal{U}_{N_{\mathrm{BS}}\times N_{\mathrm{S}}}, k = 1, 2, ..., K$, with the set of semi-unitary matrices $\mathcal{U}_{N_{\mathrm{BS}}\times N_{\mathrm{S}}} = \left\{\mathbf{U} \in \mathbb{C}^{N_{\mathrm{BS}}\times N_{\mathrm{S}}} | \mathbf{U}^*\mathbf{U} = \mathbf{I}\right\}$. Note that while the total power constraint allows the transmit power to be distributed, possibly non-uniformly, among the subcarriers and the data streams on each subcarrier, the unitary power constraint enforces an equal power allocation among the subcarriers and the data streams on each subcarrier.

At the MS, assuming perfect carrier and frequency offset synchronization, the received signal is first combined in the RF domain using the $N_{\mathrm{MS}} \times N_{\mathrm{RF}}$ combining matrix $\mathbf{W}_{\mathrm{RF}}$. Then, the cyclic prefix is removed, and the symbols are returned back

to the frequency domain with N_{RF} length-K FFT's. The symbols at each subcarrier k are then combined using the $N_{\mathrm{RF}} \times N_{\mathrm{S}}$ digital combining matrix $\mathbf{W}[k]$. The constraints on the entries of RF combiner $\mathbf{W}_{\mathrm{RF}}$ are similar to the RF precoders. Denoting the $N_{\mathrm{MS}} \times N_{\mathrm{BS}}$ channel matrix at subcarrier k as $\mathbf{H}[k]$, the received signal at subcarrier k after processing can be then expressed as

$$\mathbf{y}[k] = \mathbf{W}^*[k]\mathbf{W}^*_{\mathrm{RF}}\mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]s[k] + \mathbf{W}^*[k]\mathbf{W}^*_{\mathrm{RF}}\mathbf{n}[k], \tag{5.2}$$

where $\mathbf{n}[k] \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\mathrm{N}}\mathbf{I})$ is the Gaussian noise vector corrupting the received signal.

To incorporate the wideband and limited scattering characteristics of mmWave channels [6,7,21,36,125,126], we adopt a geometric wideband mmWave channel model with L clusters. Each cluster ℓ has a time delay $\tau_\ell \in \mathbb{R}$, and angles of arrival and departure (AoA/AoD), $\theta_\ell, \phi_\ell \in [0, 2\pi]$. Each cluster ℓ is further assumed to contribute with R_ℓ rays/paths between the BS and MS [9,125,127]. Each ray $r_\ell = 1, 2, ..., R_\ell$, has a relative time delay τ_{r_ℓ}, relative AoA/AoD shift ϑ_{r_ℓ}, φ_{r_ℓ}, and complex path gain α_{r_ℓ}. Further, let ρ_{PL} represent the path-loss between the BS and MS, and $p_{\mathrm{rc}}(\tau)$ denote a pulse-shaping function for T_{S}-spaced signaling evaluated at τ seconds [128]. Under this model, the delay-d MIMO channel matrix, $\boldsymbol{H}[d]$, can be written as [128]

$$\boldsymbol{H}[d] = \sqrt{\frac{N_{\mathrm{BS}}N_{\mathrm{MS}}}{\rho_{\mathrm{PL}}}} \sum_{\ell=1}^{L} \sum_{r_\ell=1}^{R_\ell} \alpha_{r_\ell} p_{\mathrm{rc}}\left(dT_{\mathrm{S}} - \tau_\ell - \tau_{r_\ell}\right) \mathbf{a}_{\mathrm{MS}}\left(\theta_\ell - \vartheta_{r_\ell}\right) \mathbf{a}^*_{\mathrm{BS}}\left(\phi_\ell - \varphi_{r_\ell}\right), \tag{5.3}$$

where $\mathbf{a}_{\mathrm{BS}}(\phi)$ and $\mathbf{a}_{\mathrm{MS}}(\theta)$ are the antenna array response vectors of the BS and MS, respectively. Given the delay-d channel model in (5.3), the channel at subcarrier k,

$\mathbf{H}[k]$, can be then expressed as [59]

$$\mathbf{H}[k] = \sum_{d=0}^{D-1} \boldsymbol{H}[d] e^{-j\frac{2\pi k}{K}d}. \tag{5.4}$$

While most of the results developed in this work are general for large MIMO channels, and not restricted to the channel model in (5.3), we described the wideband mmWave channel model in this section as it will be important for understanding the motivation behind the proposed construction of the hybrid analog/digital precoding codebooks in Section 5.6. Further, it will be adopted for the simulations in Section 5.8 and for drawing conclusions about the performance of the proposed precoding schemes and codebooks in wideband mmWave channels.

5.4 Problem Statement

In this work, we consider the downlink system model in Section 5.3 when the BS and MS are connected via a limited feedback link. For this setup, we assume the MS has perfect channel knowledge with which it selects the best RF and baseband precoding matrices $\mathbf{F}_{\mathrm{RF}}^{\star}$ and $\{\mathbf{F}^{\star}[k]\}_{k=1}^{K}$ from predefined quantization codebooks to maximize the achievable mutual information when used by the BS. The main objective of this work then is to develop efficient RF and baseband precoding codebooks for limited feedback wideband hybrid analog/digital precoding architectures. In this section, we first formulate the optimal hybrid precoding based mutual information when given RF and baseband precoding codebooks are used. Then, we briefly explain how the main objective of this work will be investigated in the subsequent sections.

As this chapter focuses on the limited feedback hybrid precoding design, i.e.,

the design of $\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}$, we will assume that the receiver can perform optimal nearest neighbor decoding based on the N_{MS}-dimensional received signal with fully digital hardware. This allows decoupling the transceiver design problem, and focusing on the hybrid precoders design to maximize the mutual information of the system [20], defined as

$$\mathcal{I}\left(\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}\right) = \frac{1}{K}\sum_{k=1}^{K} \log_2 \left| \mathbf{I}_{N_{\mathrm{MS}}} + \frac{\rho}{N_{\mathrm{S}}} \mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\mathbf{F}^{*}[k]\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{H}^{*}[k] \right|, \quad (5.5)$$

where $\rho = \frac{P}{K\sigma^2}$ is the SNR. As combining with fully digital hardware, though, is not a practical mmWave solution, the hybrid combining design problem needs also to be considered. The design ideas that will be given in this chapter for the hybrid precoders, however, provide direct tools for the construction of the hybrid combining matrices, $\mathbf{W}_{\mathrm{RF}}$, $\{\mathbf{W}[k]\}_{k=1}^{K}$, and is therefore omitted due to space limitations.

If the RF and baseband precoders are taken from quantized codebooks $\mathcal{F}_{\mathrm{RF}}$ and $\mathcal{F}_{\mathrm{BB}}$, respectively, then the maximum mutual information under the given hybrid precoding codebooks and the total power constraint is

$$\begin{aligned} \mathcal{I}_{\mathrm{HP}}^{\star} = \max_{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}} \quad & \mathcal{I}\left(\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}\right) \\ \text{s.t.} \quad & \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \\ & \mathbf{F}[k] \in \mathcal{F}_{\mathrm{BB}}, \quad k = 1, 2, ..., K, \\ & \sum_{k=1}^{K} \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\|_F^2 = KN_{\mathrm{S}}. \end{aligned} \quad (5.6)$$

The maximum mutual information with hybrid precoding and under the unitary power constraint is similar but with the last constraint in (5.6) replaced with $\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k] \in \mathcal{U}_{N_{\mathrm{BS}} \times N_{\mathrm{S}}}$.

Our main objective in this work is to construct efficient hybrid precoding codebooks $\mathcal{F}_{\mathrm{RF}}$ and $\mathcal{F}_{\mathrm{BB}}$ to maximize the achievable mutual information in (5.6). To get initial insights into the solution of this problem, we will first investigate a special case of the limited feedback hybrid precoding problem in Section 5.5 when only the RF precoders are taken from quantized codebooks while no quantization constraints are imposed on the baseband precoders. For this problem, we will derive the optimal hybrid precoding design for any given RF codebook $\mathcal{F}_{\mathrm{RF}}$. The results of Section 5.5 will help us developing RF and baseband precoders codebook in Section 5.6.

5.5 Optimal Hybrid Precoding Design for a Given RF Codebook

In this section, we investigate the limited feedback hybrid precoding design when only the RF precoders are taken from quantized codebooks. This problem is of a special interest for two main reasons. First, it will provide useful insights into the construction of efficient hybrid analog/digital precoding codebooks as will be summarized at the end of this section. Second, the hybrid precoding design problem with only RF precoders quantization can also be interpreted as the hybrid precoding design problem with perfect channel knowledge. The reason is that even when perfect channel knowledge is available at the transmitter, the RF precoders will be taken from a certain codebook that captures the hardware constraints such as the phase shifters quantization. With this motivation, we consider the following relaxation of the optimization in (5.7) that captures the assumption that only the RF precoders

are quantized.

$$\begin{aligned} \mathcal{I}^{\star}_{\mathrm{HP}} = \max_{\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}} \quad & \mathcal{I}\left(\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}\right) \\ \text{s.t.} \quad & \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \\ & \sum_{k=1}^{K} \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\|_F^2 = KN_{\mathrm{S}}. \end{aligned} \tag{5.7}$$

The design of the hybrid analog/digital precoders in (5.7) is non-trivial due to (i) the RF hardware non-convex constraint $\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}$, and (ii) the coupling between the analog and digital precoding matrices, which arises in the power constraint (the second constraint in (5.7)). Due to these difficulties, prior work [20, 27, 62] focused on developing heuristics designs for the hybrid analog/digital precoders in (5.7). Although these heuristic algorithms were shown to give good performance, they do not provide enough insights that help, for example, to design limited feedback hybrid precoding codebooks.

In this section, we will consider the coupling between the analog and digital precoders, and show that the optimal baseband precoders can be written as a function of the RF precoders under both the total and unitary power constraints. This will reduce the hybrid precoding design problem to an RF precoder design problem.

5.5.1 Total Power Constraint

As the RF precoding matrix $\mathbf{F}_{\mathrm{RF}}$ in (5.7) is taken from a quantized codebook $\mathcal{F}_{\mathrm{RF}}$, then the optimal mutual information in (5.7) can also be equivalently written

in the following outer-inner problems form

$$\mathcal{I}^{\star}_{\mathrm{HP}} = \max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \left\{ \begin{array}{ll} \max\limits_{\{\mathbf{F}[k]\}_{k=1}^{K}} & \mathcal{I}\left(\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}\right) \\ \\ \text{s.t.} & \sum\limits_{k=1}^{K} \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\|_F^2 = KN_{\mathrm{S}}, \end{array} \right\} \tag{5.8}$$

where the outer maximization is over the set of possible quantized RF precoding matrices, and the inner problem is over the set of feasible baseband precoders given the RF precoder, $\left\{\mathbf{F}[k] \in \mathbb{C}^{N_{\mathrm{RF}} \times N_{\mathrm{S}}} | \sum_{k=1}^{K} \|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\|_F^2 = KN_{\mathrm{S}}\right\}$.

Note that the solution of the optimal baseband precoders in the inner problem of (5.8) is not given by the simple SVD of the effective channel with the RF precoder, $\mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}$, because of the different power constraint that represents the coupling between the baseband and RF precoders. In the following proposition, we find the optimal baseband precoders of the inner problem of (5.8).

Proposition 17 *Define the SVD decompositions of the kth subcarrier channel matrix* $\mathbf{H}[k]$ *as* $\mathbf{H}[k] = \mathbf{U}[k]\mathbf{\Sigma}[k]\mathbf{V}^{*}[k]$*, and the SVD decomposition of the effective channel matrix* $\mathbf{\Sigma}[k]\mathbf{V}^{*}[k]\mathbf{F}_{\mathrm{RF}} \left(\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} = \overline{\mathbf{U}}[k]\overline{\mathbf{\Sigma}}[k]\overline{\mathbf{V}}^{*}[k]$*. Then, the baseband precoders* $\{\mathbf{F}[k]\}_{k=1}^{K}$ *that solve the inner optimization problem of* (5.8) *are given by*

$$\mathbf{F}^{\star}[k] = \left(\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}} \mathbf{\Lambda}[k], \quad k = 1, 2, ..., K, \tag{5.9}$$

where $\left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}}$ *is the* $N_{\mathrm{RF}} \times N_{\mathrm{S}}$ *matrix that gathers the* N_{S} *dominant vectors of* $\overline{\mathbf{V}}[k]$*, and* $\mathbf{\Lambda}[k]$ *is an* $N_{\mathrm{S}} \times N_{\mathrm{S}}$ *water-filling power allocation diagonal matrix with*

$$[\mathbf{\Lambda}[k]]_{i,i}^{2} = \left(\mu - \frac{N_{\mathrm{S}}}{\rho \left[\overline{\mathbf{\Sigma}}[k]\right]^{2}}\right)^{+}, i = 1, ..., N_{\mathrm{S}}, \ k = 1, ..., K, \tag{5.10}$$

and with μ satisfying

$$\sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{S}}}\left(\mu-\frac{N_{\mathrm{S}}}{\rho\left[\overline{\mathbf{\Sigma}}[k]\right]^{2}}\right)^{+}=KNs \tag{5.11}$$

Proof: See Appendix A in [34]. □

Given the optimal baseband precoder in (5.9), the optimal hybrid precoding based mutual information with the RF codebook $\mathcal{F}_{\mathrm{RF}}$ and a total power constraint can now be written as

$$\mathcal{I}^{\star}_{\mathrm{HP}}=\max_{\mathbf{F}_{\mathrm{RF}}\in\mathcal{F}_{\mathrm{RF}}}\frac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}_{N_{\mathrm{S}}}+\frac{\rho}{N_{\mathrm{S}}}\left[\overline{\mathbf{\Sigma}}[k]\right]^{2}_{1:N_{\mathrm{S}},1:N_{\mathrm{S}}}\mathbf{\Lambda}[k]^{2}\right|, \tag{5.12}$$

where $\overline{\mathbf{\Sigma}}[k]^2$ and $\mathbf{\Lambda}[k]^2$ are functions only of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{H}[k]$ as defined in Proposition 17. This means that the optimal hybrid precoding based mutual information is determined only by the RF precoders design. Hence, an exhaustive search over the RF precoders codebook $\mathcal{F}_{\mathrm{RF}}$ is sufficient to find the maximum achievable mutual information with hybrid precoding.

The hybrid precoding design in Proposition 17 can also be extended to the case when the power constraint is imposed on each subcarrier. In this case, the power constraint on the hybrid precoders is written as $\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\right\|_{\mathrm{F}}=N_{\mathrm{S}}, k=1,...,K$. The following corollary presents the optimal baseband precoder for a given RF codebook, under the per-subcarrier total power constraint.

Corollary 18 *The optimal baseband precoders that maximizes the objective of the inner optimization problem of* (5.8)*, under the constraint* $\left\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\right\|_{\mathrm{F}}=N_{\mathrm{S}}, k=$

$1, ..., K$, *are given by*

$$\mathbf{F}^{\star}[k] = \left(\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}} \mathbf{\Lambda}_{\mathrm{P}}[k], \quad k = 1, 2, ..., K, \tag{5.13}$$

where $\mathbf{\Lambda}_{\mathrm{P}}[k]$ *is an* $N_{\mathrm{S}} \times N_{\mathrm{S}}$ *water-filling power allocation diagonal matrix with*

$$\left[\mathbf{\Lambda}_{\mathrm{P}}[k]\right]_{i,i}^{2} = \left(\mu - \frac{N_{\mathrm{S}}}{\rho \left[\overline{\mathbf{\Sigma}}[k]\right]^{2}}\right)^{+}, i = 1, ..., N_{\mathrm{S}}, \tag{5.14}$$

and with μ *satisfying*

$$\sum_{i=1}^{N_{\mathrm{S}}} \left(\mu - \frac{N_{\mathrm{S}}}{\rho \left[\overline{\mathbf{\Sigma}}[k]\right]^{2}}\right)^{+} = Ns, \ k = 1, ..., K. \tag{5.15}$$

Proof: The proof is similar to Proposition 17 and is therefore omitted. □

It is worth mentioning here that most current wireless systems do not perform per subcarrier power allocation. This constraint, therefore, is not especially critical for practical systems.

An important note on the structure of the optimal hybrid precoders derived in Proposition 17 and corollary 18 is that the matrix $\mathbf{F}_{\mathrm{HP}}[k]$ representing this optimal hybrid precoders at subcarrier k can be written as

$$\mathbf{F}_{\mathrm{HP}}[k] = \mathbf{F}_{\mathrm{U}}[k]\Lambda[k], \tag{5.16}$$

where $\mathbf{F}_{\mathrm{U}}[k] = \mathbf{F}_{\mathrm{RF}}^{\star}\left(\left(\mathbf{F}_{\mathrm{RF}}^{\star}\right)^{*}\mathbf{F}_{\mathrm{RF}}^{\star}\right)^{-\frac{1}{2}} \left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}}$ is a semi-unitary matrix, as it can be verified that $\mathbf{F}_{\mathrm{U}}^{*}[k]\mathbf{F}_{\mathrm{U}}[k] = \mathbf{I}_{N_{\mathrm{S}}}$. This means that the structure of the optimal hybrid precoders is similar to that of the unconstrained SVD precoders, as it is written as a product of a semi-unitary matrix and a diagonal water-filling power allocation matrix.

5.5.2 Unitary Power Constraint

For limited feedback MIMO systems, the unitary power constraint which requires the columns of the precoding matrix $\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]$ to be orthogonal with equal power, is an alternative important constraint. Even though some performance loss should be expected with unitary constraints compared with the more relaxed total power constraint, unitary constraints usually lead to more efficient codebooks and codeword selection algorithms for limited feedback systems [70]. Further, they normally offer a close performance to the total power constraint [70]. In this subsection, we investigate the optimal hybrid precoding design under a unitary power constraint, and conclude important results for limited feedback hybrid precoding.

Similar to (5.8), the optimal mutual information with hybrid precoding under the unitary power constraint can be written in the following outer-inner problems form

$$\mathcal{I}^{\star}_{\mathrm{HP}} = \max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{J}_{\mathrm{RF}}} \left\{ \begin{array}{ll} \max\limits_{\{\mathbf{F}[k]\}_{k=1}^{K}} & \mathcal{I}\left(\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}\right) \\ \text{s.t.} & \mathbf{F}_{\mathrm{RF}}\mathbf{F}[k] \in \mathcal{U}_{N_{\mathrm{BS}} \times N_{\mathrm{S}}}, \quad k = 1, 2, ..., K. \end{array} \right\} \tag{5.17}$$

Given an RF precoder $\mathbf{F}_{\mathrm{RF}}$, we find, in the following proposition, the optimal baseband precoders of the inner problem of (5.17).

Proposition 19 *Define the SVD decompositions of the kth subcarrier channel matrix $\mathbf{H}[k]$ and the matrix $\mathbf{\Sigma}[\boldsymbol{k}]\mathbf{V}^{*}[k]\left(\mathbf{F}^{*}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}$ as in Proposition 17, then the baseband precoders $\{\mathbf{F}[k]\}_{k=1}^{K}$ that solve the inner optimization problem of (5.17) are given by*

$$\mathbf{F}^{\star}[k] = \left(\mathbf{F}^{*}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}}, \quad k = 1, 2, ..., K. \tag{5.18}$$

Proof: The proof is similar to that in Appendix A in [34], and is skipped due to space limitations. □

Given the optimal baseband precoder in (5.18), the optimal hybrid precoding based mutual information with the RF codebook $\mathcal{F}_{\mathrm{RF}}$ and the unitary power constraint can be written as

$$\mathcal{I}^{\star}_{\mathrm{HP}} = \max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_{N_{\mathrm{S}}} + \frac{\rho}{N_{\mathrm{S}}} \left[\overline{\mathbf{\Sigma}}[k] \right]^2_{1:N_{\mathrm{S}},1:N_{\mathrm{S}}} \right|, \tag{5.19}$$

where $\overline{\mathbf{\Sigma}}[k]^2$ depends only on $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{H}[k]$ as defined in Proposition 17. Next, we state an important remark on the structure of the optimal hybrid precoding design.

Remark 1. *The optimal baseband precoder* $\mathbf{F}^{\star}[k]$ *under the unitary hybrid precoding power constraint* $\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k] \in \mathcal{U}_{N_{\mathrm{BS}} \times N_{\mathrm{S}}}$ *is decomposed as* $\mathbf{F}^{\star}[k] = \mathbf{A}_{\mathrm{RF}}\mathbf{G}^{\star}[k]$*, where* $\mathbf{A}_{\mathrm{RF}} = (\mathbf{F}^{*}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}})^{-\frac{1}{2}}$ *depends only on the RF precoder, and* $\mathbf{G}^{\star}[k]$*, which we call the equivalent baseband precoder, is a semi-unitary matrix,* $\mathbf{G}^{\star}[k] \in \mathcal{U}_{N_{\mathrm{RF}} \times N_{\mathrm{S}}}$*, with the optimal design described in* (5.18).

Remark 1 shows that for the BS to achieve the optimal mutual information with the unitary power constraint and RF codebook $\mathcal{F}_{\mathrm{RF}}$, it needs to know (i) the index of the RF precoder codeword that solves (5.19), and (ii) the optimal *semi-unitary* equivalent baseband precoding matrix $\mathbf{G}^{\star}[k]$.

Although an exhaustive search over the RF codebook is still required to find the optimal mutual information in (5.12) and (5.19), the derived results are useful for several reasons. First, equations (5.12) and (5.19) provide, for the first time, the maximum achievable rate with hybrid precoding for any given RF codebooks. Therefore, these equations give a benchmark that can be used to evaluate the performance

of any heuristic/iterative hybrid analog/digital precoding algorithms, and to estimate how much additional improvement is possible. Further, the optimal mutual information in (5.12) and (5.19), depend only on the RF codebook, which will help in the design of the RF codebook as we will see in Section 5.6. Another useful finding is the special construction of the optimal baseband precoder described in Remark 1 which offers insights into the limited feedback hybrid precoding design as will be described in Section 5.6.

For the remaining part of this chapter, we will focus on the hybrid precoding design problem with the unitary constraint. In the next section, we will address the design of hybrid precoding codebooks for limited feedback wideband mmWave systems. Then, in Section 5.7, we will develop a greedy frequency selective hybrid precoding algorithm based on Gram-Schmidt orthogonalization, that relaxes the exhaustive search requirement over the RF codebook in (5.19) while providing a near-optimal performance. It is worth noting here that as shown in (5.16), the optimal hybrid precoders under total power constraints consist of a semi-unitary matrix multiplied by a diagonal water-filling power allocation matrix. Therefore, the hybrid precoding codebooks and codeword selection algorithms that will be designed under unitary power constraints can also be used for hybrid precoding under total power constraints. The water=filling power allocation can be done as a subsequent step to further improve the performance.

5.6 Codebook Design for Frequency Selective Hybrid Precoding

In this section, we consider the wideband mmWave system model in Section 5.3 with limited feedback, and develop hybrid analog and digital codebooks. First, we will consider the case $N_{\mathrm{S}} = N_{\mathrm{RF}}$ in Section 5.6.1 where we leverage the structure of the optimal hybrid precoders developed in Section 5.5 to show that the hybrid codebook design problem can be reduced to an RF codebook design problem. Then, we consider the case $N_{\mathrm{S}} < N_{\mathrm{RF}}$ in Section 5.6.2, where we develop hybrid analog and digital precoding codebooks leveraging the results in Section 5.6.1.

5.6.1 Case 1: $N_{\mathrm{S}} = N_{\mathrm{RF}}$

Given the optimal baseband precoder structure from (5.18), the optimal hybrid precoding based mutual information when $N_{\mathrm{S}} = N_{\mathrm{RF}}$, and with the RF codebook $\mathcal{F}_{\mathrm{RF}}$ can be written as $\mathcal{I}^{\star}_{\mathrm{HP}} =$

$$\max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \log_2 \Bigg| \mathbf{I}_{r(\mathbf{H}[k])} + \frac{\rho}{N_{\mathrm{S}}} \mathbf{\Sigma}[k]^2 \mathbf{V}^*[k] \mathbf{F}_{\mathrm{RF}} \left(\mathbf{F}_{\mathrm{RF}}^* \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \mathbf{G}[k] \mathbf{G}^*[k] \times \left(\mathbf{F}_{\mathrm{RF}}^* \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}} \mathbf{F}_{\mathrm{RF}}^* \mathbf{V}[k] \Bigg|. \quad (5.20)$$

Since $\mathbf{G}[k]$ is unitary for $N_S = N_{\mathrm{RF}}$, equation (5.20) can be equivalently written as

$$\mathcal{I}^{\star}_{\mathrm{HP}} = \max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_{r(\mathbf{H}[k])} + \frac{\rho}{N_{\mathrm{S}}} \mathbf{\Sigma}[k]^2 \mathbf{V}^*[k] \mathbf{F}_{\mathrm{RF}} \left(\mathbf{F}_{\mathrm{RF}}^* \mathbf{F}_{\mathrm{RF}}\right)^{-1} \mathbf{F}_{\mathrm{RF}} \mathbf{V}[k] \right|. \quad (5.21)$$

As a result, the optimal mutual information is invariant to the *optimal* equivalent baseband precoder, and depends only on the knowledge of the RF precoder $\mathbf{F}_{\mathrm{RF}}$. This

leads to the following remark.

Remark 2. *With $N_{\mathrm{S}} = N_{\mathrm{RF}}$, feeding back only the index of the optimal RF precoder that solves* (5.21) *is sufficient to achieve the optimal mutual information with limited feedback hybrid precoding.*

Remark 2 also means that no quantization of the baseband precoder is required when $N_{\mathrm{S}} = N_{\mathrm{RF}}$. Further, optimizing the limited feedback hybrid precoding performance is achieved by the optimization of the RF codebook design $\mathcal{F}_{\mathrm{RF}}$, which is addressed in the remaining part of this subsection.

RF Codebook Design Criterion: Our objective is to design the RF codebook to minimize the distortion given by the average mutual information loss of hybrid precoding compared with the optimal unconstrained per-subcarrier SVD solution. Denoting the SVD of the RF precoder as $\mathbf{F}_{\mathrm{RF}} = \mathbf{U}_{\mathrm{RF}}\mathbf{\Sigma}_{\mathrm{RF}}\mathbf{V}_{\mathrm{RF}}^{*}$, the optimal mutual information with limited feedback hybrid precoding in (5.21) can be written as

$$\mathcal{I}_{\mathrm{HP}}^{\star} = \max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_{r(\mathbf{H}[k])} + \frac{\rho}{N_{\mathrm{S}}} \mathbf{\Sigma}[k]^2 \mathbf{V}^{*}[k] \mathbf{U}_{\mathrm{RF}} \mathbf{U}_{\mathrm{RF}}^{*} \mathbf{V}[k] \right|. \tag{5.22}$$

For large mmWave MIMO systems, a reasonable assumption as stated in [20] for narrowband channels is that the hybrid precoders can be made sufficiently close to the dominant channel eigenspace. Further, the dominant channel eigenspaces of the different subcarriers may have high correlation at mmWave channels [21, 129]. This means that the eigenvalues of the matrix $\mathbf{I} - \tilde{\mathbf{V}}^{*}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\mathbf{F}^{*}[k]\mathbf{F}_{\mathrm{RF}}^{*}\tilde{\mathbf{V}}[k]$ can be made sufficiently small. Using this assumption, which will also be evaluated by simulations in Fig. 5.4-Fig. 5.6, and following similar steps to that in equations (12)-(14) of [20],

$\mathcal{I}^{\star}_{\mathrm{HP}}$ can be approximately written as

$$\mathcal{I}^{\star}_{\mathrm{HP}} \approx \max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \left(\log_2 \left| \mathbf{I}_{N_{\mathrm{RF}}} + \frac{\rho}{N_{\mathrm{S}}} \tilde{\mathbf{\Sigma}}[k]^2 \right| - \left(N_{\mathrm{RF}} - \left\| \mathbf{U}^{*}_{\mathrm{RF}} \tilde{\mathbf{V}}[k] \right\|^2_F \right) \right), \quad (5.23)$$

$$= \frac{1}{K} \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_{N_{\mathrm{RF}}} + \frac{\rho}{N_{\mathrm{S}}} \tilde{\mathbf{\Sigma}}[k]^2 \right| - \min_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \left(N_{\mathrm{RF}} - \left\| \mathbf{U}^{*}_{\mathrm{RF}} \tilde{\mathbf{V}}[k] \right\|^2_F \right), \quad (5.24)$$

where $\tilde{\mathbf{\Sigma}}[k] = [\mathbf{\Sigma}[k]]_{1:N_{\mathrm{RF}},1:N_{\mathrm{RF}}}$ and $\tilde{\mathbf{V}}[k] = [\mathbf{V}[k]]_{:,1:N_{\mathrm{RF}}}$.

When fully digital unconstrained precoding with perfect channel knowledge is possible, the optimal mutual information with per-subcarrier unitary constraint is achieved by per-subcarrier SVD precoding and is equal to

$$\mathcal{I}^{\star}_{\mathrm{UC}} = \frac{1}{K} \sum_{k=1}^{K} \log_2 \left| \mathbf{I}_{N_{\mathrm{RF}}} + \frac{\rho}{N_{\mathrm{S}}} \tilde{\mathbf{\Sigma}}[k]^2 \right|. \quad (5.25)$$

We can now define the distortion due to limited feedback hybrid precoding with the RF precoder $\mathcal{F}_{\mathrm{RF}}$ as

$$\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}\right) = \mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}} \left[\mathcal{I}^{\star}_{\mathrm{UC}} - \mathcal{I}^{\star}_{\mathrm{HP}} \right], \quad (5.26)$$

$$\approx \mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}} \left[\min_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \left(N_{\mathrm{RF}} - \left\| \mathbf{U}^{*}_{\mathrm{RF}} \tilde{\mathbf{V}}[k] \right\|^2_F \right) \right], \quad (5.27)$$

$$= \mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}} \left[\min_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} d^2_{\mathrm{chord}} \left(\mathbf{U}_{\mathrm{RF}}, \tilde{\mathbf{V}}[k] \right) \right], \quad (5.28)$$

$$= \mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}} \left[\min_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \Phi_{\mathrm{chord}} \left(\mathbf{U}_{\mathrm{RF}}, \left\{ \tilde{\mathbf{V}}[k] \right\}_{k=1}^{K} \right) \right], \quad (5.29)$$

where $d_{\mathrm{chord}}\left(\mathbf{X}, \mathbf{Y}\right)$ is the chordal distance between the two points $\mathbf{X}, \mathbf{Y}$ on the Grassmann manifold $\mathcal{G}\left(N_{\mathrm{BS}}, N_{\mathrm{RF}}\right)$. $\Phi_{\mathrm{chord}}\left(\mathbf{X}, \{\mathbf{Y}[k]\}_{k=1}^{K}\right)$ is the average of the squared chordal distances between the Grassmann points $\mathbf{X}$ and $\{\mathbf{Y}[k]\}_{k=1}^{K}$. If no constraints

are imposed on $\mathbf{X}$, the solution of $\arg\min_{\mathbf{X}\in\mathcal{G}(N_{\mathrm{BS}},N_{\mathrm{RF}})} \Phi_{\mathrm{chord}}\left(\mathbf{X}, \{\mathbf{Y}[k]\}_{k=1}^{K}\right)$ is given by the Karcher mean of the K N_{RF}-dimensional subspaces defined by the points $\{\mathbf{Y}[k]\}_{k=1}^{K}$ [130]. Our RF codebook design criterion is then to minimize the distortion function expression in (5.29)

RF Codebook Construction: Developing a closed-form solution for the RF precoders codebook that minimizes the distortion in (5.29) is non-trivial for two main reasons. First, the RF hardware constraints like the constant modulus limitation on the entries of the RF precoding matrix and the angle quantization of the phase shifters, which impose non-convex constraints on the distortion function minimization problem. Second, the lack of knowledge about the closed-form distributions of the mmWave channel matrices. These closed-form distributions usually play a key role in constructing the precoders codebook. For example, the uniform distribution of the dominant singular vectors of the IID complex Gaussian MIMO channels led to the codebook design based on isotropic packing of the Grassmann manifold [70, 131].

To overcome these challenges, we developed Algorithm 4 which is a Lloyd-type algorithm [132, 133], that first constructs a precoders codebook to minimize the distortion function in (5.29) for wideband mmWave channels while neglecting the RF hardware constraints. Then, the RF precoding codewords are designed to minimize the additional distortion results from the RF hardware constraints. One advantage of developing a Lloyd-type algorithm is that no knowledge about the closed-form distributions of the channel matrices is required, and only the knowledge of the mmWave channel parameter statistics, which are given by measurements [125], is needed. These parameter statistics are used to generate random channel realizations,

Algorithm 4 RF Codebook Construction for Frequency Selective Hybrid Precoding

1) **Initialization:** Generate random initial centroid points $\mathcal{F}_{\mathrm{U}} = \left\{\mathbf{F}_1^{\mathrm{U}}, ..., \mathbf{F}_{N_{\mathrm{CB}}}^{\mathrm{U}}\right\} \subset \mathcal{U}_{N_{\mathrm{BS}} \times N_{\mathrm{S}}}$.

2) **Source:** Generate random mmWave channels according to (5.4), $\mathcal{H} = \left\{\{\mathbf{H}[k]\}_{k=1}^{K}\right\}$, and construct the set of dominant right singular vectors corresponding to the generated channels $\mathcal{V}$ where

$$\mathcal{V} = \left\{ \left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K} \subset \mathcal{U}_{N_{\mathrm{BS}} \times N_{\mathrm{S}}} \middle| \tilde{\mathbf{V}}[k] = [\mathbf{V}[k]]_{:,1:N_{\mathrm{S}}}, \{\mathbf{H}[k]\}_{k=1}^{K} \in \mathcal{H} \right\}.$$

3) **Nearest Neighbor Partitioning:** Partition the set $\mathcal{V}$ into N_{CB} Voronoi cells $\{\mathcal{R}_1, ..., \mathcal{R}_{N_{\mathrm{CB}}}\}$ according to (5.30)-(5.31).

4) For each Voronoi cell n, $n = 1, ..., N_{\mathrm{CB}}$

a) **Karcher Mean Calculation:** Calculate the Karcher mean $\mathbf{M}_n$ of the points

$\left\{\left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}\right\}$ in $\mathcal{R}_n$ according to (5.33).

b) **Updating the Centroid:** Update the nth unconstrained codeword $\mathbf{F}_n^{\mathrm{U}} = \mathbf{M}_n$.

c) **RF Codeword Approximation:** Calculate the approximated RF codeword $\mathbf{F}_n^{\mathrm{RF}}$

according to (5.38).

5) Loop back to step 3) until convergence

according to (5.3), which are employed in constructing the RF precoders codebook as described in Algorithm 4.

The operation of Algorithm 4 can be summarized as follows.

- Initialization and source generation: In this step, N_{CB} initial codewords $\mathbf{F}_n^{\mathrm{U}}, n = 1, ..., N_{\mathrm{CB}}$ for the unconstrained codebook are randomly chosen from $\mathcal{U}_{N_{\mathrm{BS}} \times N_{\mathrm{S}}}$. Further, random wideband mmWave channel realizations are generated according to (5.3), (5.4) with the parameter statistics given from measurements,

e.g., [125]. For each channel realization, the K subcarrier channels $\{\mathbf{H}[k]\}_{k=1}^{K}$ are calculated, and their dominant right singular vectors are determined $\left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}$. Note that each element of $\mathcal{H}$ (and $\mathcal{V}$) is a set of K matrices for the K subcarriers.

- Nearest neighbor partitioning: In this step, the points in $\mathcal{V}$ are partitioned into N_{CB} Voronoi cells with respect to the codewords in $\mathcal{F}_{\mathrm{U}}$ to minimize the average distortion. To do that, we first define the quantization map $\mathbf{C}\left(\left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}\right)$, that determines the closest codeword in $\mathcal{F}_{\mathrm{U}}$ to $\left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}$ in terms of the average squared chordal distance $\Phi_{\mathrm{chord}}(.)$, as

$$\mathbf{C}\left(\left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}\right) = \arg\min_{\mathbf{X}\in\mathcal{F}_{\mathrm{U}}} \Phi_{\mathrm{chord}}\left(\mathbf{X}, \left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}\right). \tag{5.30}$$

  Once the codeword closest to each point in $\mathcal{V}$ is determined, these points can be partitioned into N_{CB} sets $\mathcal{R}_n, n = 1, 2, ..., N_{\mathrm{CB}}$ as follows

$$\mathcal{R}_n = \left\{ \left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K} \in \mathcal{V} \,\middle|\, \mathbf{C}\left(\left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}\right) = \mathbf{F}_n^{\mathrm{U}} \right\}. \tag{5.31}$$

- Centroid calculation: The centroid of each partition $\mathcal{R}_n$ is then derived to minimize the average distortion for this partition. Hence, the objective of this step is to calculate the new codeword $\mathbf{F}_n^{\mathrm{U}}$ that solves

$$\mathbf{F}_n^{\mathrm{U}} = \arg\min_{\mathbf{X}\in\mathcal{U}_{N_{\mathrm{BS}}\times N_{\mathrm{S}}}} \mathbb{E}\left[\Phi_{\mathrm{chord}}\left(\mathbf{X}, \left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}\right) \,\middle|\, \left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K} \in \mathcal{R}_n\right]. \tag{5.32}$$

  Minimizing the objective function of the problem in (5.32) is similar to minimizing the function $\Phi_{\mathrm{chord}}\left(\mathbf{X}, \left\{\tilde{\mathbf{V}}[k]\right\}_{k=1}^{K}\right)$, whose solution is found to be given

by the Karcher mean [133, 134]. Therefore, the new centroid of (5.32) can be calculated in a closed form as

$$\mathbf{M}_n = \mathrm{eig}_{1:N_\mathrm{S}} \left(\sum_{\mathcal{R}_n} \sum_{k=1}^{K} \tilde{\mathbf{V}}[k] \tilde{\mathbf{V}}^*[k] \right), \tag{5.33}$$

where $\mathrm{eig}_{1:N_\mathrm{S}}(\mathbf{X})$ represents the first N_S eigenvectors of the matrix $\mathbf{X}$ corresponding to the N_S largest eigenvalues.

- RF codewords approximation: The final objective of Algorithm 4 is to construct an RF codebook $\mathcal{F}_\mathrm{RF}$ that minimizes the distortion in (5.29). Using the triangle inequality on the chordal distances [135], the additional distortion due to the RF hardware constraints can be bounded by

$$\Phi_\mathrm{chord} \left(\mathbf{U}_n^\mathrm{RF}, \left\{ \tilde{\mathbf{V}}[k] \right\}_{k=1}^{K} \right) - \Phi_\mathrm{chord} \left(\mathbf{F}_n^\mathrm{U}, \left\{ \tilde{\mathbf{V}}[k] \right\}_{k=1}^{K} \right) \leq d_\mathrm{chord}^2 \left(\mathbf{F}_n^\mathrm{U}, \mathbf{U}_n^\mathrm{RF} \right), \tag{5.34}$$

where $\mathbf{U}_n^\mathrm{RF}$ is the N_S dominant left singular vectors of the nth RF codeword. As the chordal distance between two Grassmannian points $\mathbf{X}, \mathbf{Y} \in \mathcal{U}_{N_\mathrm{BS} \times N_\mathrm{S}}$ is invariant to the right multiplication of any of them by a unitary matrix in $\mathcal{U}_{N_\mathrm{S} \times N_\mathrm{S}}$, then we have

$$d_\mathrm{chord}^2 \left(\mathbf{F}_n^\mathrm{U}, \mathbf{U}_n^\mathrm{RF} \right) = d_\mathrm{chord}^2 \left(\mathbf{F}_n^\mathrm{U}, \mathbf{U}_n^\mathrm{RF} \mathbf{V}_n^\mathrm{RF} \right) = d_\mathrm{chord}^2 \left(\mathbf{F}_n^\mathrm{U}, \mathbf{F}_n^\mathrm{RF} \left(\mathbf{F}_n^{\mathrm{RF}^*} \mathbf{F}_n^\mathrm{RF} \right)^{-\frac{1}{2}} \right). \tag{5.35}$$

So, our objective is to solve

$$\mathbf{F}_n^{\mathrm{RF}^\star} = \arg \min_{|\mathbf{F}_n^\mathrm{RF}|_{p,q}=1} d_\mathrm{chord}^2 \left(\mathbf{F}_n^\mathrm{U}, \mathbf{F}_n^\mathrm{RF} \left(\mathbf{F}_n^{\mathrm{RF}^*} \mathbf{F}_n^\mathrm{RF} \right)^{-\frac{1}{2}} \right). \tag{5.36}$$

Finding the exact solution of (5.36) is non-trivial because of the constant-modulus constraint on the entries of $\mathbf{F}_n^{\mathrm{RF}}$. For the sake of a closed-form approximated solution, however, we make the following two approximations, that will be shown by simulations in Section 5.8 to give very good results compared with the optimal unconstrained solution. (i) For large mmWave MIMO channels, the columns of $\mathbf{F}_n^{\mathrm{RF}}$ can be chosen to be nearly orthogonal, i.e., $\left(\mathbf{F}_n^{\mathrm{RF}^*}\mathbf{F}_n^{\mathrm{RF}}\right) \approx \mathbf{I}$. (ii) The hybrid precoding and the unconstrained points $\mathbf{F}_n^{\mathrm{RF}}\left(\mathbf{F}_n^{\mathrm{RF}^*}\mathbf{F}_n^{\mathrm{RF}}\right)^{-\frac{1}{2}}$ and $\mathbf{F}_n^{\mathrm{U}}$ can be made very close [20]. Hence, by leveraging the locally Euclidean property of the Grassmann manifold, the chordal distance in (5.36) can be replaced by the Euclidean distance [20, 136]. Therefore, minimizing the distortion in (5.36) is approximately equal to the following problem

$$\mathbf{F}_n^{\mathrm{RF}^\star} = \arg \min_{|\mathbf{F}_n^{\mathrm{RF}}|_{p,q}=1} \left\|\mathbf{F}_n^{\mathrm{U}} - \mathbf{F}_n^{\mathrm{RF}}\right\|_F^2 . \tag{5.37}$$

The problem in (5.37) is a per-entry optimization problem, of which the optimal solution is given by

$$\left[\mathbf{F}_n^{\mathrm{RF}^\star}\right]_{p,q} = e^{j\measuredangle\left(\left[\mathbf{F}_n^{\mathrm{U}}\right]_{p,q}\right)}, \tag{5.38}$$

where the angle $\measuredangle\left(\left[\mathbf{F}_n^{\mathrm{U}}\right]_{p,q}\right)$ can then be approximated to the closest quantized angle of the available phase shifters.

The convergence of Algorithm 4 is shown in Fig. 5.2 for a 32×16 mmWave system with $N_{\mathrm{RF}} = 3$ RF chains, and for a codebook size 128. The monotonic convergence of Algorithm 4 to a local optimal solution is guaranteed as the precoding codewords are updated in each iteration according to the nearest neighbor and

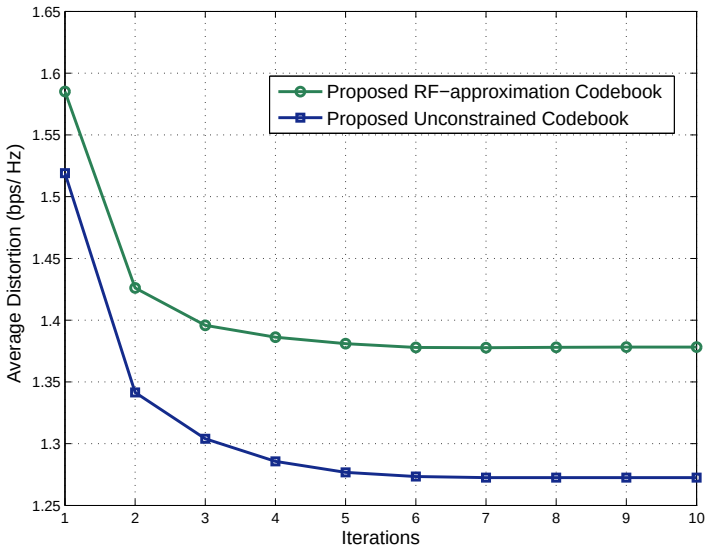


Figure 5.2: Average distortion in (5.29) of the proposed RF codebook constructed using Algorithm 4 with $N_{\mathrm{S}} = N_{\mathrm{RF}} = 3$, and codebook size $N_{\mathrm{CB}} = 128$. The rest of the channel and simulation parameters are similar to Fig. 5.4(a) described in Section 5.8. The figure shows the convergence of the unconstrained and RF approximated codebooks to small distortion values.

centroid steps (5.30) - (5.32) to make an additional reduction in the distortion function [132]. In the next subsection, we extend the developed codebook to the case when $N_{\mathrm{S}} < N_{\mathrm{RF}}$.

5.6.2 Case 2: $N_{\mathrm{S}} < N_{\mathrm{RF}}$

When $N_{\mathrm{S}} < N_{\mathrm{RF}}$, we can see from (5.19) that the optimal hybrid precoding based mutual information depends on the value of the equivalent baseband precoders, and are not invariant with respect to them because they will not have a unitary structure as when $N_{\mathrm{S}} = N_{\mathrm{RF}}$. Hence, both the RF and baseband precoders $\mathbf{F}_{\mathrm{RF}}, \{\mathbf{F}[k]\}_{k=1}^{K}$,

need to be quantized and fed back to the transmitter in this case. Inspired by the optimal structure of the baseband precoders in (5.18) and by Remark 1, we propose to quantize the equivalent baseband precoders $\{\mathbf{G}[k]\}$ instead of the baseband precoders. In addition to the intuitive good performance expected to be achieved with equivalent baseband quantization thanks to following the optimal precoders structure, one main advantage of equivalent baseband precoders quantization appears in the favorable structure of the optimal equivalent baseband codebooks as will be discussed shortly. With RF and equivalent baseband precoders quantization, the optimal mutual information is given by

$$\begin{aligned}\mathcal{I}^{\star}_{\mathrm{HP}} = \max_{\mathbf{F}_{\mathrm{RF}},\{\mathbf{F}[k]\}_{k=1}^{K}} \frac{1}{K}\sum_{k=1}^{K}\log_2\Bigg|\mathbf{I}_{N_{\mathrm{MS}}} + \frac{\rho}{N_{\mathrm{S}}}\mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}\left(\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}\mathbf{G}[k]\mathbf{G}^{*}[k] \\ \times\left(\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{H}^{*}[k]\Bigg| \\ \text{s.t.} \quad \mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \\ \mathbf{G}[k] \in \mathcal{G}_{\mathrm{BB}} \subseteq \mathcal{U}_{N_{\mathrm{RF}}\times N_{\mathrm{S}}}, \quad k = 1,2,...,K,\end{aligned} \tag{5.39}$$

where the constraint $\mathcal{G}_{\mathrm{BB}} \subseteq \mathcal{U}_{N_{\mathrm{RF}}\times N_{\mathrm{S}}}$ on the equivalent baseband precoders codebook $\mathcal{G}_{\mathrm{BB}}$ follows from the unitary power constraint on the hybrid precoders, which requires the equivalent baseband precoders to have a unitary structure. Before delving into the design of RF precoders codebook $\mathcal{F}_{\mathrm{RF}}$ and the equivalent baseband precoders codebook $\mathcal{G}_{\mathrm{BB}}$, we make the following remark on the codebook structure of the optimal equivalent baseband precoders.

Remark 3. *Regardless of the RF codebook, the optimal codebook for the equivalent baseband precoders* $\{\mathbf{G}[k]\}_{k=1}^{K}$ *under a unitary hybrid precoding constraint is unitary.*

In the remaining part of this subsection, we present the proposed design and construction of the RF and equivalent baseband precoders codebooks, $\mathcal{F}_{\mathrm{RF}}$ and $\mathcal{G}_{\mathrm{BB}}$.

Hybrid Codebook Design Criterion: The objective now is to design $\mathcal{F}_{\mathrm{RF}}$ and $\mathcal{G}_{\mathrm{BB}}$ to minimize the distortion function $\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}, \mathcal{G}_{\mathrm{BB}}\right)$ defined as the average mutual information loss of limited feedback frequency selective hybrid precoding compared with the unconstrained perfect channel knowledge solution. Formally, the distortion function $\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}, \mathcal{G}_{\mathrm{BB}}\right)$ is written as

$$\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}, \mathcal{G}_{\mathrm{BB}}\right) = \mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}}\left[\mathcal{I}_{\mathrm{UC}}^{\star} - \mathcal{I}_{\mathrm{HP}}^{\star}\right], \tag{5.40}$$

where $\mathcal{I}_{\mathrm{UC}}^{\star}$ and $\mathcal{I}_{\mathrm{HP}}^{\star}$ are as defined in (5.25) and (5.39), respectively.

The main challenge of this distortion function is that the hybrid precoding mutual information depends on the joint RF and equivalent baseband precoders codebooks as shown in (5.39), which makes the direct design of these codebooks to minimize the distortion in (5.40) non-trivial. Next, we leverage the optimal baseband prcoders structure in Section 5.5 to derive an upper bound on the limited feedback hybrid precoding distortion in (5.40). This bound will attempt to decouple the distortion impact of the RF and equivalent baseband precoding codebooks, and therefore simplify the hybrid codebook design problem. The limited feedback hybrid precoding distortion $\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}, \mathcal{G}_{\mathrm{BB}}\right)$ in (5.40) can be written as

$$\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}, \mathcal{G}_{\mathrm{BB}}\right) = \mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}}\left[\mathcal{I}_{\mathrm{UC}}^{\star} - \mathcal{I}_{\mathrm{HP}}^{\star}\right], \tag{5.41}$$

$$\overset{(a)}{=}\mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}}\underbrace{\left[\mathcal{I}^{\star}_{\mathrm{UC}}-\max_{\mathbf{F}_{\mathrm{RF}}\in\mathcal{F}_{\mathrm{RF}}}\frac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}_{N_{\mathrm{S}}}+\frac{\rho}{N_{\mathrm{S}}}\left[\overline{\mathbf{\Sigma}}[k]\right]^{2}_{1:N_{\mathrm{S}},1:N_{\mathrm{S}}}\right|\right]}_{\Delta\mathcal{I}_{\mathrm{RF}}}$$
$$+\mathbb{E}_{\{\mathbf{H}[k]\}_{k=1}^{K}}\underbrace{\left[\max_{\mathbf{F}_{\mathrm{RF}}\in\mathcal{F}_{\mathrm{RF}}}\frac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}_{N_{\mathrm{S}}}+\frac{\rho}{N_{\mathrm{S}}}\left[\overline{\mathbf{\Sigma}}[k]\right]^{2}_{1:N_{\mathrm{S}},1:N_{\mathrm{S}}}\right|-\mathcal{I}^{\star}_{\mathrm{HP}}\right]}_{\Delta\mathcal{I}_{BB|RF}}, \tag{5.42}$$

$$=\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}\right)+\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}}\left|\mathcal{F}_{\mathrm{RF}}\right.\right), \tag{5.43}$$

where (a) follows by adding and subtracting the optimal hybrid precoding based mutual information with optimal equivalent baseband precoding knowledge in (5.19). The first term is therefore the average mutual information loss due to RF codebook alone, $\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}\right)$, while the second term represents the additional loss with equivalent baseband precoders quantization $\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}}\left|\mathcal{F}_{\mathrm{RF}}\right.\right)$. Exploiting the optimal baseband precoders design in (5.18), we can bound mutual information loss due to RF quantization as

$$\Delta\mathcal{I}_{\mathrm{RF}}\overset{(a)}{=}\frac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}+\frac{\rho}{N_{\mathrm{S}}}\tilde{\mathbf{\Sigma}}[k]^2\right|$$
$$-\max_{\mathbf{F}_{\mathrm{RF}}\in\mathcal{F}_{\mathrm{RF}}}\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{S}}}\log_2\left(1+\frac{\rho}{N_{\mathrm{S}}}\lambda_i\left(\mathbf{\Sigma}[k]\mathbf{V}^{*}[k]\mathbf{U}_{\mathrm{RF}}\mathbf{V}^{*}_{\mathrm{RF}}\mathbf{V}_{\mathrm{RF}}\mathbf{U}^{*}_{\mathrm{RF}}\mathbf{V}[k]\mathbf{\Sigma}^{*}[k]\right)\right), \tag{5.44}$$

$$\overset{(b)}{\leq}\frac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}+\frac{\rho}{N_{\mathrm{S}}}\tilde{\mathbf{\Sigma}}[k]^2\right|$$
$$-\max_{\mathbf{F}_{\mathrm{RF}}\in\mathcal{F}_{\mathrm{RF}}}\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{N_{\mathrm{S}}}\log_2\left(1+\frac{\rho}{N_{\mathrm{S}}}\lambda_i\left(\tilde{\mathbf{\Sigma}}[k]\tilde{\mathbf{V}}^{*}[k]\mathbf{U}_{\mathrm{RF}}\mathbf{U}^{*}_{\mathrm{RF}}\tilde{\mathbf{V}}[k]\tilde{\mathbf{\Sigma}}^{*}[k]\right)\right), \tag{5.45}$$
$$=\frac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}+\frac{\rho}{N_{\mathrm{S}}}\tilde{\mathbf{\Sigma}}[k]^2\right|$$

$$- \max_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \log_2 \left| \mathbf{I} + \frac{\rho}{N_{\mathrm{S}}} \tilde{\mathbf{\Sigma}}[k] \tilde{\mathbf{V}}^*[k] \mathbf{U}_{\mathrm{RF}} \mathbf{U}_{\mathrm{RF}}^* \tilde{\mathbf{V}}[k] \tilde{\mathbf{\Sigma}}^*[k] \right|, \tag{5.46}$$

$$\overset{(c)}{\approx} \min_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \overline{d}^2_{\mathrm{chord}} \left(\mathbf{U}_{\mathrm{RF}}, \tilde{\mathbf{V}}[k] \right), \tag{5.47}$$

$$= \min_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \overline{\Phi}_{\mathrm{chord}} \left(\mathbf{U}_{\mathrm{RF}}, \left\{ \tilde{\mathbf{V}}[k] \right\}_{k=1}^{K} \right), \tag{5.48}$$

where (a) follows from the design of the optimal baseband precoder in (5.18). The bound in (b) follows by considering only the N_{S} dominant right singular vectors of the channel, i.e., the first N_{S} columns of $\mathbf{V}[k]$, and (c) follows by considering the large mmWave MIMO approximations used in (5.23). In (c), $\overline{d}_{\mathrm{chord}}$ is the generalized chordal distance between subspaces of different dimensions defined as [137]

$$\overline{d}^2_{\mathrm{chord}} \left(\mathbf{U}_{\mathrm{RF}}, \tilde{\mathbf{V}}[k] \right) = \min \left(N_{\mathrm{RF}}, N_{\mathrm{S}} \right) - \left\| \mathbf{U}_{\mathrm{RF}}^* \tilde{\mathbf{V}}[k] \right\|_F^2, \tag{5.49}$$

where the dimensions of $\mathbf{U}_{\mathrm{RF}}$ and $\tilde{\mathbf{V}}[k]$ are $N_{\mathrm{BS}} \times N_{\mathrm{RF}}$ and $N_{\mathrm{BS}} \times N_{\mathrm{S}}$, respectively. Finally, $\overline{\Phi}_{\mathrm{chord}}(.)$ is defined as in (5.29), but with respect to the generalized chordal distance $\overline{d}_{\mathrm{chord}}(.)$. Given the result in (5.48), we reach the following bound on $\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}\right)$

$$\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}\right) \leq \mathbb{E}\left[\min_{\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}} \frac{1}{K} \sum_{k=1}^{K} \overline{\Phi}_{\mathrm{chord}} \left(\mathbf{F}_{\mathrm{RF}}, \left\{ \tilde{\mathbf{V}}[k] \right\}_{k=1}^{K} \right) \right]. \tag{5.50}$$

Now, we derive a similar bound on the additional distortion due to the equivalent baseband quantization given a certain RF codebook $\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}} \left| \mathcal{F}_{\mathrm{RF}} \right.\right)$. Let $\mathbf{F}_{\mathrm{RF}}^{\star} \in \mathcal{F}_{\mathrm{RF}}$ be the solution of (5.19), i.e., the solution of the first term in $\Delta \mathcal{I}_{BB|RF}$ in (5.42), and $\overline{\mathbf{\Sigma}}^{\star}[k]$ be the corresponding $\overline{\mathbf{\Sigma}}[k]$. As $\mathbf{F}_{\mathrm{RF}}^{\star}$ represents a feasible (not necessarily the optimal) solution of the problem in (5.39), then $\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}} \left| \mathcal{F}_{\mathrm{RF}} \right.\right)$ in (5.43) can be

bounded as

$$\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}}\left|\mathcal{F}_{\mathrm{RF}}\right.\right) \leq \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\left(\log_2\left|\mathbf{I}_{N_{\mathrm{S}}}+\frac{\rho}{N_{\mathrm{S}}}\left[\overline{\mathbf{\Sigma}}[k]\right]^2_{1:N_{\mathrm{S}},1:N_{\mathrm{S}}}\right|\right.\right.$$
$$\left.\left.+\max_{\mathbf{G}[k]\in\mathcal{G}_{\mathrm{BB}}}\log_2\left|\mathbf{I}_{N_{\mathrm{S}}}+\frac{\rho}{N_{\mathrm{S}}}\left[\overline{\mathbf{\Sigma}}[k]\right]^2_{1:N_{\mathrm{S}},1:N_{\mathrm{S}}}\left[\overline{\mathbf{V}}[k]\right]^*_{:,1:N_{\mathrm{S}}}\mathbf{G}[k]\mathbf{G}^*[k]\left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}}\right|\right)\right], \tag{5.51}$$

$$\overset{(a)}{\approx}\mathbb{E}\left[\min_{\mathbf{G}\in\mathcal{G}_{\mathrm{BB}}}\frac{1}{K}\sum_{k=1}^{K}d_{\mathrm{chord}}\left(\mathbf{G},\left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}}\right)\right], \tag{5.52}$$

$$=\mathbb{E}\left[\min_{\mathbf{G}\in\mathcal{G}_{\mathrm{BB}}}\Phi_{\mathrm{chord}}\left(\mathbf{G},\left\{\left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{S}}}\right\}_{k=1}^{K}\right)\right], \tag{5.53}$$

where (a) follows by considering the large mmWave MIMO approximations used in (5.23). The codebook design objective is then to minimize the upper bound on the distortion function, that is given by the bounds in (5.53), (5.50) on $\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}\right)+\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}}\left|\mathcal{F}_{\mathrm{RF}}\right.\right)$.

Hybrid Codebook Construction: Given the distortion function upper bounds in (5.50) and (5.53), we will design the RF codebook $\mathcal{F}_{\mathrm{RF}}$ to minimize the derived bound on $\mathcal{D}\left(\mathcal{F}_{\mathrm{RF}}\right)$. Then, we will design the equivalent baseband codebook $\mathcal{G}_{\mathrm{BB}}$ to minimize the bound on the additional distortion of the equivalent baseband precoders quantization $\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}}\left|\mathcal{F}_{\mathrm{RF}}\right.\right)$. As the distortion bounds in (5.50) and (5.53) are similar to the expression of the RF codebook distortion in (5.29), we use Algorithm 4 to design the hybrid RF and equivalent baseband codebooks $\mathcal{F}_{\mathrm{RF}}$ and $\mathcal{G}_{\mathrm{BB}}$. For the RF codebook, Algorithm 4 will be used, but with replacing the chordal distance $d_{\mathrm{chord}}\left(.\right)$ in (5.30), (5.32), (5.34), (5.36) by the generalized chordal distance between subspaces of different dimensions in (5.48). To build the unitary equivalent baseband precoders codebook, Algorithm 4 will be also used, but without step 4-c

as no RF approximation is required. Even though the dependence of the distortion function $\mathcal{D}\left(\mathcal{G}_{\mathrm{BB}}\,|\mathcal{F}_{\mathrm{RF}}\right)$ on the RF codebook $\mathcal{F}_{\mathrm{RF}}$ is relaxed in the design, i.e., the RF and baseband codebooks are sequentially designed, the developed hybrid codebooks achieve good performance compared with the perfect channel knowledge case as will be shown in Section 5.8.

5.7 Gram-Schmidt Based Greedy Hybrid Precoding

The optimal hybrid precoding design for any given RF codebook was derived in Section 5.5. An exhaustive search over the RF codebook, however, is still required to find the optimal RF precoder in (5.19). This search may be of high complexity, especially for large antenna systems. Therefore, and inspired by the optimal baseband precoder structure in (5.18), we develop a greedy frequency selective hybrid precoding algorithm in this section based on Gram-Schmidt orthogonalization. Different from prior work that mainly depends on heuristic ideas for the joint design of the RF and baseband precoders [20, 27, 123], we will make statements on the optimality of the proposed algorithm in some cases, even though it sequentially designs the RF and baseband precoders.

Equation (5.19) showed that the optimal hybrid precoding based mutual information can be written as a function of the RF precoding alone. Hence, the remaining problem was to determine the best RF precoding matrix, i.e., the best N_{RF} beamforming vectors, from the RF precoding codebook $\mathcal{F}_{\mathrm{RF}}$. This requires making an exhaustive search over the matrices codewords in $\mathcal{F}_{\mathrm{RF}}$. A natural greedy approach to construct the hybrid precoder is to iteratively select the N_{RF} RF beamforming vectors

to maximize the mutual information. In this chapter, we call this the direct greedy hybrid precoding (DG-HP) algorithm. For simplicity of exposition, we will assume that the RF beamforming vectors of the N_{RF} RF chains are to be selected from the same vector codebook $\mathcal{F}_{\mathrm{RF}}^{\mathrm{v}} = \left\{\mathbf{f}_1^{\mathrm{RF}}, ..., \mathbf{f}_{N_{\mathrm{CB}}}^{\mathrm{RF}}\right\}$, but choosing unique codewords. Extensions to the case when each of the N_{RF} RF beamforming vectors is taken from a different codebook is straightforward.

The operation of the DG-HP algorithm then consists of N_{RF} iterations. In each iteration, the RF beamforming vector from $\mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}$ that maximizes the mutual information at this iteration will be selected. Let the $N_{\mathrm{BS}} \times (i-1)$ matrix $\mathbf{F}_{\mathrm{RF}}^{(i-1)}$ denote the RF precoding matrix at the end of the $(i-1)$th iteration. Then by leveraging the optimal baseband precoder structure in (5.18), the objective of the ith iteration is to select $\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}$ that solves

$$\mathcal{I}_{\mathrm{HP}}^{(i)} = \max_{\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}} \frac{1}{K} \sum_{k=1}^{K} \sum_{\ell=1}^{i} \log_2 \left(1 + \frac{\rho}{N_{\mathrm{S}}} \lambda_\ell \left(\mathbf{H}\left[k\right] \dot{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \left(\dot{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^*} \dot{\mathbf{F}}_{\mathrm{RF}}^{(i,n)}\right)^{-1} \dot{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^*} \mathbf{H}\left[k\right]^*\right)\right), \tag{5.54}$$

with $\dot{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} = \left[\mathbf{F}_{\mathrm{RF}}^{(i-1)}, \mathbf{f}_n^{\mathrm{RF}}\right]$. The best vector $\mathbf{f}_{n^\star}^{\mathrm{RF}}$ will be then added to the RF precoding matrix to form $\mathbf{F}_{\mathrm{RF}}^{(i)} = \left[\mathbf{F}_{\mathrm{RF}}^{(i-1)}, \mathbf{f}_{n^\star}^{\mathrm{RF}}\right]$. The achievable mutual information with this algorithm is then $\mathcal{I}_{\mathrm{HP}}^{\mathrm{DG-HP}} = \mathcal{I}_{\mathrm{HP}}^{(N_{\mathrm{RF}})}$. The main limitation of this algorithm is that it still requires an exhaustive search over $\mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}$ and eigenvalues calculation in each iteration. The objective of this section is to develop a low-complexity algorithm that has a similar (or very close) performance to this DG-HP algorithm. In the next subsection, we will make the first step towards this goal by proving that a Gram-Schmidt based algorithm can lead to exactly the same performance of the DG-HP. This will be then leveraged into the design of our algorithm in Section 5.7.2.

5.7.1 Gram-Schmidt Based Greedy Hybrid Precoding

In hybrid analog/digital precoding architectures, the effective channel seen at the baseband is through the RF precoders lens. This gives the intuition that it is better for the RF beamforming vectors to be orthogonal (or close to orthogonal), as this physically means that the effective channel will have a better coverage over the dominant subspaces belonging to the actual channel matrix. This intuition is also confirmed by the structure of the optimal baseband precoder discussed in Remark 1, as the overall matrix $\mathbf{F}_{\mathrm{RF}} \left(\mathbf{F}_{\mathrm{RF}}^{*} \mathbf{F}_{\mathrm{RF}}\right)^{-\frac{1}{2}}$ has a semi-unitary structure. Indeed, this observation can also be related to the structure of the solutions of the nearest matrix and nearest tight frame problems [138, 139]. This note means that in each iteration i of the greedy hybrid precoding algorithm in (5.54) with a selected codeword $\mathbf{f}_{n^{\star}}^{\mathrm{RF}}$, the additional mutual information gain over the previous iterations is due to the contribution of the component of $\mathbf{f}_{n^{\star}}^{\mathrm{RF}}$ that is orthogonal on the existing RF precoding matrix $\mathbf{F}_{\mathrm{RF}}^{(i-1)}$. Based on that, we modify the DG-HP algorithm by adding a Gram-Schmidt orthogonalization step in each iteration i to project the candidate beamforming codewords on the orthogonal complement of the subspace spanned by the selected codewords in $\mathbf{F}_{\mathrm{RF}}^{(i-1)}$. This can be simply done by multiplying the candidate vectors by the projection matrix $\mathbf{P}^{(i-1)^{\perp}} = \left(\mathbf{I}_{N_{\mathrm{BS}}} - \mathbf{F}_{\mathrm{RF}}^{(i-1)} \left(\mathbf{F}_{\mathrm{RF}}^{(i-1)^{*}} \mathbf{F}_{\mathrm{RF}}^{(i-1)}\right)^{-1} \mathbf{F}_{\mathrm{RF}}^{(i-1)^{*}}\right)$. Given the optimal precoder design in (5.18), the mutual information at the ith iteration of the modified Gram-Schmidt hybrid precoding (GS-HP) algorithm can be written as

$$\overline{\mathcal{I}}_{\mathrm{HP}}^{(i)} = \max_{\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}} \frac{1}{K} \sum_{k=1}^{K} \sum_{\ell=1}^{i} \log_2 \left(1 + \frac{\rho}{N_{\mathrm{S}}} \lambda_{\ell} \left(\mathbf{H}\left[k\right] \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \left(\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^{*}} \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)}\right)^{-1} \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^{*}} \mathbf{H}\left[k\right]^{*}\right)\right), \tag{5.55}$$

$$\overset{(a)}{=} \max_{\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}} \frac{1}{K} \sum_{k=1}^{K} \sum_{\ell=1}^{i} \log_2 \left(1 + \frac{\rho}{N_{\mathrm{S}}} \lambda_\ell \left(\mathbf{T}^{(i-1)} + \mathbf{H}[k] \mathbf{P}^{(i-1)^\perp} \mathbf{f}_n^{\mathrm{RF}} \mathbf{f}_n^{\mathrm{RF}^*} \mathbf{P}^{(i-1)^{\perp^*}} \mathbf{H}^*[k] \right) \right), \tag{5.56}$$

with $\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} = \left[\mathbf{F}_{\mathrm{RF}}^{(i-1)}, \mathbf{P}^{(i-1)^\perp} \mathbf{f}_n^{\mathrm{RF}} \right]$, $\mathbf{T}^{(i-1)} = \mathbf{H}[k] \mathbf{F}_{\mathrm{RF}}^{(i-1)} \left(\mathbf{F}_{\mathrm{RF}}^{(i-1)^*} \mathbf{F}_{\mathrm{RF}}^{(i-1)} \right)^{-1} \mathbf{F}_{\mathrm{RF}}^{(i-1)^*} \mathbf{H}[k]^*$. Note that $\mathbf{T}^{(i-1)}$ is a constant matrix at iteration i, and (a) follows from the Gram-Schmidt orthogonalization which allows the matrix $\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \left(\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^*} \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \right)^{-\frac{1}{2}}$ at iteration i to be written as $\left[\mathbf{F}_{\mathrm{RF}}^{(i-1)} \left(\mathbf{F}_{\mathrm{RF}}^{(i-1)^*} \mathbf{F}_{\mathrm{RF}}^{(i-1)} \right)^{-\frac{1}{2}}, \mathbf{P}^{(i-1)^\perp} \mathbf{f}_n^{\mathrm{RF}} \right]$. Hence, the eigenvalues calculation in (5.56) can be calculated as a rank-1 update of the previous iteration eigenvalues, which reduces the overall complexity [140]. The best vector $\mathbf{f}_{n^\star}^{\mathrm{RF}}$ will be then added to the RF precoding matrix to form $\mathbf{F}_{\mathrm{RF}}^{(i)} = \left[\mathbf{F}_{\mathrm{RF}}^{(i-1)}, \mathbf{f}_{n^\star}^{\mathrm{RF}} \right]$. At the end of the N_{RF} iterations, the achieved mutual information is $\mathcal{I}_{\mathrm{HP}}^{\mathrm{GS-HP}} = \overline{\mathcal{I}}_{\mathrm{HP}}^{(N_{\mathrm{RF}})}$. In the following proposition, we prove that this Gram-Schmidt hybrid precoding algorithm is exactly equivalent to the DG-HP algorithm.

Proposition 20 *The achieved mutual information of the direct greedy hybrid precoding algorithm in* (5.54) *and the Gram-Schmidt based hybrid precoding algorithm in* (5.55) *are exactly equal, i.e.,* $\mathcal{I}_{\mathrm{HP}}^{\mathrm{DG-HP}} = \mathcal{I}_{\mathrm{HP}}^{\mathrm{GS-HP}}$.

Proof: See Appendix B in [34]. □

5.7.2 Approximate Gram-Schmidt Based Greedy Hybrid Precoding

The main advantage of the Gram-Schmidt hybrid precoding design in Section 5.7.1 is that it leads to a near-optimal low-complexity design of the frequency selective hybrid precoding as will be discussed in this section. Given the optimal

Algorithm 5 Approximate Gram-Schmidt Based Frequency Selective Hybrid Precoding

Initialization

1) Construct $\mathbf{\Pi} = \tilde{\mathbf{V}}_{\mathbf{H}}\tilde{\mathbf{\Sigma}}_{\mathbf{H}}$, with $\tilde{\mathbf{\Sigma}}_{\mathbf{H}} = \text{diag}\left(\tilde{\mathbf{\Sigma}}_1, ..., \tilde{\mathbf{\Sigma}}_{\text{K}}\right)$ and $\tilde{\mathbf{V}}_{\mathbf{H}} = \left[\tilde{\mathbf{V}}_1, ..., \tilde{\mathbf{V}}_{\text{K}}\right]$. Set $\mathbf{F}_{\text{RF}} =$ Empty Matrix. Define $\mathbf{A}_{\text{CB}} = \left[\mathbf{f}_1^{\text{RF}}, ..., \mathbf{f}_{N_{\text{CB}}^{\text{v}}}^{\text{RF}}\right]$, where $\mathbf{f}_n^{\text{RF}}, n = 1, ..., N_{\text{CB}}^{\text{v}}$ are the codewords in $\mathcal{F}_{\text{RF}}$

RF Precoder Design

2) For $i, i = 1, ..., N_{\text{RF}}$

a) $\mathbf{\Psi} = \mathbf{\Pi}^* \mathbf{A}_{\text{CB}}$

b) $n^\star = \arg\max_{n=1,2,..N_{\text{CB}}^{\text{v}}} \left\| [\mathbf{\Psi}]_{:,n} \right\|_2$.

c) $\mathbf{F}_{\text{RF}}^{(i)} = \left[\mathbf{F}_{\text{RF}}^{(i-1)} \mathbf{f}_{n^\star}^{\text{RF}}\right]$

d) $\mathbf{\Pi} = \left(\mathbf{I}_{N_{\text{BS}}} - \mathbf{F}_{\text{RF}}^{(i)} \left(\mathbf{F}_{\text{RF}}^{(i)^*} \mathbf{F}_{\text{RF}}^{(i)}\right)^{-1} \mathbf{F}_{\text{RF}}^{(i)^*}\right) \mathbf{\Pi}$

Digital Precoder Design

3) $\mathbf{F}[k] = \mathbf{F}_{\text{RF}}^{(N_{\text{RF}})} \left(\mathbf{F}_{\text{RF}}^{(N_{\text{RF}})^*} \mathbf{F}_{\text{RF}}^{(N_{\text{RF}})}\right)^{-\frac{1}{2}} \left[\overline{\mathbf{V}}[k]\right]_{:,1:N_{\text{S}}}, k = 1, ..., K$, with $\overline{\mathbf{V}}[k]$ defined in (5.18)

baseband precoding solution in (5.18), the mutual information at the ith iteration in (5.55) can be written as

$$\overline{\mathcal{I}}_{\text{HP}}^{(i)} = \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}^{\text{v}}} \frac{1}{K} \sum_{k=1}^{K} \sum_{\ell=1}^{i} \log_2 \left(1 + \frac{\rho}{N_{\text{S}}} \lambda_\ell \left(\mathbf{H}\left[k\right] \overline{\mathbf{F}}_{\text{RF}}^{(i,n)} \left(\overline{\mathbf{F}}_{\text{RF}}^{(i,n)^*} \overline{\mathbf{F}}_{\text{RF}}^{(i,n)}\right)^{-1} \overline{\mathbf{F}}_{\text{RF}}^{(i,n)^*} \mathbf{H}\left[k\right]^*\right)\right) \tag{5.57}$$

$$\overset{(a)}{\geq} \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}^{\text{v}}} \frac{1}{K} \sum_{k=1}^{K} \sum_{\ell=1}^{i} \log_2 \left(1 + \frac{\rho}{N_{\text{S}}} \lambda_\ell \left(\tilde{\mathbf{\Sigma}}[k] \tilde{\mathbf{V}}^*[k] \overline{\mathbf{F}}_{\text{RF}}^{(i,n)} \left(\overline{\mathbf{F}}_{\text{RF}}^{(i,n)^*} \overline{\mathbf{F}}_{\text{RF}}^{(i,n)}\right)^{-1} \right.\right.$$
$$\left.\left. \times\ \overline{\mathbf{F}}_{\text{RF}}^{(i,n)^*} \tilde{\mathbf{V}}[k] \tilde{\mathbf{\Sigma}}^*[k]\right)\right), \tag{5.58}$$

$$\overset{(b)}{\approx} \frac{1}{K} \sum_{k=1}^{K} \left(\log_2 \left|\mathbf{I} + \frac{\rho}{N_{\text{S}}} \tilde{\mathbf{\Sigma}}[k]^2\right| - \text{tr}\left(\tilde{\mathbf{\Sigma}}[k]\right)\right) \tag{5.59}$$

$$+ \max_{\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}} \frac{1}{K} \sum_{k=1}^{K} \left\| \tilde{\mathbf{\Sigma}}[k] \tilde{\mathbf{V}}^*[k] \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \left(\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^*} \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \right)^{-\frac{1}{2}} \right\|_F^2, \tag{5.60}$$

where the bound in (a) is by considering only the first N_{S} dominant singular values of $\mathbf{H}[k]$, and (b) follows from using the large mmWave MIMO approximations used in (5.23). The objective then of the ith iteration is to select $\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}$ that solves

$$\mathbf{f}_{n^\star}^{\mathrm{RF}} = \arg\max_{\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}} \frac{1}{K} \sum_{k=1}^{K} \left\| \tilde{\mathbf{\Sigma}}[k] \tilde{\mathbf{V}}^*[k] \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \left(\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^*} \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \right)^{-\frac{1}{2}} \right\|_F^2, \tag{5.61}$$

$$\overset{(a)}{=} \left\| \tilde{\mathbf{\Sigma}}_{\mathbf{H}} \tilde{\mathbf{V}}_{\mathbf{H}}^* \mathbf{F}_{\mathrm{RF}}^{(i-1)} \left(\mathbf{F}_{\mathrm{RF}}^{(i-1)^*} \mathbf{F}_{\mathrm{RF}}^{(i-1)} \right)^{-\frac{1}{2}} \right\|_F^2 + \arg\max_{\mathbf{f}_n^{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}} \left\| \tilde{\mathbf{\Sigma}}_{\mathbf{H}} \tilde{\mathbf{V}}_{\mathbf{H}}^* \mathbf{P}^{(i-1)^\perp} \mathbf{f}_n^{\mathrm{RF}} \right\|_2^2, \tag{5.62}$$

where $\tilde{\mathbf{\Sigma}}_{\mathbf{H}} = \mathrm{diag}\left(\tilde{\mathbf{\Sigma}}[1], ..., \tilde{\mathbf{\Sigma}}[K] \right)$, $\tilde{\mathbf{V}}_{\mathbf{H}} = \left[\tilde{\mathbf{V}}[1], ..., \tilde{\mathbf{V}}[K] \right]$, and (a) follows from Gram-Schmidt which makes $\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \left(\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)^*} \overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} \right)^{-\frac{1}{2}}$, with $\overline{\mathbf{F}}_{\mathrm{RF}}^{(i,n)} = \left[\mathbf{F}_{\mathrm{RF}}^{(i-1)}, \mathbf{P}^{(i-1)^\perp} \mathbf{f}_n^{\mathrm{RF}} \right]$ at the ith iteration equals to the matrix $\left[\mathbf{F}_{\mathrm{RF}}^{(i-1)} \left(\mathbf{F}_{\mathrm{RF}}^{(i-1)^*} \mathbf{F}_{\mathrm{RF}}^{(i-1)} \right)^{-\frac{1}{2}}, \mathbf{P}^{(i-1)^\perp} \mathbf{f}_n^{(i)} \right]$. The problem in (5.62) is simple to solve with just a maximum projection step. We call this algorithm the approximate Gram-Schmidt hybrid precoding (Approximate GS-HP) algorithm. As shown in Algorithm 5, the developed algorithm sequentially build the RF and baseband precoding matrices in two separate stages. First, the RF beamforming vectors are iteratively selected to solve (5.62). Then, the baseband precoder is optimally designed according to (5.18). Despite its sequential design of the RF and baseband precoders, which reduces the complexity when compared with prior solutions that mostly depend on the joint design of the baseband and RF precoding matrices [20, 27], Algorithm 5 achieves a significant gain over prior solutions, and gives a very close performance to the optimal solution in (5.19), as will be shown in Section 5.8. In fact, for some special cases like the case when $N_{\mathrm{S}} = N_{\mathrm{RF}}$,

Algorithm 5 can be proved to provide the optimal baseband and RF precoding design of the problem $\max_{\mathbf{F}_{\mathrm{RF}}\in\mathcal{F}_{\mathrm{RF}},\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\|_F^2\leq N_{\mathrm{RF}}} \frac{1}{K}\sum_{k=1}^{K}\left\|\tilde{\mathbf{\Sigma}}[k]\,\tilde{\mathbf{V}}^*[k]\,\mathbf{F}_{\mathrm{RF}}\mathbf{F}[k]\right\|_F^2$ which has been an important optimization objective for many hybrid precoding papers [20, 27].

5.7.3 Total Feedback Overhead

In this subsection, we summarize the feedback overhead associated with the proposed hybrid precoding strategies in Section 5.6 and Section 5.7 as illustrated in Table 5.1.

In Section 5.6, we develop an algorithm to construct efficient codebooks $\mathcal{F}_{\mathrm{RF}}$ for RF precoding *matrices*. Hence, the MS will need to feedback the index of the best RF precoding codeword, i.e., $B_{\mathrm{RF}} = \log_2|\mathcal{F}_{\mathrm{RF}}|$ bits. For the baseband precoders, we found in Section 5.5 that the optimal baseband precoder can be written in terms of the RF precoder and a semi-unitary matrix (the equivalent baseband precoder). In the case when $N_{\mathrm{S}} = N_{\mathrm{RF}}$, though, this equivalent baseband precoder takes a unitary structure, and the spectral efficiency is invariant to this equivalent baseband precoder. Hence, no baseband feedback bits are needed in this case. If $N_{\mathrm{S}} < N_{\mathrm{RF}}$, $\log_2|\mathcal{G}_{\mathrm{BB}}|$ bit will be needed for each subcarrier.

In Section 5.7.2, we developed a greedy hybrid precoding algorithm that requires only a per RF beamforming *vector* codebook $\mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}$. Hence, the index of the best codeword for each RF beamforming vectors will be fed back, i.e., a total of $B_{\mathrm{RF}} = N_{\mathrm{RF}}\log_2|\mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}|$ bits. The baseband feedback bits are similar to the other scheme.

Table 5.1: Total feedback overhead with the proposed limited feedback hybrid precoding strategies

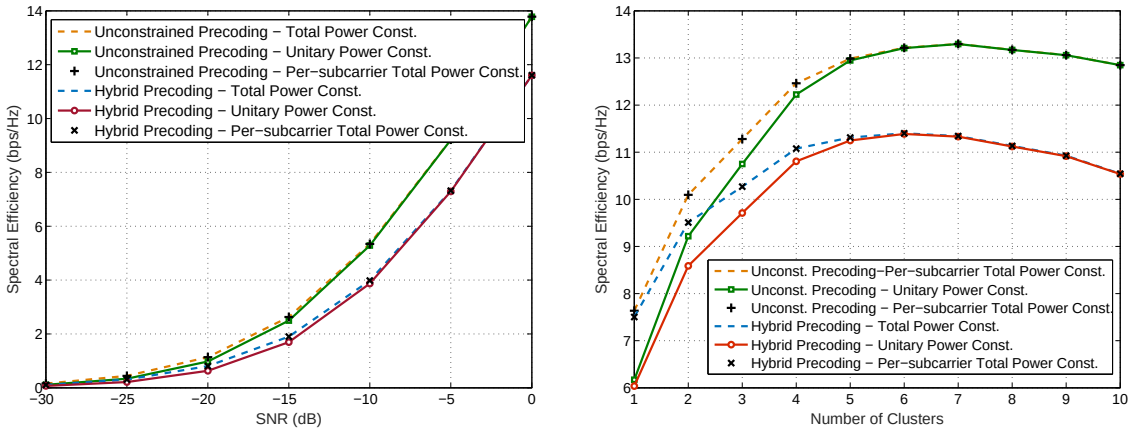
Hybrid Precoders Quantization Scheme	RF Bits B_{RF}	Baseband Bits B_{BB}	
		$N_{\mathrm{S}} = N_{\mathrm{RF}}$	$N_{\mathrm{S}} < N_{\mathrm{RF}}$
RF matrices quantization with $\mathcal{F}_{\mathrm{RF}}$ and baseband matrices quantization with $\mathcal{G}_{\mathrm{BB}}$ codebook as in Section 5.6	$\log_2 \lvert\mathcal{F}_{\mathrm{RF}}\rvert$	0	$K \log_2 \lvert\mathcal{G}_{\mathrm{BB}}\rvert$
RF vectors quantization with $\mathcal{F}^{\mathrm{v}}_{\mathrm{RF}}$ and baseband matrices quantization with $\mathcal{G}_{\mathrm{BB}}$ codebook as in Section 5.6	$N_{\mathrm{RF}} \log_2 \lvert\mathcal{F}^{\mathrm{v}}_{\mathrm{RF}}\rvert$	0	$K \log_2 \lvert\mathcal{G}_{\mathrm{BB}}\rvert$

5.8 Simulation Results

In this section, we validate our analytical results and evaluate the performance of the proposed codebooks and hybrid precoding designs using numerical simulations. We adopt the wideband mmWave channel as in (5.3)-(5.4), where a raised-cosine filter is adopted for the pulse shaping function [128], i.e., $p_{\mathrm{rc}}(t)$ is modeled as

$$p_{\mathrm{rc}}(t) = \begin{cases} \frac{\pi}{4}\ \mathrm{sinc}\left(\frac{1}{2\beta}\right), & t = \pm\frac{T_{\mathrm{s}}}{2\beta} \\ \mathrm{sinc}\left(\frac{t}{T_{\mathrm{s}}}\right) \frac{\cos\left(\frac{\pi\beta t}{T_{\mathrm{s}}}\right)}{1-\left(\frac{2\beta t}{T_{\mathrm{s}}}\right)^2}, & \text{otherwise,} \end{cases} \tag{5.63}$$

with T_{s} the sampling time and the roll-off factor $\beta = 1$. The number of clusters is assumed to be $L = 6$, and the center AoAs/AoDs of the L clusters θ_ℓ, ϕ_ℓ are assumed to be uniformly distributed in $[0, 2\pi)$. Each cluster has $R_\ell = 5$ rays with Laplacian distributed AoAs/AoDs [20, 127], and angle spread of 10^{o}. The number of system subcarriers K equals 512, and the cyclic prefix length is $D = 128$, which is similar to 802.11ad [36]. The paths delay is uniformly distributed in $[0, DT_{\mathrm{s}}]$. While the proposed algorithms and codebooks are general for large MIMO channels, we assume



(a) (b)

Figure 5.3: The performance of the optimal hybrid precoding design under different power constraints in Proposition 17, Corollary 18, and Proposition 19 versus the SNR in (a) and versus the number of channel clusters with SNR = 0 dB in (b). The adopted system model has $N_{\mathrm{BS}} = 32$ antennas, $N_{\mathrm{MS}} = 8$ antennas, and $N_{\mathrm{S}} = N_{\mathrm{RF}} = 3$.

in these simulations that both the BS and MS has a ULA with $N_{\mathrm{RF}} = 3$. Hence, $\mathbf{a}_{\mathrm{BS}}(\phi)$ is defined as

$$\mathbf{a}_{\mathrm{BS}}(\phi) = \frac{1}{\sqrt{N_{\mathrm{BS}}}}\left[1, e^{\mathrm{j}\frac{2\pi}{\lambda}d_{\mathrm{s}}\sin(\phi)}, ..., e^{\mathrm{j}(N_{\mathrm{BS}}-1)\frac{2\pi}{\lambda}d_{\mathrm{s}}\sin(\phi)}\right]^{\mathrm{T}}, \tag{5.64}$$

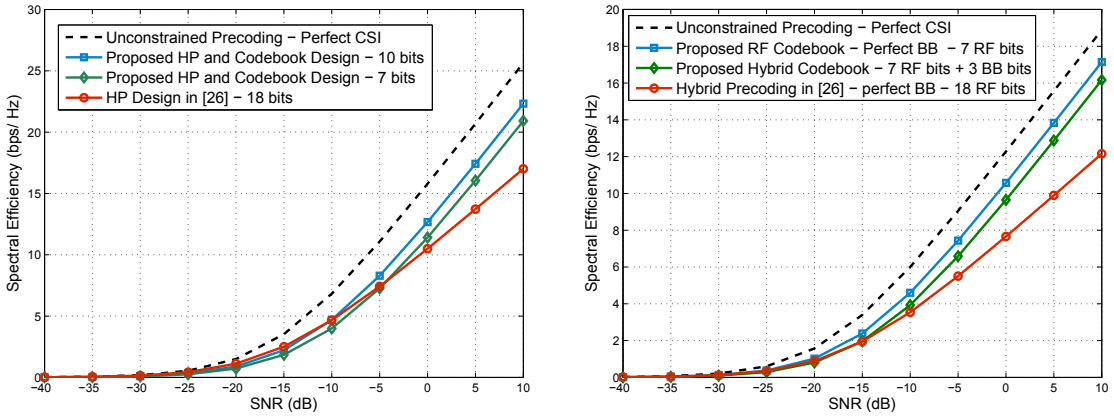
where λ is the signal wavelength, and d_{s} is the distance between antenna elements with $d_{\mathrm{s}} = \lambda/2$. The array response vectors at the MS, $\mathbf{a}_{\mathrm{MS}}(\theta)$, can be written in a similar fashion.

5.8.1 Optimal Hybrid Precoders and Codebook Designs

First, we compare the performance of the optimal hybrid precoders and the fully-digital unconstrained precoders for different power constraints in Fig. 5.3(a)-Fig. 5.3(b). For these figures, we adopt the system model in Section 5.3 with $N_{\mathrm{BS}} = 32$

antennas, $N_{\mathrm{MS}} = 8$ antennas, and $N_{\mathrm{S}} = N_{\mathrm{RF}} = 3$ streams. In Fig. 5.3(a), the spectral efficiencies of the optimal hybrid precoders under total power constraints, per-subcarrier total power constraints, and unitary power constraints are plotted, and compared with the spectral efficiencies of the SVD unconstrained precoding under the same power constraints. Fig. 5.3(a) shows that the gain of total power constraints over unitary constraints is limited, and decreases with the SNR. The same setup is adopted again in Fig. 5.3(b) where the optimal hybrid precoders under different power constraints are compared for different numbers of channel clusters, assuming that each cluster contributes with a single ray, and fixing the number of transmitted streams at $N_{\mathrm{S}} = 3$. Fig. 5.3(b) illustrates that the gain of total power constraints over unitary constraints increases when the channel is very sparse, i.e., when a very small number of clusters exist. This gain, though, is very small if the channel has more than 4-5 clusters.

Next, we evaluate the performance of the proposed hybrid precoding codebooks in Fig. 5.4, adopting the system model in Fig. 5.1 with $N_{\mathrm{BS}} = 32$ antennas, and $N_{\mathrm{MS}} = 16$ antennas. In Fig. 5.4(a), the case $N_{\mathrm{S}} = N_{\mathrm{RF}} = 3$ is considered, the RF codebook is constructed using Algorithm 4 with different sizes, and the hybrid precoders are designed according to (5.19). Fig. 5.4(a) shows that the proposed codebook improves the performance compared with the prior work in [20], even though much smaller numbers of feedback bits are needed, namely 10 and 7 bits compared with 18 bits in the case of beamsteering codebooks in [20,27]. In Fig. 5.4(b), the same setup is considered again, but with $N_{\mathrm{S}} = 2$ streams and $N_{\mathrm{RF}} = 3$ RF chains. In this case, the hybrid codebooks are constructed as explained in Section 5.6.2, with an RF



(a) (b)

Figure 5.4: The performance of the proposed hybrid codebook design in Algorithm 4, compared with the unconstrained SVD solution, and the prior work, for the case when $N_{\mathrm{S}} = N_{\mathrm{RF}} = 3$ in (a), and the case $N_{\mathrm{S}} = 2, N_{\mathrm{RF}} = 3$ in (b).

codebook of size 128 and an equivalent baseband precoders codebook of size 8. The figure shows that a very good performance can be also achieved with the designed codebook, despite the relatively small codebook sizes. Further, Fig. 5.4 shows that the proposed limited feedback hybrid precoding codebooks achieve a good slope with the SNR relative to the unconstrained with perfect channel knowledge solution.

5.8.2 Low-Complexity Gram-Schmidt Based Greedy Hybrid Precoding

In Fig. 5.5 and Fig. 5.6, we validate the result in Proposition 20, in addition to evaluating the approximate Gram-Schmidt based hybrid precoding algorithm. In Fig. 5.5, the same setup of Fig. 5.4(a) is adopted, and the hybrid precoders are greedily constructed using the direct greedy hybrid precoding algorithm in (5.54), the Gram-Schmidt hybrid precoding in (5.55), and the low-complexity approximate Gram-Schmidt hybrid precoding design in Algorithm 5. The spectral efficiencies

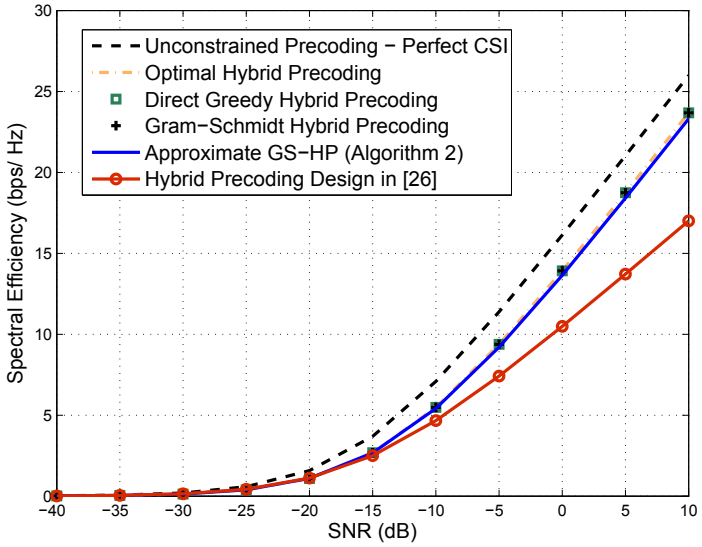


Figure 5.5: The performance of the approximate Gram-Schmidt hybrid precoding design in Algorithm 5 compared with the optimal hybrid precoding solution in (5.19), the unconstrained SVD solution, and the prior work. The system has $N_{\mathrm{BS}} = 32$ antennas, $N_{\mathrm{MS}} = 16$ antennas, and $N_{\mathrm{S}} = N_{\mathrm{RF}} = 3$.

achieved by these greedy algorithms are compared with the optimal hybrid precoding design in (5.19) where the RF precoders are selected through an exhaustive search over the RF beamforming vectors codebook. The rates are also compared with the prior solution in [20]. For a fair comparison, we assume that each RF beamforming vector is selected from a beamsteering codebook with a size $N_{\mathrm{CB}}^{\mathrm{v}} = 64$. First, Fig. 5.5 shows that the direct greedy and Gram-Schmidt based hybrid precoding algorithms achieve exactly the same performance which verifies Proposition 20. Their performance is also shown to be almost equal to the optimal solution given by (5.19). Despite its low-complexity, the developed approximate Gram-Schmidt hybrid precoding design in Algorithm 5 achieves a very close performance to the optimal solution. We emphasize

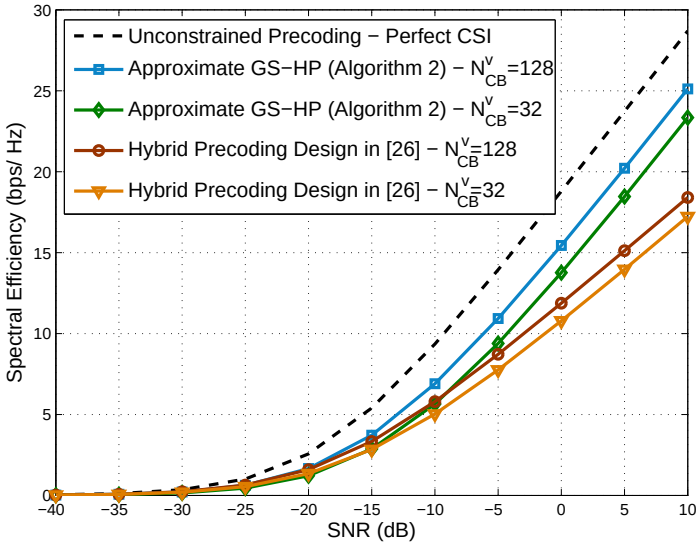


Figure 5.6: The performance of the approximate Gram-Schmidt hybrid precoding design in Algorithm 5 for different codebook sizes, compared with the unconstrained SVD solution, and the prior work. The system has $N_{\mathrm{BS}} = 64$ antennas, $N_{\mathrm{MS}} = 16$ antennas, and $N_{\mathrm{S}} = N_{\mathrm{RF}} = 3$.

here that any hybrid precoding design can not perform better that the shown optimal hybrid precoding solution with the considered RF codebook, which confirms the near-optimal result of the proposed algorithm. This is also clear in the considerable gain obtained by the proposed algorithm compared with the prior solution in [20]. Also, it is worth mentioning that the developed hybrid precoding algorithms in this work can be applied to any large MIMO system (not specifically mmWave systems). The same setup is considered in Fig. 5.6, but with $N_{\mathrm{BS}} = 64$ antennas. Fig. 5.6 illustrates the gain achieved by Algorithm 5 compared with the designs in [20] for different codebook sizes.

In Fig. 5.7, we evaluate the performance of the proposed approximate Gram-

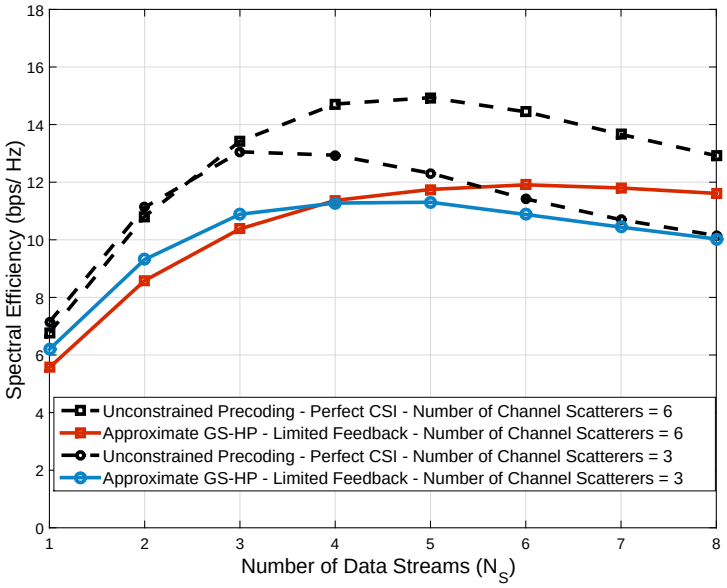


Figure 5.7: The performance of the approximate Gram-Schmidt hybrid precoding design in Algorithm 5 compared with the fully-digital SVD solution for different numbers of data streams. The system has $N_{\mathrm{BS}} = 32$ antennas, $N_{\mathrm{MS}} = 8$ antennas, and $N_{\mathrm{RF}} = N_{\mathrm{S}}$.

Schmidt hybrid precoding algorithm compared with the digital unconstrained solution for different numbers of transmitted data streams. In this figure, we adopt the same setup of Fig. 5.4(a), but with $N_{\mathrm{BS}} = 32$ antennas and $N_{\mathrm{MS}} = 8$ antennas. Further, each RF beamforming vector is selected from a beamsteering codebook with a size $N_{\mathrm{CB}}^{\mathrm{v}} = 128$. The number of RF chains are assumed to be equal to the number of data streams. First, Fig. 5.7 shows that the performance of the both the unconstrained precoding and the hybrid precoding increases then decreases again with the number of data streams. This decrease with large numbers of transmitted data streams is a result of the sparse mmWave channels and the equal power allocation among

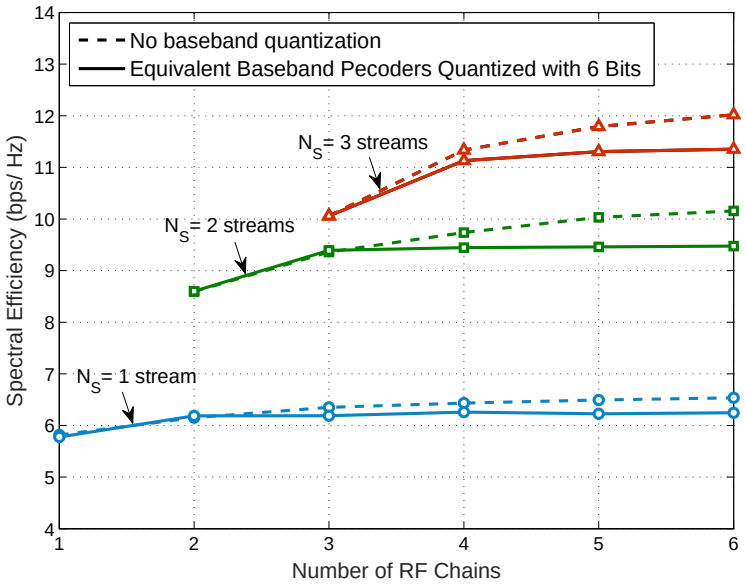


Figure 5.8: The performance of the approximate Gram-Schmidt hybrid precoding design in Algorithm 5 compared with the fully-digital SVD solution for different numbers of data streams. The system has $N_{\mathrm{BS}} = 32$ antennas, $N_{\mathrm{MS}} = 8$ antennas, and $N_{\mathrm{RF}} = N_{\mathrm{S}}$.

the different streams, which causes some power to be allocated to less important multi-path components. The solution to this problem is what is called multi-mode precoding [141, 142]. Further, this figure illustrates that the difference between the proposed hybrid precoding algorithm and digital SVD solution is small at both the small number of streams and the large number of streams regimes, which also follows from the sparsity of mmWave channels.

5.8.3 Gain of RF Chains

In Table 5.1, we summarize the required feedback overhead for the limited feedback operation of OFDM-based hybrid precoding systems. Table 5.1 shows that

when the number of transmitted streams equals the number of RF chains, $N_{\mathrm{S}} = N_{\mathrm{RF}}$, then only the feedback bits that correspond to the RF precoding codeword need to be fed back to the transmitter. Note that the reason is not that we only need RF beamforming for this case, but because the optimal baseband precoder, as obtained by Proposition 19, can be written as a matrix that depends only on the RF precoder multiplied by a unitary matrix. Table 5.1 also shows that the number of feedback bits scales linearly with the number of subcarriers if $N_{\mathrm{S}} < N_{\mathrm{RF}}$. It is, therefore, interesting to evaluate the gain of employing more RF chains than the number of streams, as achieving this gain requires considerable feedback overhead. In Fig. 5.8, we plot the spectral efficiency achieved with the proposed Gram-Schmidt greedy hybrid precoding versus the number of RF chains for $N_{\mathrm{S}} = 1, 2$, and 3 streams. The RF beamforming vectors are quantized with 5 bits and the equivalent baseband precoders are quantized with 6 bits. This figure shows that the spectral efficiency gain of having more RF chains saturates after a few RF chains. Fig. 5.8 also illustrates that having a number of RF chains $N_{\mathrm{RF}} = 2N_{\mathrm{S}}$ achieves less than 20% gain, with the cost of much more feedback overhead. The required numbers of feedback bits and the achievable spectral efficiency are also listed in Table 5.2 for the case $N_{\mathrm{S}} = 2$ streams. Table 5.2 shows that 3102 bits are needed to achieve 9.7 bps/Hz spectral efficiency when 6 RF chains are employed, while only 10 bits are enough to obtain 8.6 bps/Hz when $N_{\mathrm{RF}} = N_{\mathrm{S}} = 2$. These results indicate that activating a number of RF chains equals to the number of data streams may be a good technique to reduce the feedback overhead and increase the feasibility of limited feedback operation in hybrid precoding based wideband mmWave systems.

Table 5.2: Required feedback overhead for the hybrid precoding transmission in Fig. 5.8 with $N_{\mathrm{S}} = 2$ streams

Number of RF chains N_{RF}	2	4	6
RF feedback bits $N_{\mathrm{RF}} \log_2 \lvert\mathcal{F}_{\mathrm{RF}}^{\mathrm{v}}\rvert$	10 bits	20 bits	30 bits
Baseband feedback bits per subcarrier	0 bits	6 bits	6 bits
Total feedback bits for 512 subcarriers	10 bits	3092 bits	3102 bits
Spectral efficiency with $N_{\mathrm{S}} = 2$ streams	8.6 bps/Hz	9.5 bps/Hz	9.7 bps/Hz

5.9 Conclusion

In this chapter, we investigated limited feedback hybrid precoding design for wideband mmWave systems. First, we derived the optimal hybrid precoding design that maximizes the achievable mutual information for any given RF codebook, and showed that the optimal baseband structure can be decomposed into an RF precoder dependent matrix and a unitary matrix. This indicated that when the number of data streams equals to the number of RF chains, only the feedback of the RF precoder index is sufficient to achieve the maximum mutual information. Exploiting the structure of the optimal hybrid precoders, we also showed that the codebook of the equivalent baseband precoders should have a unitary structure. These notes led to efficient hybrid analog/digital precoders codebooks for spatial multiplexing in wideband mmWave systems. Further, we developed a novel greedy hybrid precoding algorithm based on Gram-Schmidt orthogonalization. Thanks to this Gram-Schmidt orthogonalization, we showed that only sequential design of the RF and baseband precoders is required to achieve the same performance of more sophisticated algorithms that requires a joint design of the RF and baseband precoders in each step.

Simulation results illustrated that the proposed codebook and precoding algorithms improve over prior work and stay within a small gap from the unconstrained perfect channel knowledge solutions.

Chapter 6

Concluding Remarks

6.1 Summary

This dissertation focused on developing precoding and channel estimation solutions that address key challenges in massive MIMO systems, namely, the hardware constraints, the channel acquisition overhead, and the precoding design complexity. First, we proposed efficient mmWave channel estimation algorithms leveraging the sparse nature of the channels. Then, we developed hybrid analog/digital precoding techniques for multi-user mmWave MIMO systems. This solution is then generalized to consider out-of-cell interference resulting, what we called, multi-layer precoding. Finally, we considered frequency-selective channels and developed hybrid precoding and codebook designs that leverage the large-scale MIMO channel characteristics to realize low-complexity solutions.

For the mmWave channel estimation problem, we developed a sparse formulation leveraging the sparse nature of the channels. Based on this formulation, we proposed a channel estimation algorithms that efficiently detect the different parameters of the mmWave channel with a low training overhead. Employing hybrid analog/digital architectures, we designed a hierarchical precoding codebook that constructs training beams with arbitrary beamwidth. The performance of the proposed

algorithm is analytically evaluated for the single-path channel case, and some insights into efficient training power distributions are obtained. Despite its low-complexity, results showed that the proposed mmWave channel estimation solution can achieve good estimation quality, which yields close spectral efficiency to that obtained by exhaustive search techniques.

For downlink multi-user mmWave MIMO systems, we proposed a low-complexity yet near-optimal hybrid analog/digital precoding algorithm. The proposed solution leverages the characteristics of the mmWave channels and the massive MIMO systems. The performance of the proposed algorithm was analyzed when the channels are single-path and when the system dimensions are very large. In these cases, the asymptotic optimality of the proposed algorithm was established, and the gains over beamsteering solutions were illustrated. Results showed indicated that interference management in multi-user mmWave systems is required even when the number of antennas is very large. With limited feedback channels, results showed that the hybrid precoding gains are sensitive to RF angles quantization. Further, good quantization of the effective lower-dimensional baseband precoders is required to maintain good gains over analog-only solutions.

For massive MIMO cellular systems, we proposed a general precoding framework, called multi-layer precoding. We developed a specific multi-layer precoding design for full-dimensional and massive MIMO systems that efficiently manages different kinds of interference. Further, the developed algorithm leverages the large-scale channel statistics and the directional structure of the massive MIMO channels to reduce the required training overhead. The performance of multi-layer precoding is

analyzed, and its asymptotic optimality is proved for one-ring and single-path channel models. Results indicated that multi-layer precoding can achieve a close performance, in terms of rate and coverage, to the single-user case. Further, results showed that multi-layer precoding achieves clear gains over conventional massive MIMO precoding techniques.

For frequency-selective channels, we designed and OFDM-based hybrid precoding and codebook designs. Different than fully-digital MIMO precoding, the RF precoding is common between all subcarriers in the hybrid architectures. We derived the optimal hybrid precoding design that maximizes the achievable mutual information for any given RF codebook, and showed that the optimal baseband structure can be decomposed into an RF precoder dependent matrix and a unitary matrix. This indicated that when the number of data streams equals to the number of RF chains, only the feedback of the RF precoder index is sufficient to achieve the maximum mutual information. Exploiting the optimal precoding structure, we developed efficient codebooks for the hybrid analog and digital precoders. Results showed that the proposed codebooks and precoding designs achieve close spectral efficiencies to that obtained by fully-digital solutions, despite adopting cost and power efficient architectures.

6.2 Future Work

There are several possible directions for future research

Radom Measurements for mmWave Channel Estimation: In Chapter 2, we proposed a low-complexity mmWave channel estimation algorithm that

leverages the sparse nature of the channel and adaptive compressed sensing tools. Extending this solution to multi-user system, however, is non-trivial. This is mainly due to the adaptive nature of the solution that causes the training overhead to scale with the number of users. One way to prevent this scaling is by adopting random beamforming/measurement vectors. Random beamforming/measurement allows all users to simultaneously estimate the channels, and hence decreases the associated training overhead. Therefore, it would be interesting to develop random compressed sensing based channel estimation algorithms for mmWave systems leveraging the sparse formulation developed in Chapter 2.

Several challenges, though, face channel estimation with random compressive sensing at hybrid architecture based mmWave systems. First, the hardware constraints makes it challenging to derive closed-form recovery guarantees. This, therefore, requires developing new analytical tools to overcome the challenges and to enable formulating solid statements on the mmWave channel estimation performance under random compressive sensing. Further, given that the final communication objective is the data rate, it is interesting to optimize the compressive measurement tools and algorithms for hybrid architecture based mmWave systems while adopting the system sum-rate as a performance metric. One difficulty of doing that come from the mmWave geometric channel model that leaves a small gap for tractable analysis. This opens new rooms for future work.

Multi-Layer Precoding Operation with Hybrid Architectures: The multiplicative structure of the proposed multi-layer precoding framework in Chapter 4 makes it feasible for hybrid architecture implementations. While we highlighted

one main idea for this implementation, several open problems still need careful investigations. For example, it is important to develop novel and optimized solutions for implementing the inter-cell interference avoidance stage in the RF beamforming. One initial way to do that is by employing antenna elements that have electrically-controlled directional patterns. This, however, needs further research in the optimization and the actual implementation.

Another important challenge with hybrid architecture based multi-layer precoding is the channel acquisition. In Chapter 4, we outlined a three-stage channel acquisition procedure to obtain the required channel knowledge for designing the proposed multi-layer precoding algorithm. Implementing this procedure with hybrid architecture, though, is not straight forward and needs some research. For example, estimating the elevation interference covariance matrix when the channel is seen through the RF lens is an interesting problem. Therefore, extending the hybrid architecture based channel estimation solution in Chapter 2 to the multi-layer precoding is an interesting problem for future work.

Analyzing hybrid architecture with accurate RF circuit models: Prior research on hybrid architectures based precoding and channel estimation algorithms assumed ideal RF circuit components and antenna. Practical circuits and antennas, however, have non-ideal characteristics that can affect these designs and resulting conclusions. It is therefore important to study the impact of hardware impairments on the performance of low-frequency and mmWave massive MIMO systems. For example, it is known that the fully-connected hybrid architecture has an array gain over the array-of-subarray one, but it is not clear what is the net gain if the insertion

losses in the power dividers, combiners, and phase shifters are taken into consideration. Analyzing different hybrid architectures under practical RF circuit models is critical to provide an accurate evaluation of the different architectures.

It would also be interesting to develop impairment-aware precoding and channel estimation solutions. Based on the insights that will be obtained from analyzing hybrid architectures under practical circuit models, it might be possible to design new optimized hybrid architectures, precoding, and channel estimation solutions that take these insights into consideration. Further, developing solutions that are more robust to the hardware impairments can boost the actual performance of hybrid architectures under practical conditions. Therefore, developing such hybrid architectures and associated signal processing solutions and analyzing their performance is important for future mmWave and massive MIMO systems.

Appendices

Appendix A

Proof of Lemma 13

To prove the achievable rate in (4.18), it is sufficient to prove that the power normalization factor $[\boldsymbol{\Upsilon}]_{k,k}$ that satisfies the multi-layer precoding power constraint $\left\| \left[\mathbf{F}_c^{(1)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} \right]_{:,k} \right\|^2 = 1$ is given by $[\boldsymbol{\Upsilon}]_{k,k} = \sqrt{\left(\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1}_{k,k} \right)^{-1}}$. Using this values of $[\boldsymbol{\Upsilon}]_{k,k}$, the multi-layer precoding power constraint can be written as

$$\begin{aligned}
\left\| \left[\mathbf{F}_c^{(1)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} \right]_{:,k} \right\|^2 &= [\boldsymbol{\Upsilon}]^*_{:,k} \mathbf{F}_c^{(3)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(1)} \mathbf{F}_c^{(1)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} [\boldsymbol{\Upsilon}]_{:,k} && \text{(A.1)} \\
&\overset{(a)}{=} [\boldsymbol{\Upsilon}]^*_{:,k} \mathbf{F}_c^{(3)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} [\boldsymbol{\Upsilon}]_{:,k} && \text{(A.2)} \\
&= [\boldsymbol{\Upsilon}]^*_{:,k} \left(\overline{\mathbf{H}}_c^* \overline{\mathbf{H}}_c \right)^{-1} \overline{\mathbf{H}}_c^* \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)} \overline{\mathbf{H}}_c \left(\overline{\mathbf{H}}_c^* \overline{\mathbf{H}}_c \right)^{-1} [\boldsymbol{\Upsilon}]_{:,k} && \text{(A.3)} \\
&\overset{(b)}{=} [\boldsymbol{\Upsilon}]^*_{:,k} \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} \mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)} \\
&\quad \times \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} [\boldsymbol{\Upsilon}]_{:,k} && \text{(A.4)} \\
&= [\boldsymbol{\Upsilon}]^*_{:,k} \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} \mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \\
&\quad \times \left(\mathbf{W}_c \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c^* \right)^{-1} \mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \\
&\quad \times \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} [\boldsymbol{\Upsilon}]_{:,k} && \text{(A.5)} \\
&= [\boldsymbol{\Upsilon}]^*_{:,k} \left(\mathbf{W}_c \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c^* \right)^{-1} [\boldsymbol{\Upsilon}]_{:,k} && \text{(A.6)} \\
&\overset{(c)}{=} 1, && \text{(A.7)}
\end{aligned}$$

where (a) follows by noting that $\mathbf{F}_c^{(1)}$ has a semi-unitary structure. The effective channel matrix $\overline{\mathbf{H}}_c = [\overline{h}_{c1}, ..., \overline{h}_{cK}]$ with $\overline{h}_{ck} = \left[\mathbf{G}_{c,(k,1)}, ..., \mathbf{G}_{c,(k,1)}\right]^* \overline{\mathbf{w}}_{cck}, k = 1, ..., K$ can also be written as $\overline{\mathbf{H}}_c = \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c$ with $\mathbf{W}_c = \mathbf{I}_K \circ [\overline{\mathbf{w}}_{cc1}, ..., \overline{\mathbf{w}}_{ccK}]$, which leads to (b). Finally, (c) follows by substituting with the value of $[\boldsymbol{\Upsilon}]_{k,k}$.

Appendix B

Proof of Theorem 14

Considering the system and channel models in Section 4.3 and applying the multi-layer precoding algorithm in Section 4.5, the achievable rate by user k at cell c is given by Lemma 13

$$R_{ck} = \log_2 \left(1 + \frac{\mathsf{SNR}}{\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1}_{k,k}} \right). \tag{B.1}$$

If $\mathbf{G}_{c,(k,m)} = \mathbf{0}, \forall m \neq k$ and $\mathbf{G}_{c,(k,k)} = \mathbf{I}$, then by noting that the matrix $\mathbf{W}_c$ has a block diagonal structure and using the matrix inversion lemma [143], we get $\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)^*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1}_{k,k} = \|\overline{\mathbf{w}}_{cck}\|^{-2}$. Therefore, to complete the proof, it is sufficient to prove that (i) $\lim_{N_{\mathrm{V}}, N_{\mathrm{H}} \to \infty} \mathbf{G}_{c,(k,m)} = \mathbf{0}, \forall m \neq k$ and (ii) $\lim_{N_{\mathrm{V}}, N_{\mathrm{H}} \to \infty} \mathbf{G}_{c,(k,k)} = \mathbf{I}$. To do that, we will first present the following useful lemma, which is a modified version of Lemma 3 in [86].

Lemma 21 *Consider a user k at cell c with an azimuth angle ϕ_{ck}. Adopt the one-ring channel model in* (4.19) *with an azimuth angular spread Δ_{A} and correlation matrix $\mathbf{R}^{\mathrm{A}}_{cck}$. Define the unit-norm azimuth array response vector associated with an azimuth angle ϕ_m and elevation angle θ_m as $\mathbf{u}_m = \frac{\mathbf{a}(\phi_m, \theta_m)}{\sqrt{(N_{\mathrm{H}})}}$, where $\mathbf{a}(\phi_m, \theta_m) = \left[1, ..., e^{jkD(N_{\mathrm{H}}-1)\sin(\theta_x)\sin(\phi_x)}\right]$. If the angle $\phi_x \notin [\phi_{ck} - \Delta_{\mathrm{A}}, \phi_{ck} + \Delta_{\mathrm{A}}]$, then*

$$\mathbf{u}_x \in Null\left(\mathbf{R}^{\mathrm{A}}_{cck}\right), \quad as \quad N_{\mathrm{H}} \to \infty. \tag{B.2}$$

Proof: First, note that $\left[\mathbf{R}^{\mathrm{A}}_{cck}\right]_{n_1,n_2}$ in (4.19), can also be written as

$$\left[\mathbf{R}^{\mathrm{A}}_{cck}\right]_{n_1,n_2} = \frac{1}{2\Delta_{\mathrm{A}}}\int_{-\Delta_{\mathrm{A}}}^{\Delta_{\mathrm{A}}} \left[\mathbf{a}\left(\phi_{ck}+\alpha,\theta_{ck}\right)\mathbf{a}^{*}\left(\phi_{ck}+\alpha,\theta_{ck}\right)\right]_{n_1,n_2} d\alpha \tag{B.3}$$

Then, we have

$$\mathbf{u}_m^{*}\mathbf{R}\mathbf{u}_m = \frac{1}{2\Delta_{\mathrm{A}}N_{\mathrm{H}}}\int_{-\Delta_{\mathrm{A}}}^{\Delta_{\mathrm{A}}} \mathbf{a}^{*}(\phi_m,\theta_m)\mathbf{a}\left(\phi_{ck}+\alpha,\theta_{ck}\right)\mathbf{a}^{*}\left(\phi_{ck}+\alpha,\theta_{ck}\right)\mathbf{a}(\phi_m,\theta_m)d\alpha \tag{B.4}$$

$$= \frac{1}{2\Delta_{\mathrm{A}}}\int_{-\Delta_{\mathrm{A}}}^{\Delta_{\mathrm{A}}} \frac{1}{N_{\mathrm{H}}}\left|\mathbf{a}^{*}(\phi_m,\theta_m)\mathbf{a}\left(\phi_{ck}+\alpha,\theta_{ck}\right)\right|^2 d\alpha. \tag{B.5}$$

Using Lemma 1 in [144], we reach

$$\lim_{N_H\to\infty} \mathbf{u}_m^{*}\mathbf{R}\mathbf{u}_m = 0, \quad \forall \phi_m \notin \left[\phi_{ck}-\Delta_{\mathrm{A}},\phi_{ck}+\Delta_{\mathrm{A}}\right]. \tag{B.6}$$

□

Now, to prove that $\mathbf{G}_{c,(k,m)} = \overline{\mathbf{U}}^{*}_{cck}\overline{\mathbf{U}}_{ccr} = \mathbf{U}^{\mathrm{A}^*}_{cck}\mathbf{U}^{\mathrm{A}}_{ccr} \otimes \overline{\mathbf{U}}^{\mathrm{E}^*}_{cck}\overline{\mathbf{U}}^{\mathrm{E}}_{ccr} = \mathbf{0}$, we need to prove that either $\mathbf{U}^{\mathrm{A}^*}_{cck}\mathbf{U}^{\mathrm{A}}_{ccr} = \mathbf{0}$ or $\overline{\mathbf{U}}^{\mathrm{E}^*}_{cck}\overline{\mathbf{U}}^{\mathrm{E}}_{ccr} = \mathbf{0}$. If $|\phi_{ck}-\phi_{cm}| \geq 2\Delta_{\mathrm{A}}$, then the columns of $\mathbf{U}^{\mathrm{A}}_{ccm} \in \mathrm{Span}\left\{\frac{\mathbf{a}(\phi_m)}{\sqrt{(N_{\mathrm{H}})}}\,|\phi_m \in [\phi_{cm}-\Delta_{\mathrm{A}},\phi_{cm}+\Delta_{\mathrm{A}}]\right\} \subseteq \mathrm{Null}\left(\mathbf{R}^{\mathrm{A}}_{cck}\right)$ as $N_{\mathrm{H}}\to\infty$, which follows from Lemma 21. This leads to $\lim_{N_{\mathrm{H}}\to\infty}\mathbf{U}^{\mathrm{A}}_{cck}{}^{*}\mathbf{U}^{\mathrm{A}}_{ccm} = \mathbf{0}$. Similarly, if $|\theta_{ck}-\theta_{cm}| \geq 2\Delta_{\mathrm{E}}$, then $\lim_{N_{\mathrm{V}}\to\infty}\mathbf{U}^{\mathrm{E}^*}_{cck}\mathbf{U}^{\mathrm{E}}_{ccm} = \mathbf{0}$. Further, since $d \leq d_{\max}$, we have $|\theta_{ck}-\theta_I| \geq 2\Delta_{\mathrm{E}}$, for any elevation angle θ_I of another cell user. This implies that $\mathbf{U}^{\mathrm{E}}_{cck} \in \mathrm{Range}\left\{\mathbf{U}_{\mathrm{NI}}\right\}$ as $N_{\mathrm{V}}\to\infty$ by Lemma 21, and $\exists\mathbf{A}_{ck}$ such that $\mathbf{U}^{\mathrm{E}}_{cck} = \mathbf{U}_{\mathrm{NI}}\mathbf{A}_{ck}$. For the $\mathbf{U}_{ccm}$, it can be generally expressed as $\mathbf{U}_{ccm} = \mathbf{U}_{\mathrm{NI}}\mathbf{A}_{cm} + \mathbf{U}_{\mathrm{I}}\mathbf{B}_{cm}$ for some matrices $\mathbf{A}_{cm},\mathbf{B}_{cm}$ of proper dimensions. As $\lim_{N_{\mathrm{V}}\to\infty}\mathbf{U}^{\mathrm{E}^*}_{cck}\mathbf{U}^{\mathrm{E}}_{ccm} = \mathbf{0}$, we have $\lim_{N_{\mathrm{V}}\to\infty}\mathbf{A}^{*}_{ck}\mathbf{A}_{cm} = \mathbf{0}$. Then, $\overline{\mathbf{U}}^{\mathrm{E}^*}_{cck}\overline{\mathbf{U}}^{\mathrm{E}}_{ccm} = \mathbf{A}^{*}_{ck}\mathbf{A}_{cm} = \mathbf{0}$ as $N_{\mathrm{V}}\to\infty$. This completes the proof of the first condition, $\mathbf{G}_{c,(k,m)} = \mathbf{0}$ if $|\phi_{ck}-\phi_{cm}| \geq 2\Delta_{\mathrm{A}}$ or $|\theta_{ck}-\theta_{cm}| \geq 2\Delta_{\mathrm{E}}$, $\forall m \neq k$.

To prove that $\lim_{N_{\mathrm{V}},N_{\mathrm{H}}\to\infty}\mathbf{G}_{c,(k,k)}=\mathbf{I}$, we need to show that $\overline{\mathbf{U}}^{\mathrm{E}^*}_{cck}\overline{\mathbf{U}}^{\mathrm{E}}_{cck}=\mathbf{I}$. Since $\mathbf{U}^{\mathrm{E}}_{cck}$ can be written as $\mathbf{U}^{\mathrm{E}}_{cck}=\mathbf{U}_{\mathrm{NI}}\mathbf{A}_{ck}$ when $N_{\mathrm{V}}\to\infty$, then we have $\mathbf{U}^{\mathrm{E}^*}_{cck}\mathbf{U}^{\mathrm{E}}_{cck}=\mathbf{A}^*_{ck}\mathbf{A}_{ck}=\mathbf{I}$. This results in $\overline{\mathbf{U}}^{\mathrm{E}^*}_{cck}\overline{\mathbf{U}}^{\mathrm{E}}_{cck}=\mathbf{A}^*_{ck}\mathbf{A}_{ck}=\mathbf{I}$ as $N_{\mathrm{V}}\to\infty$, which completes the proof.

Bibliography

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

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

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

[4] 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 Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, April 2016.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[6] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June

2011.

[7] T. Bai, A. Alkhateeb, and R. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 70–77, Sept. 2014.

[8] S. Rangan, T. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[9] M. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Sel. Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[10] J. G. Andrews, T. Bai, M. Kulkarni, A. Alkhateeb, A. Gupta, and R. W. Heath Jr, "Modeling and analyzing millimeter wave cellular systems," *submitted to IEEE Transactions on Communications, arXiv preprint arXiv:1605.04283*, 2016.

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

[12] J. Singh, S. Ponnuru, and U. Madhow, "Multi-gigabit communication: the ADC bottleneck," in *Proc. of IEEE International Conference on Ultra-Wideband*

(ICUWB), Vancouver, BC, Sept. 2009, pp. 22–27.

[13] J. Wang, Z. Lan, C. Pyo, T. Baykas, C. Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, Nov. 2009.

[14] Y. Tsang, A. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmwave communication system," in *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM)*, Houston, TX, Dec. 2011, pp. 1–6.

[15] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *Proc. of 6th International ICST Conference on Communications and Networking in China*, China, 2011, pp. 361–365.

[16] A. Abbaspour-Tamijani and K. Sarabandi, "An affordable millimeter-wave beam-steerable antenna using interleaved planar subarrays," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 9, pp. 2193–2202, Sept. 2003.

[17] B. Biglarbegian, M. Fakharzadeh, D. Busuioc, M. Nezhad-Ahmadi, and S. Safavi-Naeini, "Optimized microstrip antenna arrays for emerging millimeter-wave wireless applications," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 5, pp. 1742–1747, May 2011.

[18] X. Zhang, A. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband

codesign for MIMO antenna selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[19] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[20] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[21] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[22] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *Journal of Communications and Networks*, vol. 15, no. 4, pp. 338–351, Aug 2013.

[23] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[24] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks,"

IEEE Transactions on Communications, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[25] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5213–5226, Sept 2013.

[26] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, December 2010.

[27] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[28] ——, "Single-sided adaptive estimation of multi-path millimeter wave channels," in *Signal Processing Advances in Wireless Communications (SPAWC), 2014 IEEE 15th International Workshop on*, June 2014, pp. 125–129.

[29] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine,*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

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

[31] A. Alkhateeb, R. W. Heath, and G. Leus, "Achievable rates of multi-user millimeter wave systems with hybrid precoding," in *IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 1232–1237.

[32] A. Alkhateeb, G. Leus, and R. W. Heath, "Multi-layer precoding for full-dimensional massive MIMO systems," in *Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 815–819.

[33] ——, "Multi-layer precoding: A potential solution for full-dimensional massive MIMO systems," *under preparation*.

[34] A. Alkhateeb and R. W. Heath Jr, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *submitted to IEEE Transactions on Communications, arXiv preprint arXiv:1510.00609*, 2015.

[35] IEEE Standard 802.15.3c, "Wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs), amendement 2: Millimeter-wave-based alternative physical layer extension," Oct. 2009.

[36] IEEE 802.11ad, "IEEE 802.11ad standard draft D0.1." [Online]. Available: www.ieee802.org/11/Reports/tgad update.htm

[37] O. El Ayach, R. Heath, S. Rajagopal, and Z. Pi, "Multimode precoding in millimeter wave mimo transmitters with multiple antenna sub-arrays," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, Dec 2013, pp. 3476–3480.

[38] S. Hur, S. Baek, B. Kim, Y. Chang, A. F. Molisch, T. S. Rappaport, K. Haneda, and J. Park, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 454–469, Apr. 2016.

[39] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. of Information Theory and Applications Workshop (ITA)*, Feb 2013, pp. 1–5.

[40] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[41] A. J. Laub, *Matrix analysis for scientists and engineers*. Society for Industrial and Applied Mathematics, 2004.

[42] H. L. Van Trees, "Optimum array processing (detection, estimation, and modulation theory, part iv)," *Wiley-Interscience, Mar*, no. 50, p. 100, 2002.

[43] H. Zhu, G. Leus, and G. Giannakis, "Sparse regularized total least squares for sensing applications," in *Proc. of IEEE Eleventh International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Marrakech, Morocco, July 2010, pp. 1–5.

[44] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Transactions on Signal*

Processing, vol. 59, no. 10, pp. 4735–4744, 2011.

[45] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive parameter estimation in awgn," *IEEE Transactions on Signal Processing*, vol. 62, no. 8, pp. 2012–2027, April 2014.

[46] M. Rossi, A. Haimovich, and Y. Eldar, "Spatial compressive sensing for MIMO radar," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 419–430, 2014.

[47] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[48] M. Malloy and R. Nowak, "Near-optimal adaptive compressed sensing," *Information Theory, IEEE Transactions on*, vol. 60, no. 7, pp. 4001–4012, July 2014.

[49] M. Iwen and A. Tewfik, "Adaptive strategies for target detection and localization in noisy environments," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2344–2353, 2012.

[50] T. Bai, R. Vaze, and R. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 5070–5083, Sept. 2014.

[51] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[52] J. Andrews, F. Baccelli, and R. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, November 2011.

[53] S. Akoum, O. El Ayach, and R. W. Heath, "Coverage and capacity in mmwave cellular systems," in *Proc. of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, CA, 2012, pp. 688–692.

[54] T. Bai and R. W. Heath Jr., "Coverage analysis for millimeter wave cellular networks with blockage effects," in *Proc. of IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, Dec. 2013.

[55] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[56] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[57] M. Iwen, "Group testing strategies for recovery of sparse signals in noise," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, Nov 2009, pp. 1561–1565.

[58] A. Soni and J. Haupt, "On the fundamental limits of recovering tree sparse vectors from noisy linear measurements," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 133–149, Jan. 2014.

[59] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.

[60] J. Murdock, E. Ben-Dor, Y. Qiao, J. Tamir, and T. Rappaport, "A 38 GHz cellular outage study for an urban outdoor campus environment," in *Proc. of Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, Apr. 2012, pp. 3085–3090.

[61] A. Sayeed and J. Brady, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," in *Proc. of 2013 IEEE Global Telecommunications Conference (GLOBECOM)*, Atlanta, GA, 2013.

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

[63] P. Xia, S.-K. Yong, J. Oh, and C. Ngo, "A practical SDMA protocol for 60 GHz millimeter wave communications," in *Proc. of 42nd Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Oct. 2008, pp. 2019–2023.

[64] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[65] C.-B. Chae, D. Mazzarese, T. Inoue, and R. Heath, "Coordinated beamforming for the multiuser MIMO broadcast channel with limited feedforward," *IEEE*

Transactions on Signal Processing, vol. 56, no. 12, pp. 6044–6056, Dec. 2008.

[66] C.-B. Chae, D. Mazzarese, N. Jindal, and R. Heath, "Coordinated beamforming with limited feedback in the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1505–1515, Oct. 2008.

[67] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[68] F. Boccardi and H. Huang, "A near-optimum technique using linear precoding for the MIMO broadcast channel," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, Apr. 2007, pp. III–17–III–20.

[69] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[70] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[71] M. Trivellato, H. Huang, and F. Boccardi, "Antenna combining and codebook design for the MIMO broadcast channel with limited feedback," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, Nov. 2007, pp. 302–308.

[72] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proc. of 2011 IEEE GLOBECOM Workshops (GC Wkshps)*, Houston, TX, 2011, pp. 253–257.

[73] H. Yang and T. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[74] P. Smith, C. Neil, M. Shafi, and P. Dmochowski, "On the convergence of massive MIMO systems," in *Proc. of IEEE International Conference on Communications (ICC)*, June 2014, pp. 5191–5196.

[75] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," vol. 24, no. 3, pp. 528–541, Mar. 2006.

[76] L. V. Kantorovich, "Functional analysis and applied mathematics," *Uspekhi Matematicheskikh Nauk*, vol. 3, no. 6, pp. 89–185, 1948.

[77] Z. Bai and G. H. Golub, "Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices," *Annals of Numerical Mathematics*, vol. 4, pp. 29–38, 1996.

[78] Z. Hong, K. Liu, R. Heath, and A. Sayeed, "Spatial multiplexing in correlated fading via the virtual channel representation," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 856–866, June 2003.

[79] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: Basic theory and applications," in *Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept. 2010, pp. 1196–1203.

[80] V. Va, J. Choi, and R. W. Heath Jr, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *arXiv preprint arXiv:1511.02937*, 2015.

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

[82] E. Bjrnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[83] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[84] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, August 2011.

[85] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, July 2012, pp. 1137–1141.

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

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

[88] T. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, Dec. 2014, pp. 4066–4071.

[89] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, Dec 2014.

[90] A. Kammoun, H. Khanfir, Z. Altman, M. Debbah, and M. Kamoun, "Preliminary results on 3D channel modeling: From theory to standardization," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1219–1229, June 2014.

[91] Z. Zhong, X. Yin, X. Li, and X. Li, "Extension of ITU IMT-advanced channel models for elevation domains and line-of-sight scenarios," in *Proc. of IEEE*

Vehicular Technology Conference, Sept 2013, pp. 1–5.

[92] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. of IEEE International Conference on Communications (ICC)*, June 2014, pp. 5865–5870.

[93] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[94] G. T. 36.873, "Study on 3D channel model for lte," Sep. 2014.

[95] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, Mar 2000.

[96] P. Petrus, J. H. Reed, and T. S. Rappaport, "Geometrical-based statistical macrocell channel model for mobile environments," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 495–502, Mar 2002.

[97] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 550–560, Apr 2002.

[98] M. Zhang, P. J. Smith, and M. Shafi, "An extended one-ring MIMO channel model," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2759–2764, August 2007.

[99] E. Bjrnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, Nov 2014.

[100] J. C. Shen, J. Zhang, E. Alsusa, and K. B. Letaief, "Compressed CSI acquisition in FDD massive MIMO: How much training is needed?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4145–4156, June 2016.

[101] H. Lutkepohl, "Handbook of matrices." *Computational Statistics and Data Analysis*, vol. 2, no. 25, p. 243, 1997.

[102] N. Seifi, J. Zhang, R. Heath, T. Svensson, and M. Coldrey, "Coordinated 3D beamforming for interference management in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5396–5410, Oct 2014.

[103] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Pearson Education, 2014.

[104] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[105] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 168–178, Jan. 2015.

[106] G. MacCartney and T. Rappaport, "73 GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York City," in *Proc. of IEEE International Conference on Communications (ICC)*, June 2014, pp. 4862–4867.

[107] T. Bai and R. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.

[108] S. Singh, M. Kulkarni, A. Ghosh, and J. Andrews, "Tractable model for rate in self-backhauled millimeter wave cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2196–2211, Oct. 2015.

[109] Y. Zhu, Z. Zhang, Z. Marzi, C. Nelson, U. Madhow, B. Y. Zhao, and H. Zheng, "Demystifying 60 GHz outdoor picocells," in *Proc. of the 20th annual international conference on Mobile computing and networking.* ACM, 2014, pp. 5–16.

[110] B. Walke and R. Briechle, "A local cellular radio network for digital voice and data transmission at 60 GHz," in *Proc. of the International Cellular & Mobile Comm.*, vol. 0. London: Online, Nov 1985, pp. 215–225.

[111] WirelessHD Standard, "WirelessHD specification version 1.0 overview," *WirelessHD*, Oct. 2007.

[112] A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "Communication challenges in high-density deployments of wearable wireless devices," *IEEE*

Wireless Communications, vol. 22, no. 1, pp. 12–18, Feb. 2015.

[113] K. Venugopal, M. C. Valenti, and R. W. Heath Jr, "Device-to-device millimeter wave communications: Interference, coverage, rate, and finite topologies," June 2015.

[114] J. Wehling, "Multifunction millimeter-wave systems for armored vehicle application," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 3, pp. 1021–1025, Mar. 2005.

[115] M. Heddebaut, F. Elbahhar, C. Loyez, N. Obeid, N. Rolland, A. Rivenq, and J.-M. Rouvaen, "Millimeter-wave communicating-radars for enhanced vehicle-to-vehicle communications," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 3, pp. 440–456, 2010.

[116] P. Kumari, N. Gonzlez Prelcic, and R. W. Heath Jr., "Investigating the IEEE 802.11ad standard for millimeter wave automotive radar," in *Proc. of the IEEE Vehicular Technology Conference (VTC 2015-Fall)*, Sept. 2015.

[117] T. A. Khan, A. Alkhateeb, and R. W. Heath Jr, "Millimeter wave energy harvesting," *submitted to IEEE Transactions on Wireless Communications, arXiv preprint arXiv:1509.01653*, Sept. 2015.

[118] I. Krikidis, S. Timotheou, S. Nikolaou, G. Zheng, D. Ng, and R. Schober, "Simultaneous wireless information and power transfer in modern communication systems," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 104–110, Nov 2014.

[119] S. Bi, C. Ho, and R. Zhang, "Wireless powered communication: opportunities and challenges," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 117–125, Apr. 2015.

[120] H. Sun, Y.-X. Guo, and Z. Wang, "60-GHz circularly polarized u-slot patch antenna array on LTCC," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 1, pp. 430–435, Jan. 2013.

[121] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale MIMO systems," in *Proc. of the IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia*, Apr. 2015.

[122] C. Mendez-Rial, R. Rusu, N. Gonzalez-Prelcic, and R. W. Heath, "Dictionary-free hybrid precoders and combiners for mmwave MIMO systems," in *Proc. of the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015.

[123] C.-E. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Communications Letters*, vol. 4, no. 3, pp. 285–288, June 2015.

[124] Y. w. Liang, R. Schober, and W. Gerstacker, "Transmit beamforming for frequency-selective channels with decision-feedback equalization," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4401–4411, December 2007.

[125] M. Samimi and T. Rappaport, "Ultra-wideband statistical channel model for non line of sight millimeter-wave urban channels," in *Proc. of the IEEE Global*

Communications Conference (GLOBECOM), Dec 2014, pp. 3483–3489.

[126] T. Rappaport, Y. Qiao, J. Tamir, J. Murdock, and E. Ben-Dor, "Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems," in *Radio and Wireless Symposium*, Santa Clara, CA, Jan. 2012, pp. 151–154.

[127] A. Forenza, D. Love, and R. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 1924–1934, July 2007.

[128] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter wave communications: The sparse way," in *the Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2014.

[129] A. Ghosh, T. Thomas, M. Cudak, R. Ratasuk, P. Moorut, F. Vook, T. Rappaport, G. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, June 2014.

[130] P.-A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of grassmann manifolds with a view on algorithmic computation," *Acta Applicandae Mathematica*, vol. 80, no. 2, pp. 199–220, 2004.

[131] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8,

pp. 2967–2976, Aug. 2005.

[132] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar 1982.

[133] R.-A. Pitaval and O. Tirkkonen, "Joint Grassmann-Stiefel quantization for MIMO product codebooks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 210–222, January 2014.

[134] B. Mondal, S. Dutta, and R. Heath, "Quantization on the grassmann manifold," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4208–4216, Aug 2007.

[135] R.-A. Pitaval, O. Tirkkonen, and S. Blostein, "Density and bounds for Grassmannian codes with chordal distance," in *Proc. of the IEEE International Symposium on Information Theory Proceedings (ISIT)*, July 2011, pp. 2298–2302.

[136] J. Lee, *Introduction to smooth manifolds*. Springer Science & Business Media, 2012, vol. 218.

[137] K. Ye and L.-H. Lim, "Distance between subspaces of different dimensions," *arXiv preprint arXiv:1407.0900*, 2014.

[138] J. Choi, B. Mondal, and R. Heath, "Interpolation based unitary precoding for spatial multiplexing MIMO-OFDM with limited feedback," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4730–4740, Dec 2006.

[139] J. Tropp, I. Dhillon, R. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 188–209, Jan 2005.

[140] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, vol. 415, no. 1, pp. 20 – 30, 2006.

[141] D. Love and R. Heath, "Multimode precoding for MIMO wireless systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3674–3687, Oct 2005.

[142] J. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 3:1–3:10, Mar. 2009.

[143] F. Zhang, *The Schur complement and its applications*. Springer Science & Business Media, 2006, vol. 4.

[144] O. El Ayach, R. W. Heath Jr, S. Abu-Surra, S. Rajagopal, and P. Zhouyue, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cesme, Turkey, June 2012, pp. 100–104.

Publications

Publications related to the dissertation

1. T. Bai, A. Alkhateeb, and R. W. Heath Jr, "Coverage and Capacity of Millimeter Wave Cellular Networks," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 70-77, Sept. 2014.

2. A. Alkhateeb, O. ElAyach, G. Leus, and R. W. Heath Jr, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831-846, Oct. 2014.

3. A. Alkhateeb, J. Mo, N. Gozalez-Prelcic, and R. W. Heath Jr, "MIMO Precoding and Combining Solutions for Millimeter Wave Systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122-131, Dec. 2014.

4. A. Alkhateeb, G. Leus, and R. W. Heath Jr, "Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481-6494, Nov. 2015.

5. A. Alkhateeb and R. W. Heath Jr, "Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems," submitted to *IEEE Transactions on Communications* (Invited Paper), vol. 64, no. 5, pp. 1801-1818, May 2016.

6. A. Alkhateeb, G. Leus, and R. W. Heath Jr, "Multi-Layer Precoding: A Potential Solution for Full-Dimensional Massive MIMO Systems," *under preparation*, May, 2016.

7. A. Alkhateeb, O. ElAyach, G. Leus, and R. W. Heath Jr, "Hybrid Precoding for Millimeter Wave Cellular Systems with Partial Channel Knowledge," in Proc. of *Information Theory and Applications Workshop (ITA)*, pp. 1-5, Feb. 2013.

8. A. Alkhateeb, O. ElAyach, G. Leus, and R. W. Heath Jr, "Single-Sided Adaptive Estimation of Multi-path Millimeter Wave Channels," in Proc. of the *International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 125-129, June 2014.

9. A. Alkhateeb, G. Leus, and R. W. Heath Jr, "Multi-Layer Precoding for Full-Dimensional Massive MIMO Systems," in Proc. of *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp. 815-819, Nov. 2014.

10. A. Alkhateeb, G. Leus, and R. W. Heath Jr, "Compressed-Sensing Based Multi-User Millimeter Wave Systems: How Many Measurements Are Needed?," in Proc. of the *IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, pp. 2909-2913, Apr. 2015.

11. A. Alkhateeb, R. W. Heath Jr., and G. Leus, "Achievable Rates of Multi-User Millimeter Wave Systems with Hybrid Precoding," in Proc. of the *IEEE International Conf. on Communications Workshops (ICC)*, London, UK, pp. 1232-1237, June 2015.

12. A. Alkhateeb, and R. W. Heath Jr, "Gram Schmidt Based Hybrid Precoding for Frequency Selective Millimeter Wave MIMO Systems," in Proc. of the *IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Mar., 2016.

Other publications

1. A. Alkhateeb, Y. Nam, J. Zhang, R. W. Heath Jr., "Massive MIMO Combining with Switches," to appear in *IEEE Wireless Communication Letters*, Jan., 2016.

2. Roi Mendez-Rial, Cristian Rusu, Nuria Gozalez-Prelcic, Ahmed Alkhateeb, and Robert W. Heath Jr, "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?," *IEEE Access*, vol. 4, pp. 247-267, 2016.

3. A. Alkhateeb, Y. Nam, M. Rahman, J. Zhang, R. W. Heath Jr., "Initial Beam Association for Millimeter Wave Cellular Networks: Analysis and Design Insights," submitted to *IEEE Transactions on Wireless Communications*, Feb. 2016.

4. T. A. Khan, A. Alkhateeb, and R. W. Heath Jr, "Millimeter Wave Energy Harvesting," to appear in *IEEE Transactions on Wireless Communications*, May 2016.

5. Abhishek Gupta, Ahmed Alkhateeb, Jeffrey Andrews, and Robert W. Heath Jr, "Gains of Restricted Secondary Licensing in Millimeter Wave Cellular Systems," submitted to *IEEE Journal on Selected Areas in Communications*, May, 2016

6. Jeffrey Andrews, Tianyang Bai, Mandar Kulkarni, Ahmed Alkhateeb, Abhishek Gupta, and Robert W. Heath Jr, "Modeling and Analyzing Millimeter Wave Cellular Systems," submitted to *IEEE Transactions on Communications* (Invited Paper), May, 2016

7. Jianhua Mo, Ahmed Alkhateeb, Shadi Abu-Surra, and Robert W. Heath Jr, "Hybrid Architectures with Few-Bit ADC Receivers: Achievable Rates and Energy-Rate Tradeoffs," submitted to *IEEE Transactions on Wireless Communications*, April, 2016

8. Sungwoo Park, Ahmed Alkhateeb, and Robert W. Heath Jr, "Dynamic Subarrays for Hybrid Precoding in Wideband mmWave MIMO Systems," submitted to *IEEE Transactions on Wireless Communications*, June, 2016

9. V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial Beamforming for mmWave Communications," in Proc. of *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2014.

10. R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. Gozalez-Prelcic, and R. W. Heath Jr, "Channel Estimation and Hybrid Combining for mmWave: Phase Shifters or Switches? ," in Proc. of *Information Theory and Applications Workshop (ITA)*, Feb. 2015.

11. M. Kulkarni, A. Alkhateeb, and J. Andrews, "A Tractable Model for Per User Rate in Multiuser Millimeter Wave Cellular Networks," to appear in *Asilomar Conference on Signals, Systems and Computers*, Nov. 2015.

12. M. E. Eltayeb, A. Alkhateeb, R. W. Heath Jr., and T. Y. Al-Naffouri, "Opportunistic Beam Training with Hybrid Analog/Digital Codebooks for mmWave Systems," to appear in the *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, Dec. 2015.

13. T. A. Khan, A. Alkhateeb, and R. W. Heath Jr, "Energy Coverage in Millimeter Wave Energy Harvesting Networks," to appear in *IEEE Globecom Workshops*, Dec. 2015.

14. Jianhua Mo, Ahmed Alkhateeb, Shadi Abu-Surra, and Robert W. Heath Jr., "Achievable Rates of Hybrid Architectures with Few-Bit ADC Receivers," in Proc. of *ITG workshop on smart antennas*, Mar., 2016

Vita

Ahmed Alkhateeb received his B.S. (with highest honors) and M.S. degrees from Cairo University, Egypt, in 2008 and 2012, respectively. He is currently working towards the Ph.D. degree at the University of Texas at Austin. During the summers of 2013 and 2014, he was a research intern at FutureWei Technologies (Huawei), where he developed schemes for initial beamforming in mmWave systems. During the summer of 2015, he was an intern at Samsung R&D Dallas. At Samsung, he performed research on how the inital beamforming can impact the network-level performance of mmWave systems. His current research interests are in the design and analysis of low-complexity precoding and channel estimation schemes for massive MIMO systems.

Permanent address: aalkhateeb@utexas.edu

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.