

**Multi-Layer Precoding
A Potential Solution for Full-Dimensional Massive MIMO Systems**

Alkhateeb, Ahmed; Leus, Geert; Heath, Robert W.

DOI

[10.1109/TWC.2017.2716362](https://doi.org/10.1109/TWC.2017.2716362)

Publication date

2017

Document Version

Final published version

Published in

IEEE Transactions on Wireless Communications

Citation (APA)

Alkhateeb, A., Leus, G., & Heath, R. W. (2017). Multi-Layer Precoding: A Potential Solution for Full-Dimensional Massive MIMO Systems. *IEEE Transactions on Wireless Communications*, 16(9), 5810-5824. <https://doi.org/10.1109/TWC.2017.2716362>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Multi-Layer Precoding: A Potential Solution for Full-Dimensional Massive MIMO Systems

Ahmed Alkhateeb, Geert Leus, and Robert W. Heath, Jr.

Abstract—Massive MIMO systems achieve high sum spectral efficiency by simultaneously serving large numbers of users. In time division duplexing systems, however, the reuse of uplink training pilots among cells results in channel estimation errors, which causes downlink inter-cell interference. Handling this interference is challenging due to the large channel dimensionality and the high complexity associated with implementing large precoding/combining matrices. In this paper, we propose multi-layer precoding to enable efficient and low-complexity operation in full-dimensional massive MIMO, where a large number of antennas are used in two dimensions. In multi-layer precoding, the precoding matrix of each base station is written as a product of a number of precoding matrices. Multi-layer precoding: 1) leverages the directional characteristics of large-scale MIMO channels to manage inter-cell interference with low channel knowledge requirements and 2) allows for an efficient implementation using hybrid analog/digital architectures. We present and analyze a specific multi-layer precoding design for full-dimensional massive MIMO systems. The asymptotic optimality of the proposed design is then proved for some special yet important channels. Numerical simulations verify the analytical results and illustrate the potential gains of multi-layer precoding compared with other multi-cell precoding solutions.

Index Terms—Multi-layer precoding, massive MIMO, full-dimensional MIMO, pilot decontamination.

I. INTRODUCTION

MASSIVE MIMO¹ promises significant spectral efficiency gains for cellular systems. Scaling up the number of antennas, however, faces several challenges that prevent the corresponding scaling of the gains [2]–[5]. First, the training and feedback of the large channels has high overhead in frequency division duplexing (FDD) systems. To overcome that, channel reciprocity in conjunction with time division duplexing (TDD) systems is used [6], [7]. Reusing

the uplink training pilots among cells, however, causes channel estimation errors which in turn lead to downlink inter-cell interference and pilot contamination [6]. Managing this inter-cell interference using traditional network MIMO techniques requires high coordination overhead, which could limit the overall system performance [8]. Another challenge with the large number of antennas lies in the hardware implementation [4], [9]. Traditional MIMO precoding techniques normally assume complete 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 techniques that can overcome the challenges of inter-cell interference and complete baseband processing is of great interest.

A. Prior Work

Inter-cell interference is a critical problem for general MIMO systems. Typical solutions for managing this interference require some sort of collaboration between the base stations (BSs) [10]. The overhead of this cooperation, though, can limit the system performance [8]. When the number of antennas grows to infinity, the performance of the network becomes limited by pilot contamination [6], 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 [11]–[14]. In [11], [12], multi-cell zero-forcing and MMSE MIMO precoding strategies were developed to cancel or reduce the inter-cell interference. The solutions in [11], [12], however, require global channel knowledge at every BS, which makes them feasible only for small numbers of antennas [15]. Pilot contamination precoding was proposed in [13] to overcome the pilot contamination problem, relying on the channel covariance knowledge. The technique in [13], though, requires sharing the transmitted messages between all BSs, which is difficult to achieve in practice. In [14], 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.

Precoding approaches that divide the processing between two stages have been developed in [16]–[20] for mmWave and massive MIMO systems. Motivated by the high cost and power consumption of the RF, [16] developed hybrid analog/digital

Manuscript received October 13, 2016; revised April 1, 2017; accepted June 2, 2017. Date of publication June 21, 2017; date of current version September 8, 2017. This work was supported by the National Science Foundation under Grant NSF-CCF-1319556. A shorter version of this paper was presented at Asilomar Conference on Signals, Systems, and Computers, November 2014 [1]. The associate editor coordinating the review of this paper and approving it for publication was C. Shen. (*Corresponding author: Robert W. Heath, Jr.*)

A. Alkhateeb and R. W. Heath, Jr., are with The University of Texas at Austin, Austin, TX 78712 USA (e-mail: aalkhateeb@utexas.edu; rheath@utexas.edu).

G. Leus is with the Delft University of Technology, 2628 CD Delft, Netherlands (e-mail: g.j.t.leus@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2017.2716362

¹In this paper, we use the term “massive MIMO” to describe any system with large numbers of antennas, and the term “full-dimensional massive MIMO” to specify systems with rectangular antenna arrays, and with large numbers of antennas in the two dimensions of the rectangular arrays.

precoding algorithms 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 [17] 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 [18], [19], with the general objective of maximizing the system sum-rate. In [20], 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 [16]–[20], however, did not consider out-of-cell interference, which ultimately limits the performance of massive MIMO systems.

B. Contribution

In this paper, we introduce a general framework, called multi-layer precoding, that (i) coordinates inter-cell interference in full-dimensional massive MIMO systems leveraging large channel characteristics and (ii) allows for efficient implementations using hybrid analog/digital architectures. Note that most of the literature on full-dimensional MIMO systems did not assume massive MIMO [21], [22], and the two systems were studied independently using different tools and theories. In this paper, we refer to full-dimensional massive MIMO as a two-dimensional MIMO system, which adopts large numbers of antennas in the two dimensions. 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 decouples 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, we prove that the proposed solution achieves an asymptotically optimal rate for two special channel models: the one-ring and the single-path models. This means that the pilot contamination problem can be overcome, in these cases, using multi-layer precoding that requires much less channel knowledge compared to previous solutions such as pilot contamination precoding [13] and multi-cell MMSE precoding [12]. Further, we characterize the deterministic equivalent rate as the number of antennas and users grow with a fixed ratio. These results can be leveraged for network optimization and user scheduling objectives.

The developed multi-layer precoding solutions are also evaluated by numerical simulations. Results show the

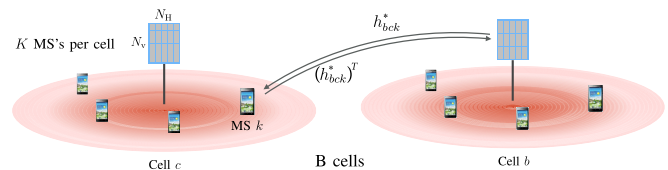


Fig. 1. A full-dimensional MIMO cellular model where each BS has a 2D antenna array and serves K users.

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 good rate and coverage gains can be obtained by multi-layer precoding compared to multi-cell MMSE precoding despite the requirement for less channel knowledge.

We use the following notation throughout this paper: \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. $[\mathbf{A}]_{r,:}$ and $[\mathbf{A}]_{:,c}$ are the r th row and c th column of the matrix \mathbf{A} , respectively. $(\mathbf{AB})_{k,k}^{-1}$ is the element in the k th row and k th column of the matrix $(\mathbf{AB})^{-1}$. $\text{diag}(\mathbf{a})$ 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. $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of \mathbf{A} and \mathbf{B} , and $\mathbf{A} \circ \mathbf{B}$ is their Khatri-Rao product. $\mathcal{N}_{\mathbb{C}}(\mathbf{m}, \mathbf{R})$ is a complex Gaussian random vector with mean \mathbf{m} and covariance \mathbf{R} . $\mathbb{E}[\cdot]$ is used to denote expectation.

II. SYSTEM AND CHANNEL MODELS

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

A. System Model

Consider a cellular system model consisting of B cells, as shown in Fig. 1, where every cell has one BS and K MSs. Each BS is equipped with a two-dimensional (2D) antenna array of N elements, $N = N_v N_h$, where N_v and N_h are the numbers of the antennas at the vertical and horizontal dimensions of the 2D array. Each MS is assumed to have a single antenna. We assume that all BSs and MSs are synchronized on the level of the frames, and operate a TDD protocol with universal frequency reuse. In the downlink, each BS $b = 1, 2, \dots, B$, applies an $N \times K$ precoder \mathbf{F}_b to transmit a symbol for each user, with a power constraint $\|\mathbf{F}_b\|_{:,k}^2 = 1, k = 1, 2, \dots, K$. Uplink and downlink channels are assumed to be reciprocal. If \mathbf{h}_{bck}^* denotes the $1 \times N$ downlink channel from BS b to user k in cell c , then the received signal by this user can be written as

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

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

TABLE I
SYSTEM MODEL PARAMETERS

Notation	Description
N, N_V, N_H	The total number of antennas, the number of vertical antennas, and the number of horizontal antennas.
B, K	The number of cells, and the number of users per cell.
\mathbf{F}_b	The downlink precoder of the BS at cell b .
\mathbf{h}_{bck}	The $N \times 1$ uplink channel from user k in cell c to the b -th cell BS.
$\mathbf{R}_{bck}, \mathbf{R}_{bck}^A, \mathbf{R}_{bck}^E$	the covariance, azimuth covariance, and elevation covariance of the channel \mathbf{h}_{bck} .
\mathbf{R}_c^I	The averaged inter-cell interference covariance matrix defined in (11).
$\mathbf{U}_{bck}^A, \mathbf{U}_{bck}^E$	The matrices containing the singular vectors of the azimuth and elevation covariance matrices of \mathbf{h}_{bck} .
$\mathbf{U}_c^I, \mathbf{U}_c^{NI}$	The matrices containing the eigenvectors of \mathbf{R}_c^I that correspond to the non-zero and zero eigenvalues.
$\bar{\mathbf{U}}_{cck}$	The effective eigenvectors of the covariance matrix of user k at cell c , $\bar{\mathbf{U}}_{cck} = \mathbf{U}_{cck}^A \otimes \mathbf{U}_c^{NI} \mathbf{U}_{cck}^E$.

to expand (1) as

$$y_{ck} = \underbrace{\mathbf{h}_{cck}^* [\mathbf{F}_c]_{:,k} s_{c,k}}_{\text{Desired signal}} + \underbrace{\sum_{m \neq k} \mathbf{h}_{cck}^* [\mathbf{F}_c]_{:,m} s_{c,m}}_{\text{Intra-cell interference}} + \underbrace{\sum_{b \neq c} \mathbf{h}_{bck}^* \mathbf{F}_b s_b}_{\text{Inter-cell interference}} + n_{ck}, \quad (2)$$

to illustrate the different components of the received signal.

B. 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 [23], [24]. One candidate is the Kronecker product correlation model, which provides a reasonable approximation of 3D covariance matrices [25]. In this model, the covariance of the 3D channel \mathbf{h}_{bck} , defined as $\mathbf{R}_{bck} = \mathbb{E}[\mathbf{h}_{bck} \mathbf{h}_{bck}^*]$, is approximated by

$$\mathbf{R}_{bck} = \mathbf{R}_{bck}^A \otimes \mathbf{R}_{bck}^E, \quad (3)$$

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

$$\mathbf{h}_{bck} = \left[\mathbf{U}_{bck}^A \mathbf{\Lambda}_{bck}^A \frac{1}{2} \otimes \mathbf{U}_{bck}^E \mathbf{\Lambda}_{bck}^E \frac{1}{2} \right] \mathbf{w}_{bck}, \quad (4)$$

where $\mathbf{w}_{bck} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a rank(\mathbf{R}_{bck}^A) rank(\mathbf{R}_{bck}^E) $\times 1$ vector, with rank(\mathbf{A}) representing the rank of the matrix \mathbf{A} . Without loss of generality, and to simplify the notation, 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 . To facilitate following the paper, we added the key system and channel parameters to Table I.

III. 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 II, the signal-to-interference-plus-noise ratio (SINR) at user k in cell c is

$$\text{SINR}_{ck} = \frac{\frac{P}{K} |\mathbf{h}_{cck}^* [\mathbf{F}_c]_{:,k}|^2}{\frac{P}{K} \sum_{m \neq k} |\mathbf{h}_{cck}^* [\mathbf{F}_c]_{:,m}|^2 + \frac{P}{K} \sum_{b \neq c} \|\mathbf{h}_{bck}^* \mathbf{F}_b\|^2 + \sigma^2}}, \quad (5)$$

where $|\mathbf{h}_{cck}^* [\mathbf{F}_c]_{:,k}|^2$ is the desired signal power, $\sum_{m \neq k} |\mathbf{h}_{cck}^* [\mathbf{F}_c]_{:,m}|^2$ is the intra-cell multi-user interference, and $\sum_{b \neq c} \|\mathbf{h}_{bck}^* \mathbf{F}_b\|^2$ is the inter-cell interference. Designing one precoding matrix per BS 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 [10]. Also, coordinating inter-cell interference between BS's typically results in high cooperation overhead that makes the value of this cooperation limited [8]. 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 , $b = 1, 2, \dots, B$, 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 [17], i.e., with a small number of RF chains compared to the number of antennas. 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 [17] and joint spatial division multiplexing [20], and leveraging the directional characteristics of large-scale MIMO channels [14], 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

$$\mathbf{F}_c = \mathbf{F}_c^{(1)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)}, \quad (6)$$

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 is 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], [16], [17], [26] 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.

IV. 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 II. Following the proposed multi-layer precoding criteria explained in Section III, we propose to design the $N_V N_H \times K$ precoding matrix \mathbf{F}_b of cell b , $b = 1, \dots, B$ as

$$\mathbf{F}_b = \mathbf{F}_b^{(1)} \mathbf{F}_b^{(2)} \mathbf{F}_b^{(3)}, \quad (7)$$

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 for 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 (7), we get

$$y_{ck} = \underbrace{\mathbf{h}_{cck}^* \mathbf{F}_c^{(1)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} \mathbf{s}_c}_{\text{received signal from serving BS}} + \underbrace{\sum_{b \neq c} \mathbf{h}_{bck}^* \mathbf{F}_b^{(1)} \mathbf{F}_b^{(2)} \mathbf{F}_b^{(3)} \mathbf{s}_b}_{\text{received signal from other BSs}} + n_{ck}. \quad (8)$$

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

A. 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 (8). Exploiting the Kronecker structure of the channel model in (4), we propose to construct the first layer as

$$\mathbf{F}_b^{(1)} = \mathbf{F}_b^{\text{A}(1)} \otimes \mathbf{F}_b^{\text{E}(1)}. \quad (9)$$

Adopting the channel model in (4) with $\bar{\mathbf{w}}_{bck} = \left(\Lambda_{bck}^{\text{A} \frac{1}{2}} \otimes \Lambda_{bck}^{\text{E} \frac{1}{2}} \right) \mathbf{w}_{bck}$ and employing the Kronecker precoding structure in (9), the second term of the received signal y_{ck} in (8) can be expanded as

$$\begin{aligned} & \sum_{b \neq c} \mathbf{h}_{bck}^* \mathbf{F}_b^{(1)} \mathbf{F}_b^{(2)} \mathbf{F}_b^{(3)} \mathbf{s}_b \\ &= \sum_{b \neq c} \bar{\mathbf{w}}_{bck}^* \left(\mathbf{U}_{bck}^{\text{A}*} \mathbf{F}_b^{\text{A}(1)} \otimes \mathbf{U}_{bck}^{\text{E}*} \mathbf{F}_b^{\text{E}(1)} \right) \mathbf{F}_b^{(2)} \mathbf{F}_b^{(3)} \mathbf{s}_b. \end{aligned} \quad (10)$$

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

The focus on the elevation direction in managing inter-cell interference has two main reasons. First, with large numbers of vertical antennas and for several channel models, the elevation covariance matrices of the desired users and other cell users will occupy different subspaces, or will have slightly overlapped subspaces, as will be shown in Section V. This means that the desired user and other cell user channels can be more easily separated in the elevation dimension as compared to the azimuth dimension. Further, 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 $\text{Null} \left(\sum_{b \neq c} \sum_{k \in \mathcal{X}_b} \mathbf{R}_{cbk}^{\text{E}} \right)$ of different scheduled users \mathcal{X}_b will have a large overlap. This implies that designing $\mathbf{F}_c^{\text{E}(1)}$ based on the interference elevation 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

$$\mathbf{R}_c^{\text{I}} = \frac{1}{(B-1)} \sum_{b \neq c} \mathbb{E}_{\mathcal{X}_b} \left[\frac{1}{K} \sum_{k \in \mathcal{X}_b} \mathbf{R}_{cbk}^{\text{E}} \right]. \quad (11)$$

where the expectation is taken over different sets \mathcal{X}_b of scheduled users in cell $b, b \neq c$. In this paper, we manage the inter-cell interference in the elevation space, and therefore, set $\mathbf{F}_b^{\text{A}(1)} = \mathbf{I}_{N_H}$. Let $[\mathbf{U}_c^{\text{I}} \ \mathbf{U}_c^{\text{NI}}] \Lambda_c [\mathbf{U}_c^{\text{I}} \ \mathbf{U}_c^{\text{NI}}]^*$ represent the eigen-decomposition of \mathbf{R}_c^{I} with the $N_V \times r_1$ matrix \mathbf{U}_c^{I} and $N_V \times r_{\text{NI}}$ matrix \mathbf{U}_c^{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

$$\mathbf{F}_c^{(1)} = \mathbf{I}_{N_H} \otimes \mathbf{U}_c^{\text{NI}}, \quad (12)$$

which is an $N_V N_H \times r_{\text{NI}} N_H$ matrix.

Given the design of the first precoding layer in (12), and defining the $r_{\text{NI}} \times r_E$ effective elevation eigen matrix $\bar{\mathbf{U}}_{cck}^E = \mathbf{U}_c^{\text{NI}*} \mathbf{U}_{cck}^E$, the received signal at user k of cell c in (8) becomes

$$y_{ck} = \bar{\mathbf{w}}_{cck}^* \left(\mathbf{U}_{cck}^{A*} \otimes \bar{\mathbf{U}}_{cck}^{E*} \right) \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} \mathbf{s}_c + n_{ck}. \quad (13)$$

Note that the first precoding layer in (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 [27], [28]. Relaxations of the precoding design in (9) are proposed in Section VI to compromise between inter-cell interference avoidance and desired signal power degradation.

Required Channel Knowledge: The design of the first precoding layer in (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. The channel covariance knowledge can normally be considered constant over a much longer time period compared to the channel coherence time [20], [29], which enables its efficient estimation. Further, this overhead is expected to be smaller in our case since we only need an averaged interference covariance knowledge. Moreover, as the first layer precoding design requires the knowledge of the *subspace* of this averaged interference covariance matrix and not the full matrix, efficient massive MIMO subspace estimation approaches, [29], [30], can be leveraged.

Note that one key advantage of the decoupled multi-layer precoding structure that allows dedicating one layer for canceling the out-of-cell interference based on 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 approaches that rely 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.

B. 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 $\bar{\mathbf{U}}_{cck} = \left(\mathbf{U}_{cck}^A \otimes \bar{\mathbf{U}}_{cck}^E \right)$, then we design the second precoding

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

$$\mathbf{F}_c^{(2)} = \left[\bar{\mathbf{U}}_{cck1}, \dots, \bar{\mathbf{U}}_{cckK} \right], \quad (14)$$

which has $N_H r_{\text{NI}} \times K r_A r_E$ dimensions. Given the second precoding layer design, and defining $\mathbf{G}_{c,(k,r)} = \bar{\mathbf{U}}_{cck}^* \bar{\mathbf{U}}_{cckr}$, the received signal by user k in cell c can be written as

$$y_{ck} = \bar{\mathbf{w}}_{cck}^* \left[\mathbf{G}_{c,(k,1)}, \dots, \mathbf{G}_{c,(k,K)} \right] \mathbf{F}_c^{(3)} \mathbf{s}_c + n_{ck}. \quad (15)$$

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 large numbers of antennas are employed. This reduces the overhead associated with training the effective channels, which is particularly important for FDD systems [17], [20]. Second, this precoding layer supports the multiplicative structure of multi-layer precoding with successive dimensionality reduction, which simplifies its implementation using hybrid analog/digital architectures, as will be briefly discussed in Section VI.

Required Channel Knowledge: The design of the second precoding layer requires only the knowledge of the effective eigenvector matrices $\bar{\mathbf{U}}_{cck}, k = 1, \dots, K$, which depends on the large-scale channel statistics. It is worth noting that during the uplink training of the matrices $\bar{\mathbf{U}}_{cck}$, 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. Obtaining the matrices $\bar{\mathbf{U}}_{cck}$ can be achieved by estimating the azimuth and elevation subspaces $\mathbf{U}_{cck}^A, \bar{\mathbf{U}}_{cck}^E$, which can be done using efficient subspace estimation techniques [29], [30], as discussed in Section IV-A.

C. 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 $\bar{\mathbf{h}}_{ck} = \left[\mathbf{G}_{c,(k,1)}, \dots, \mathbf{G}_{c,(k,K)} \right]^* \bar{\mathbf{w}}_{cck}$, and let $\bar{\mathbf{H}}_c = [\bar{\mathbf{h}}_{c1}, \dots, \bar{\mathbf{h}}_{cK}]$, then we construct the third precoding layer $\mathbf{F}_c^{(3)}$ as a zero-forcing matrix

$$\mathbf{F}_c^{(3)} = \bar{\mathbf{H}}_c \left(\bar{\mathbf{H}}_c^* \bar{\mathbf{H}}_c \right)^{-1} \boldsymbol{\Upsilon}_c, \quad (16)$$

where $\boldsymbol{\Upsilon}_c$ is a diagonal power normalization matrix that ensures satisfying the precoding power constraint $\| [\mathbf{F}_b]_{:,k} \|^2 = 1$. The dimensions of the third precoding layer $\mathbf{F}_c^{(3)}$ is $K r_A r_E \times K$. Note that this zero-forcing design requires that the effective channel matrix $\bar{\mathbf{H}}_c$ to be a tall matrix, which is always satisfied as the dimensions of this effective channel are $K r_A r_E \times K$. Given the design of the precoding matrix $\mathbf{F}_c^{(3)}$, the received signal at user k in cell c can be expressed as

$$y_{ck} = [\boldsymbol{\Upsilon}_c]_{k,k} s_{c,k} + n_{ck}. \quad (17)$$

1) *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 should have much smaller dimensions compared to the original channels in massive MIMO systems, which reduces the required training overhead.

2) *Example on the Successive Dimensionality Reduction*: To illustrate the successive reduction in the dimensions of the precoding layers, consider an example of the system model in Section II-A with a BS employing a rectangular array of $N_V = 100$, $N_H = 30$ antennas and serving $K = 15$ users. If the azimuth and elevation covariance matrices of every user have ranks $r_A = 4$, $r_E = 4$, and the rank of the elevation inter-cell interference null-space is $r_{NI} = 40$ (note that r_{NI} should be less than N_V). Then, the dimensions of the first, second, and third precoding layers are $N_V N_H \times r_{NI} N_H = 3000 \times 1200$, $r_{NI} N_H \times K r_A r_E = 1200 \times 240$, and $K r_A r_E \times K = 240 \times 15$, respectively. Further, the dimensions of the effective channels of every desired user seen after the first layer are 1200×1 , and after the second layer become 240×1 , which are successively reduced compared to the dimensions of the original channels, 3000×1 .

V. PERFORMANCE ANALYSIS

The proposed multi-layer precoding design in Section IV eliminates inter-cell interference as well as multi-user intra-cell interference, assuming that every BS b has the knowledge of its users' effective channels and channel covariance $\bar{\mathbf{H}}_b$, $\{\mathbf{R}_{cck}\}$ and the averaged inter-cell interference covariance in the elevation dimension \mathbf{R}_c^I . 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 (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 1. Then, we will show that this precoding design can achieve optimal performance for some special yet important channel models in Section V-A and Section V-B.

Lemma 1: Consider the system and channel models in Section II and the multi-layer precoding design in Section IV. The achievable rate by user k in cell c is given by

$$R_{ck} = \log_2 \left(1 + \frac{\text{SNR}}{\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)_{k,k}^{-1}} \right), \quad (18)$$

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

Proof: See Appendix A \blacksquare

Note that the achievable rate in (18) is upper bounded by the single-user rate—the rate when the user is solely served in the network—which is given by $\bar{R}_{ck} = \log_2 (1 + \text{SNR} \|\bar{\mathbf{w}}_{cck}\|^2)$. Therefore, Lemma 1 indicates that the proposed multi-layer precoding can achieve optimal performance if $\mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} = \mathbf{I}$. To achieve that, it is sufficient to satisfy the following 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.

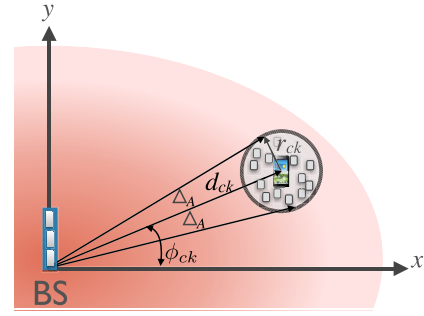


Fig. 2. An illustration of the one-ring channel model in the azimuth direction. The BS that has a UPA in the y - z plane, serves a mobile user in the x - y plane at distance d_{ck} . The user is surrounded by scatterers on a ring of radius r_{ck} , and its channel experiences an azimuth angular spread Δ_A .

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

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.

A. 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 [31]–[34]. 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 [14], [20], [35], [36].

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_{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. 2. If the mobile is surrounded by scatterers on a ring of radius r_{ck} in the azimuth dimension, then the azimuth angular spread Δ_A can be approximated as $\Delta_A = \arctan\left(\frac{r_{ck}}{d_{ck}}\right)$. 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

$$\left[\mathbf{R}_{cck}^A \right]_{n_1, n_2} = \frac{1}{2\Delta_A} \int_{-\Delta_A}^{\Delta_A} e^{-j\frac{2\pi}{\lambda} d(n_2 - n_1) \sin(\phi_{ck} + \alpha) \sin(\theta_{ck})} d\alpha. \quad (19)$$

The elevation correlation matrix can be similarly defined for the user k , in terms of its elevation angular spread Δ_E .

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

Proposition 2: Consider the full-dimensional cellular system model in Section II-A with cells of radius r_{cell} , and the

channel model in Section II-B with the one-ring correlation matrices in (19). Let ϕ_{ck}, θ_{ck} denote the azimuth and elevation angles of user k at cell c , and let Δ_A, Δ_E represent the azimuth and elevation angular spread. Define the maximum distance with no blockage on the desired signal power as $d_{\max} = H_{\text{BS}} \tan\left(\arctan\left(\frac{r_{\text{cell}}}{H_{\text{BS}}}\right) - 2\Delta_E\right)$. If $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_A$ or $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_E, \forall m \neq k$, 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 IV, satisfies

$$\lim_{N_V, N_H \rightarrow \infty} R_{ck} = \bar{R}_{ck} = \log_2\left(1 + \text{SNR} \|\bar{\mathbf{w}}_{cck}\|^2\right). \quad (20)$$

Proof: See Appendix B ■

Proposition 2 indicates that the achievable rate with multi-layer precoding converges to the optimal single-user rate for the users that are not at the cell edge ($r_{\text{cell}} - d_{\max}$ 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 $\Delta_E = 3^\circ$, then all the users within $\sim 80\text{m}$ distance from the BS achieve 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 Proposition 2 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 VII 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 proposition, we derive a lower bound on the achievable rate with multi-layer precoding for the cell-edge users.

Proposition 3: Consider the same system and channel models described in Proposition 2. If $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_A$ or $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_E, \forall m \neq k$, and $d_{\max} \leq d_{ck} \leq r_{\text{cell}}$, then the achievable rate of user k at cell c , when applying the multi-layer precoding algorithm in Section IV, satisfies

$$\lim_{N_V, N_H \rightarrow \infty} R_{ck} \geq \log_2\left(1 + \text{SNR} \|\bar{\mathbf{w}}_{cck}\|^2 \sigma_{\min}^2\left(\bar{\mathbf{U}}_{cck}^E\right)\right), \quad (21)$$

where $\sigma_{\min}(\mathbf{A})$ denotes the minimum singular value of the matrix \mathbf{A} .

Proof: See Appendix C ■

Proposition 3 indicates that cell edge users experience some degradation in their SNRs as a cost for the perfect inter-cell interference avoidance. In Section VI, 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.

One main advantage of employing large numbers of antennas at the base station is to simultaneously serve large numbers

of users. In the next proposition, we use the deterministic equivalent approach [37] to derive the achievable rate by the proposed multi-layer precoding algorithm, when the number of users scales linearly with the number of antennas.

Proposition 4: Consider the full-dimensional cellular model in Section II-A with a total power constraint $\|\mathbf{F}_c\|_F^2 \leq P$, and the channel model in Section II-B with the one-ring correlation matrices in (19). If $N_V, N_H \rightarrow \infty$, and assuming that the number of users K scales linearly with N , we get

$$R_{ck} - \dot{R}_c \xrightarrow{N \rightarrow \infty} 0, \quad (22)$$

where \dot{R}_c is a deterministic value for all the users, and can be calculated for any finite number of antennas as

$$\dot{R}_c = \log_2\left(1 + \frac{P/\sigma^2}{\frac{1}{N} \sum_{k=1}^K \frac{b_k}{a_k}}\right), \quad (23)$$

with a_k and b_k computed from the fixed-point equations

$$a_k = \frac{1}{N} \text{tr}(\mathbf{Q}_k \mathbf{T}), \quad (24)$$

$$b_k = \frac{1}{N} \text{tr}\left(\frac{1}{N} \sum_{m=1}^K \frac{b_m}{a_m} \mathbf{Q}_k \mathbf{T} \mathbf{Q}_m \mathbf{T} + \mathbf{Q}_k \mathbf{T} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{T}\right), \quad (25)$$

$$\mathbf{T} = \left(\frac{1}{N} \sum_{m=1}^K \frac{\mathbf{Q}_m}{a_m} + \mathbf{I}\right)^{-1}, \quad (26)$$

with $\mathbf{Q}_k = \mathbf{F}_c^{(2)*} \bar{\mathbf{U}}_{cck} \Lambda_{cck} \bar{\mathbf{U}}_{cck}^* \mathbf{F}_c^{(2)}, k = 1, \dots, K$.

Proof: The proof exploits and extends the results and theorems in [37], and is omitted due to space limitations. ■

The deterministic equivalent results can be used for the system optimization and user scheduling, as well as deriving insights into the relation between the different system parameters [38].

B. 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 channels, such as mmWave systems [28], [39], [40]. 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} . Then, the channel vector can be expressed as

$$\mathbf{h}_{cck} = \rho_{cck}^{\frac{1}{2}} \beta_{ck} \mathbf{a}_A(\phi_{ck}, \theta_{ck}) \otimes \mathbf{a}_E(\phi_{ck}, \theta_{ck}), \quad (27)$$

where $\mathbf{a}_A(\phi_{ck}, \theta_{ck})$ and $\mathbf{a}_E(\phi_{ck}, \theta_{ck})$ 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 5: Consider the full-dimensional cellular system model in Section II-A, and the single-path channel model

in (27). When applying the multi-layer precoding algorithm in Section IV, the achievable rate of user k at cell c satisfies

$$\lim_{N_V, N_H \rightarrow \infty} R_{ck} = \bar{R}_{ck} = \log_2 \left(1 + \text{SNR} \|\mathbf{h}_{cck}\|^2 \right). \quad (28)$$

Proof: The proof is similar to that of Proposition 2, and is omitted due to space limitations. ■

Corollary 5 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 VII.

VI. 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.

A. The Order of the Precoding Layers

In this paper, we proposed a three-layer precoding design in Section IV, which adopts a certain order of the precoding functions. The main intuition behind this order is the time scale of the channel knowledge needed for these functions. Managing the inter-cell interference in the elevation subspace can be achieved using very slowly varying channel knowledge—the average of the other cell user covariance matrices. This is why the first precoding layer is responsible for handling the inter-cell interference. For the intra-cell multi-user interference management, it needs to be done on a faster time scale and requires very accurate channel knowledge. This is the motivation of assigning the multi-user interference management to the third precoding layers. Note that every precoding layer needs to be designed based on the effective channel, that includes the impact of the preceding layers. That is why it is important to have the precoding layer that does not require frequent updates first. For the proposed multi-layer precoding design, we showed its asymptotic optimality in some channel models in Section V. Investigating different orders and functions for the precoding layers, though, is an interesting research direction.

B. 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 channel covariance knowledge 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 dimensionality 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 [20] 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 on 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.

C. Multi-Layer Precoding With Augmented Vertical Dimensions

As explained in Section IV, 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 the structure of large channels. This is summarized as follows. Consider a rectangular array in the $y-z$ plane. Then, leveraging Lemma 2 in [14], the rank of the one-ring elevation covariance matrix can be related to its angular range $[\theta_{\min}, \theta_{\max}]$, with θ_{\min} and θ_{\max} representing the minimum and maximum angles measured from the positive z -axis, as

$$\text{rank}(\mathbf{R}^E) = \frac{N_V D}{\lambda} (\cos(\theta_{\min}) - \cos(\theta_{\max})) \text{ as } N_V \rightarrow \infty. \quad (29)$$

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_c^I , as

$$\theta_c^I = \arccos \left(-\frac{\text{rank}(\mathbf{R}_c^I) \lambda}{N_V D} \right). \quad (30)$$

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

$$\left[\bar{\mathbf{R}}_c^I \right]_{n_1, n_2} = \frac{1}{\theta_c^I - \delta_E - \pi/2} \int_{\frac{\pi}{2}}^{\theta_c^I - \delta_E} e^{jkD(n_2 - n_1) \cos(a)} da. \quad (31)$$

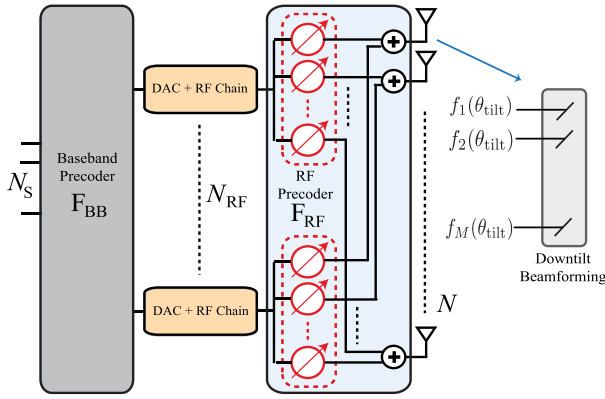


Fig. 3. The figure shows a hybrid analog/digital architecture, at which baseband precoding, RF precoding, and antenna downtilt beamforming can be utilized to implement the multi-layer precoding algorithm.

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

$$\mathbf{F}_c^{(1)} = \mathbf{I} \otimes \bar{\mathbf{U}}_c^{NI}. \quad (32)$$

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 VII.

D. Multi-Layer Precoding Using Hybrid Architectures

Thanks to the multiplicative structure and the specific multi-layer precoding design in Section IV, 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], [16], [17], [26], 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 IV, we propose to implement the first and second layers in the RF domain and perform the third layer precoding at baseband, as depicted in Fig. 3. Given the successive dimensionality 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 [22], [23]. For example, the 3GPP antenna port elevation gain $G^E(\theta)$ is defined as [23]

$$G^E(\theta) = G_{\max}^E - \min \left\{ 12 \left(\frac{\theta - \theta_{\text{tilt}}}{\theta_{3\text{dB}}} \right)^2, \text{SL} \right\}, \quad (33)$$

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

Once $\mathbf{F}_c^{(1)}$ is implemented, the second precoding layer $\mathbf{F}_c^{(2)}$ can be designed similar to [17], 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.

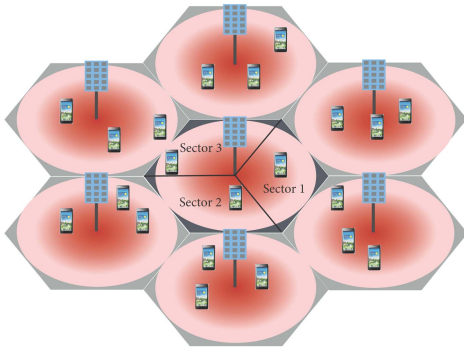
The proposed multi-layer precoding design requires the knowledge of the subspaces of the averaged inter-cell interference covariance and the desired users' channel covariance matrices. Estimating channel covariance matrices with hybrid architectures is generally more challenging compared with fully-digital transceivers, as the channel is seen at the baseband through the RF beamforming lens. The estimation and tracking of the covariance subspace, though, is generally easier than obtaining the full covariance matrices, and several efficient techniques have been developed in the last few years to address this problem [29], [30], [41], [42]. This remains, however, an active and interesting research area that deserves further investigation.

VII. 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, with three 120° sectors per cell, as depicted in Fig. 4(a), and calculate the performance for the cell in the center. For every sector, the BS has a rectangular array at a height $H_{\text{BS}} = 35\text{m}$, and serving users at cell radius $r_{\text{cell}} = 100\text{m}$. Users are randomly and uniformly dropped in the cells, and every cell sector 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. Note that with these system parameters, link budget analysis yields a receive SNR of about 10dB for a user at a distance 80m from the BS. Two channel models are assumed, namely, the single-path and the one-ring channel models.

The simulation results compare four setups: the single-user case, multi-layer precoding, multi-cell MMSE precoding, and single-cell conjugate beamforming. In the single-user case, which represents an upper bound, only one user is assumed to be served in the system. This user is randomly selected from those assigned to the BS. For multi-layer precoding, the desired BS is assumed to know (i) the covariance matrices of the channels to its users which are obtained directly from the system geometry, (ii) the averaged elevation inter-cell interference covariance matrix \mathbf{R}_c^I by averaging the elevation covariance matrices of the channels between the desired BS and the other cell users over 40 realizations of scheduled



(a) FD massive MIMO cellular model

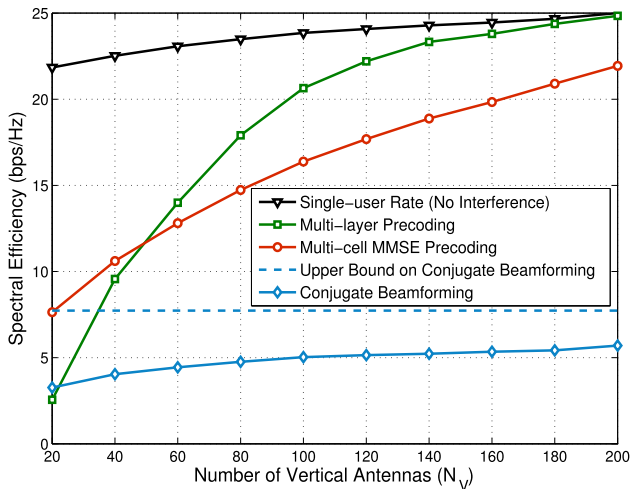

 (b) $N_H = 30$

Fig. 4. The adopted single-tier (7-cells) sectored 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, multi-cell MMSE precoding, and single-cell conjugate beamforming, for different numbers of vertical antennas. The number of BS horizontal antennas is $N_H = 30$, and the users are assumed to have single-path channels.

users, and (iii) the effective channel vectors to the desired BS, which are obtained by adding the effective channels of all the users in the system sharing the same pilot. This simulates the interference of the other cells' users in the channel estimation phase. Note that we assume a universal pilot reuse, i.e., all the cells randomly assign the same K orthogonal pilots to its users. Given this channel knowledge, the BSs apply the multi-layer precoding algorithm in Section IV. In the multi-cell MMSE precoding, the precoding matrices are constructed according to [12], assuming perfect knowledge of the covariance matrices to all the users in the system. The perfect channel covariance knowledge is also used to generate the instantaneous channel vectors as [38, Section II-C]. For the single-cell conjugate beamforming, the channels are obtained by adding the co-pilot user channels, which simulates the result of using matched filters during the channel estimation phase [6]. Note that the performance of conjugate beamforming can be further improved, compared to what we simulate in this section, by leveraging the channel covariance knowledge, [14].

A. Results With Single-Path Channels

In this section, we adopt a single-path model for the user channels as described in (27). The azimuth and elevation angles are geometrically determined based on users' locations relative to the BSs, and the complex path gains $\beta_{ck} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

1) *Optimality With Large Antennas:* In Fig. 4(b), we compare the per-user achievable rate of multi-layer precoding with the single-user rate, and the rates with multi-cell MMSE precoding, [12], and single-cell conjugate beamforming. The BSs are assumed to employ UPAs that have $N_H = 30$ horizontal antennas and different numbers of vertical antennas. First, we note that the per-user achievable rates 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 5. Note that the single-user rate is the rate if only this user is served in the network, i.e., with no inter-cell or multi-user intra-cell interference. The rate of In the figure, we also plot the achievable rate with conventional conjugate beamforming. As a function of the path-loss ρ_{bck} in (27), the conjugate beamforming rate is theoretically bounded from above by [6]

$$\bar{R}_{ck}^{CB} = \log_2 \left(1 + \text{SNR} \frac{\rho_{cck}^2}{\sum_{b \neq c} \rho_{bck}^2} \right), \quad (34)$$

which limits its rate from growing with the number of antennas 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. The intuition behind that lies in the inter-cell interference avoidance using averaged channel covariance knowledge 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 antenna numbers, Fig. 4(b) shows it can still achieve gain over conventional massive MIMO beamforming schemes at much lower number of antennas. Fig. 4(b) also shows that multi-cell MMSE approaches the single-use rate, but at higher number of antennas. At a small number of antennas, though, multi-cell MMSE achieves a higher rate compared to multi-layer precoding, as the first precoding layer filters the cell-edge users power when the number of antennas is not very large.

2) *Impact of Antenna Heights and Cell Radii:* In Fig. 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 as in Fig. 4(b). In Fig. 5(a), the achievable rates are plotted 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

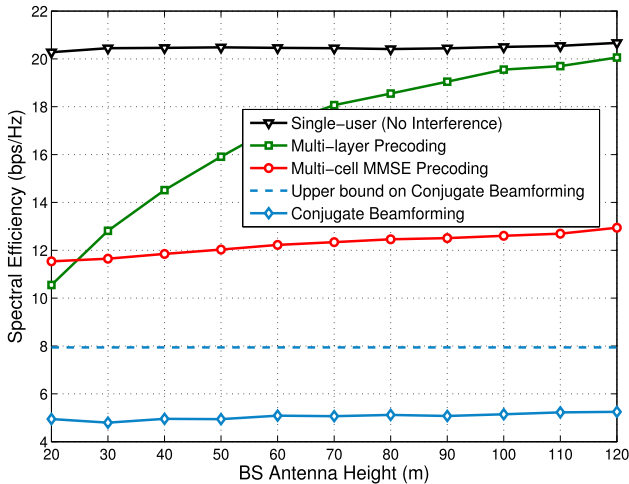
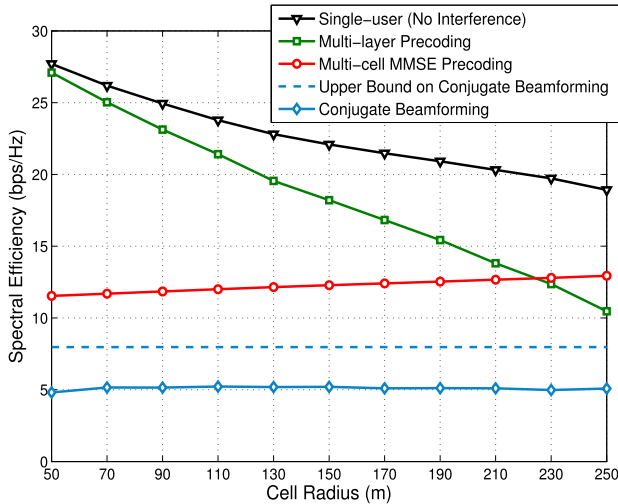
(a) $r_{\text{cell}} = 200\text{m}$ (b) $H_{\text{BS}} = 35\text{m}$

Fig. 5. The achievable rate of the proposed multi-layer precoding is compared to the single-user rate and the rate with multi-cell MMSE precoding and single-cell conjugate beamforming for different cell radii. The BSs are assumed to employ 120×30 UPAs, and the users have single-path channels.

when large arrays are deployed. These achievable rates are again compared in Fig. 5(b), but for different cell radii. This figure illustrates that a 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 smaller 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 multi-cell MMSE precoding and single-cell conjugate beamforming.

3) *Rate Coverage*: If R_{ck} denotes the rate of a randomly selected user k in cell c , and R_{th} (bps/Hz) represents a certain rate threshold, then the rate coverage at this threshold is defined as $\mathbb{P}(R_{ck} \geq R_{\text{th}})$. To evaluate the rate coverage of multi-layer precoding, we plot Fig. 6. The same setup of Fig. 4(b) is adopted again with cells of radius 100m, and

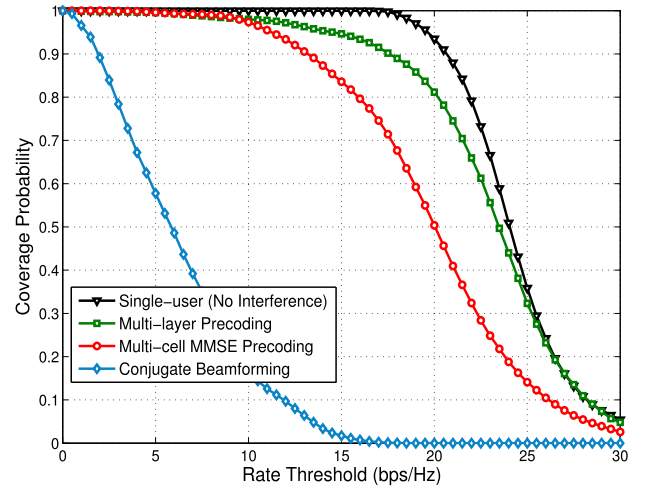


Fig. 6. The rate coverage gain of the proposed multi-layer precoding algorithm over conventional conjugate beamforming and zero-forcing is illustrated. This rate coverage is also shown to be close to the single-user case. The BSs are assumed to employ 120×30 UPAs at heights $H_{\text{BS}} = 35\text{m}$, the cell radius is $r_{\text{cell}} = 100\text{m}$, and the users have single-path channels.

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, $\sim 50\%$ 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 that multi-layer precoding achieves a close coverage to multi-cell MMSE precoding, despite the requirement of less channel knowledge.

B. Results With One-Ring Channels

In this section, we adopt a one-ring model for the user channels as described in (19). The azimuth and elevation angles are geometrically determined based on users' locations relative to the BSs, and the angular spread is set to $\Delta_A = 5^\circ$, $\Delta_E = 3^\circ$. Every BS randomly selects $K = 20$ users to be served, i.e., no scheduling is done to guarantee the angular separation condition in Proposition 2 and Proposition 3.

1) *Rate Coverage*: In Fig. 7, we compare the rate coverage of multi-layer precoding, single-user, multi-cell MMSE precoding, and single-cell conjugate beamforming. Fig. 7 considers the system model in Section II with 120×30 BS UPAs and one-ring channel model. First, the figure shows that multi-layer precoding achieves close coverage to the single-user case at the cell center. For the cell edge, though, multi-layer precoding users experience some blockage, and achieves less rate compared with multi-cell MMSE precoding. This blockage results from the elevation inter-cell interference avoidance which also affects the desired users. This blockage can be reduced using the modified augmented elevation subspace algorithm in Section VI-C. Fig. 7 also illustrates that multi-layer precoding achieves a reasonable gain over multi-cell MMSE precoding at higher rate thresholds.

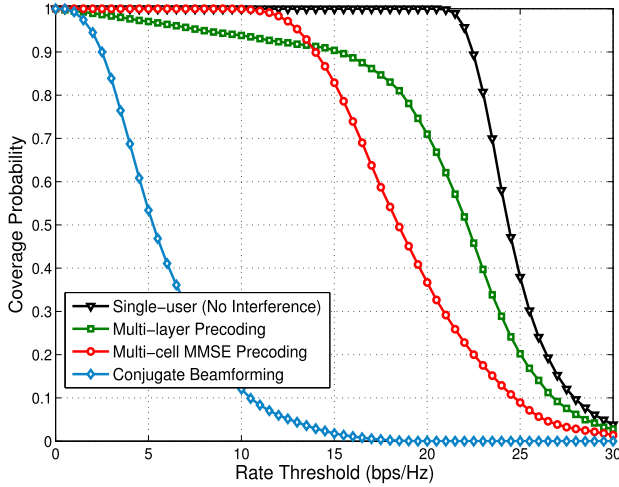


Fig. 7. The rate coverage of the proposed multi-layer precoding algorithms compared to multi-cell MMSE precoding and single-cell conjugate beamforming. This rate coverage is shown to be close to the single-user case. The BSs are assumed to be at heights $H_{BS} = 35\text{m}$, the cell radius is $r_{cell} = 100\text{m}$. The users have one-ring channel models of azimuth and elevation angular spread $\Delta_A = 5^\circ$, $\Delta_E = 3^\circ$.

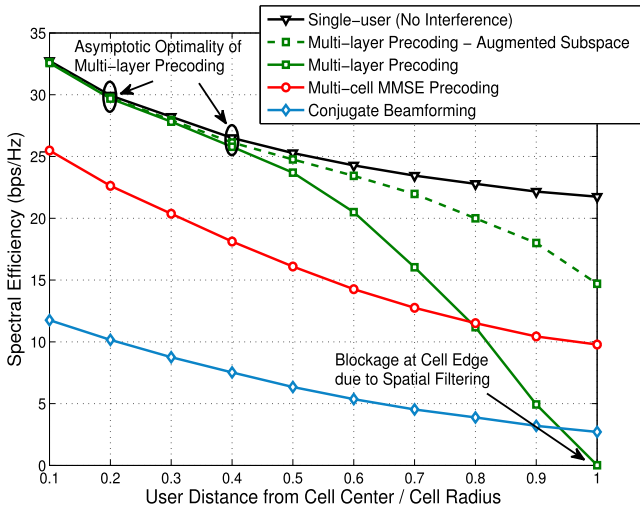


Fig. 8. The achievable rates of the proposed multi-layer precoding algorithms are compared to the single-user rate and the rates with multi-cell MMSE and single-cell conjugate beamforming, for different distances from the cell center. The BSs are assumed to employ 80×30 UPAs at heights $H_{BS} = 35\text{m}$ and the cell radius is $r_{cell} = 100\text{m}$. The users have one-ring channel models of azimuth and elevation angular spread $\Delta_A = 5^\circ$, $\Delta_E = 3^\circ$.

2) *Rates at the Cell-Interior and Cell-Edge:* To illustrate the performance at cell-interior and cell-edge users, we plot the achievable rates of multi-layer precoding, single-user, multi-cell MMSE, and conventional conjugate beamforming in Fig. 8. The rates are plotted versus the user distance to the BS, normalized by the cell radius $r_{cell} = 100\text{m}$. The figure confirms the asymptotic optimal performance of multi-layer precoding at the cell-interior, given in Proposition 2. At the cell edge, users experience some blockage that can be fixed with the augmented vertical dimension modification in Section VI-C. Compared to the multi-cell MMSE and conventional conjugate beamforming, multi-layer precoding with augmented vertical dimensions has better performance, especially at the cell center users.

VIII. 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.

APPENDIX A

Proof of Lemma 1: To prove the achievable rate in (18), it is sufficient to prove that the power normalization factor $[\mathbf{Y}]_{k,k}$ 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 $[\mathbf{Y}]_{k,k} = \sqrt{\left(\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} \right)^{-1}}$. Using this values of $[\mathbf{Y}]_{k,k}$, the multi-layer precoding power constraint can be written as

$$\left\| \left[\mathbf{F}_c^{(1)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} \right]_{:,k} \right\|^2 = [\mathbf{Y}]_{:,k}^* \mathbf{F}_c^{(3)*} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(1)*} \mathbf{F}_c^{(1)} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} [\mathbf{Y}]_{:,k} \quad (35)$$

$$\stackrel{(a)}{=} [\mathbf{Y}]_{:,k}^* \mathbf{F}_c^{(3)*} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(3)} [\mathbf{Y}]_{:,k} \quad (36)$$

$$= [\mathbf{Y}]_{:,k}^* \left(\overline{\mathbf{H}}_c^* \overline{\mathbf{H}}_c \right)^{-1} \overline{\mathbf{H}}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \overline{\mathbf{H}}_c \left(\overline{\mathbf{H}}_c^* \overline{\mathbf{H}}_c \right)^{-1} [\mathbf{Y}]_{:,k} \quad (37)$$

$$\stackrel{(b)}{=} [\mathbf{Y}]_{:,k}^* \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} \times \mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \times \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} [\mathbf{Y}]_{:,k} \quad (38)$$

$$= [\mathbf{Y}]_{:,k}^* \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} \times \mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \times \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} \mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \times \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} [\mathbf{Y}]_{:,k} \quad (39)$$

$$= [\mathbf{Y}]_{:,k}^* \left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)^{-1} [\mathbf{Y}]_{:,k} \quad (40)$$

$$\stackrel{(c)}{=} 1, \quad (41)$$

where (a) follows by noting that $\mathbf{F}_c^{(1)}$ has a semi-unitary structure. The effective channel matrix $\bar{\mathbf{H}}_c = [\bar{h}_{c1}, \dots, \bar{h}_{cK}]$ with $\bar{h}_{ck} = [\mathbf{G}_{c,(k,1)}, \dots, \mathbf{G}_{c,(k,K)}]^* \bar{\mathbf{w}}_{cck}$, $k = 1, \dots, K$ can also be written as $\bar{\mathbf{H}}_c = \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c$ with $\mathbf{W}_c = \mathbf{I}_K \circ [\bar{\mathbf{w}}_{cc1}, \dots, \bar{\mathbf{w}}_{ccK}]$, which leads to (b). Finally, (c) follows by substituting with $[\mathbf{Y}]_{k,k} = \sqrt{\left(\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)_{k,k}^{-1} \right)^{-1}}$. ■

APPENDIX B

Proof of Proposition 2: Considering the system and channel models in Section II and applying the multi-layer precoding algorithm in Section IV, the achievable rate by user k at cell c is given by Lemma 1

$$R_{ck} = \log_2 \left(1 + \frac{\text{SNR}}{\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)_{k,k}^{-1}} \right). \quad (42)$$

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 [43], we get $\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)_{k,k}^{-1} = \|\bar{\mathbf{w}}_{cck}\|^{-2}$. Therefore, to complete the proof, it is sufficient to prove that (i) $\lim_{N_V, N_H \rightarrow \infty} \mathbf{G}_{c,(k,m)} = \mathbf{0}$, $\forall m \neq k$ and (ii) $\lim_{N_V, N_H \rightarrow \infty} \mathbf{G}_{c,(k,k)} = \mathbf{I}$. To do that, we will first present the following useful lemma, which is a modified version of Lemma 3 in [14].

Lemma 6: Consider a user k at cell c with an azimuth angle ϕ_{ck} . Adopt the one-ring channel model in (19) with an azimuth angular spread Δ_A and correlation matrix \mathbf{R}_{cck}^A . Define the unit-norm azimuth array response vector associated with an azimuth angle ϕ_m and elevation angle θ_m as $\mathbf{u}_m = \frac{\mathbf{a}(\phi_m, \theta_m)}{\sqrt{N_H}}$, where $\mathbf{a}(\phi_m, \theta_m) = [1, \dots, e^{jkD(N_H-1)\sin(\theta_x)\sin(\phi_x)}]$. If the angle $\phi_x \notin [\phi_{ck} - \Delta_A, \phi_{ck} + \Delta_A]$, then

$$\mathbf{u}_x \in \text{Null}(\mathbf{R}_{cck}^A), \quad \text{as } N_H \rightarrow \infty. \quad (43)$$

Proof: First, note that $[\mathbf{R}_{cck}^A]_{n_1, n_2}$ in (19), can also be written as

$$\begin{aligned} & [\mathbf{R}_{cck}^A]_{n_1, n_2} \\ &= \frac{1}{2\Delta_A} \int_{-\Delta_A}^{\Delta_A} \mathbf{a}(\phi_{ck} + \alpha, \theta_{ck}) \mathbf{a}^*(\phi_{ck} + \alpha, \theta_{ck})_{n_1, n_2} d\alpha \end{aligned} \quad (44)$$

Then, we have

$$\begin{aligned} \mathbf{u}_m^* \mathbf{R} \mathbf{u}_m &= \frac{1}{2\Delta_A N_H} \int_{-\Delta_A}^{\Delta_A} \mathbf{a}^*(\phi_m, \theta_m) \mathbf{a}(\phi_{ck} + \alpha, \theta_{ck}) \\ &\quad \times \mathbf{a}^*(\phi_{ck} + \alpha, \theta_{ck}) \mathbf{a}(\phi_m, \theta_m) d\alpha \end{aligned} \quad (45)$$

$$\begin{aligned} &= \frac{1}{2\Delta_A} \int_{-\Delta_A}^{\Delta_A} \frac{1}{N_H} |\mathbf{a}^*(\phi_m, \theta_m) \\ &\quad \times \mathbf{a}(\phi_{ck} + \alpha, \theta_{ck})|^2 d\alpha. \end{aligned} \quad (46)$$

Using Lemma 1 in [44], we reach

$$\lim_{N_H \rightarrow \infty} \mathbf{u}_m^* \mathbf{R} \mathbf{u}_m = 0, \quad \forall \phi_m \notin [\phi_{ck} - \Delta_A, \phi_{ck} + \Delta_A]. \quad (47)$$

Now, to prove that $\mathbf{G}_{c,(k,m)} = \bar{\mathbf{U}}_{cck}^* \bar{\mathbf{U}}_{ccr} = \mathbf{U}_{cck}^{A*} \mathbf{U}_{ccr}^A \otimes \bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{ccr}^E = \mathbf{0}$, we need to prove that either $\mathbf{U}_{cck}^{A*} \mathbf{U}_{ccr}^A = \mathbf{0}$ or $\bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{ccr}^E = \mathbf{0}$. If $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_A$, then the columns of $\mathbf{U}_{ccm}^A \in \text{Span} \left\{ \frac{\mathbf{a}(\phi_m)}{\sqrt{N_H}} \mid \phi_m \in [\phi_{cm} - \Delta_A, \phi_{cm} + \Delta_A] \right\} \subseteq \text{Null}(\mathbf{R}_{cck}^A)$ as $N_H \rightarrow \infty$, which follows from Lemma 6. This leads to $\lim_{N_H \rightarrow \infty} \mathbf{U}_{cck}^{A*} \mathbf{U}_{ccm}^A = \mathbf{0}$. Similarly, if $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_E$, then $\lim_{N_V \rightarrow \infty} \bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{ccm}^E = \mathbf{0}$. Further, since $d \leq d_{\max}$, we have $|\theta_{ck} - \theta_l| \geq 2\Delta_E$, for any elevation angle θ_l of another cell user. This implies that $\bar{\mathbf{U}}_{cck}^E \in \text{Range}\{\bar{\mathbf{U}}_c^{Nl}\}$ as $N_V \rightarrow \infty$ by Lemma 6, and $\exists \mathbf{A}_{ck}$ such that $\bar{\mathbf{U}}_{cck}^E = \bar{\mathbf{U}}_c^{Nl} \mathbf{A}_{ck}$. For the \mathbf{U}_{ccm} , it can be generally expressed as $\mathbf{U}_{ccm} = \bar{\mathbf{U}}_c^{Nl} \mathbf{A}_{cm} + \bar{\mathbf{U}}_c^l \mathbf{B}_{cm}$ for some matrices \mathbf{A}_{cm} , \mathbf{B}_{cm} of proper dimensions. As $\lim_{N_V \rightarrow \infty} \bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{ccm}^E = \mathbf{0}$, we have $\lim_{N_V \rightarrow \infty} \mathbf{A}_{ck}^* \mathbf{A}_{cm} = \mathbf{0}$. Then, $\bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{ccm}^E = \mathbf{A}_{ck}^* \mathbf{A}_{cm} = \mathbf{0}$ as $N_V \rightarrow \infty$. This completes the proof of the first condition, $\mathbf{G}_{c,(k,m)} = \mathbf{0}$ if $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_A$ or $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_E$, $\forall m \neq k$.

To prove that $\lim_{N_V, N_H \rightarrow \infty} \mathbf{G}_{c,(k,k)} = \mathbf{I}$, we need to show that $\bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{cck}^E = \mathbf{I}$. Since $\bar{\mathbf{U}}_{cck}^E$ can be written as $\bar{\mathbf{U}}_{cck}^E = \bar{\mathbf{U}}_c^{Nl} \mathbf{A}_{ck}$ when $N_V \rightarrow \infty$, then we have $\bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{cck}^E = \mathbf{A}_{ck}^* \mathbf{A}_{ck} = \mathbf{I}$. This results in $\bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{cck}^E = \mathbf{A}_{ck}^* \mathbf{A}_{ck} = \mathbf{I}$ as $N_V \rightarrow \infty$, which completes the proof. ■

APPENDIX C

Proof of Proposition 3: Similar to the proof of Proposition 2, if $|\phi_{ck} - \phi_{cm}| \geq 2\Delta_A$ or $|\theta_{ck} - \theta_{cm}| \geq 2\Delta_E$, $\forall m \neq k$, then $\lim_{N_V, N_H \rightarrow \infty} \mathbf{G}_{c,(k,m)} = \mathbf{0}$. Using the matrix inversion lemma and leveraging the block diagonal structure of \mathbf{W}_c , we get $\left(\mathbf{W}_c^* \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{F}_c^{(2)*} \mathbf{F}_c^{(2)} \mathbf{W}_c \right)_{k,k}^{-1} = \left(\bar{\mathbf{w}}_{cck}^* \mathbf{G}_{c,(k,k)} \bar{\mathbf{w}}_{cck} \right)^{-1}$. Note that since $d > d_{\max}$, \mathbf{U}_{cck} is not guaranteed to be in $\text{Range}(\bar{\mathbf{U}}_c^{Nl})$, and $\bar{\mathbf{U}}_{cck}^{E*} \bar{\mathbf{U}}_{cck}^E \neq \mathbf{I}$ in general. The achievable rate of user k at cell c can therefore be written as

$$\lim_{N_V, N_H \rightarrow \infty} R_{ck} = \log_2 \left(1 + \text{SNR} \bar{\mathbf{w}}_{cck}^* \mathbf{G}_{c,(k,k)} \bar{\mathbf{w}}_{cck} \right) \quad (48)$$

$$\stackrel{(a)}{\geq} \log_2 \left(1 + \text{SNR} \|\bar{\mathbf{w}}_{cck}\|^2 \sigma_{\min}^2 \left(\mathbf{I} \otimes \bar{\mathbf{U}}_{cck}^E \right) \right) \quad (49)$$

$$\stackrel{(b)}{=} \log_2 \left(1 + \text{SNR} \|\bar{\mathbf{w}}_{cck}\|^2 \sigma_{\min}^2 \left(\bar{\mathbf{U}}_{cck}^E \right) \right), \quad (50)$$

where (a) follows by applying the Rayleigh-Ritz theorem [45], and (b) results from the properties of the Kronecker product. ■

REFERENCES

- [1] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Multi-layer precoding for full-dimensional massive MIMO systems," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 815–819.
- [2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

- [4] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [5] K. T. Truong and R. W. Heath, Jr., "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, 2013.
- [6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [7] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.
- [8] A. Lozano, R. W. Heath, Jr., and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.
- [9] J. Singh, S. Ponnuru, and U. Madhow, "Multi-gigabit communication: The ADC bottleneck," in *Proc. ICUBW*, Vancouver, BC, Canada, Sep. 2009, pp. 22–27.
- [10] 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 J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [11] 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 Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.
- [12] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang. (2015). "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression." [Online]. Available: <https://arxiv.org/abs/1505.03682>
- [13] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. ISIT*, Jul. 2012, pp. 1137–1141.
- [14] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.
- [15] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
- [16] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [17] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [18] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *Proc. GLOBECOM*, Dec. 2014, pp. 4066–4071.
- [19] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.
- [20] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [21] Y.-H. Nam *et al.*, "Full-dimension MIMO (FD-MIMO) for next generation cellular technology," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 172–179, Jun. 2013.
- [22] N. Seifi, J. Zhang, R. W. Heath, Jr., T. Svensson, and M. Coldrey, "Coordinated 3D beamforming for interference management in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5396–5410, Oct. 2014.
- [23] A. Kammoun, H. Khanfir, Z. Altman, M. Debbah, and M. Kamoun, "Preliminary results on 3D channel modeling: From theory to standardization," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1219–1229, Jun. 2014.
- [24] 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. VTC*, Sep. 2013, pp. 1–5.
- [25] 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. ICC*, Jun. 2014, pp. 5865–5870.
- [26] 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 Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
- [27] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [28] S. Hur *et al.*, "Proposal on millimeter-wave channel modeling for 5G cellular system," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 454–469, Apr. 2016.
- [29] S. Haghghatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 303–318, Jan. 2017.
- [30] S. Haghghatshoar and G. Caire. (2016). "Low-complexity massive MIMO subspace estimation and tracking from low-dimensional projections." [Online]. Available: <https://arxiv.org/abs/1608.02477>
- [31] 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 Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
- [32] P. Petrus, J. H. Reed, and T. S. Rappaport, "Geometrical-based statistical macrocell channel model for mobile environments," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 495–502, Mar. 2002.
- [33] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550–560, Apr. 2002.
- [34] M. Zhang, P. J. Smith, and M. Shafi, "An extended one-ring MIMO channel model," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2759–2764, Aug. 2007.
- [35] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
- [36] J.-C. Shen, J. Zhang, E. Alsusa, and K. B. Letaief, "Compressed CSI acquisition in FDD massive MIMO: How much training is needed?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4145–4156, Jun. 2016.
- [37] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.
- [38] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [39] T. Bai, A. Alkhateeb, and R. W. Heath, Jr., "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sep. 2014.
- [40] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [41] H. Ghanch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.
- [42] R. Méndez-Rial, N. González-Prelcic, and R. W. Heath, Jr., "Adaptive hybrid precoding and combining in mmWave multiuser MIMO systems based on compressed covariance estimation," in *Proc. IEEE 6th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2015, pp. 213–216.
- [43] F. Zhang, Ed., *The Schur Complement and Its Applications*, vol. 4, Springer, 2006.
- [44] O. El Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. SPAWC*, Cesme, Turkey, Jun. 2012, pp. 100–104.
- [45] H. Lütkepohl, *Handbook of Matrices*. Hoboken, NJ, USA: Wiley, 1996.

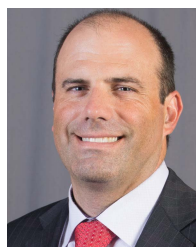


Ahmed Alkhateeb (S'08– M'17) received the B.S. degree (Hons.) and the M.S. degree from Cairo University, Egypt, in 2008 and 2012, respectively, and the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin (UT Austin), in 2016. He is currently a Wireless Communications Researcher with Facebook, CA, USA. His research interests are in the broad area of communication theory, applied math, and signal processing. In the context of wireless communication, his interests include mmWave communication and massive MIMO systems. He was a recipient of the 2012–2013 UT Austin's MCD Fellowship and the 2016 IEEE Signal Processing Society Young Author Best Paper Award.



Geert Leus received the M.Sc. and Ph.D. degrees in applied sciences from Katholieke Universiteit Leuven, Belgium, in 1996 and 2000, respectively. He currently is an Antoni van Leeuwenhoek Full Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands. His research interests are in signal processing for communications. He is a fellow of the IEEE and a fellow of the EURASIP. He received the 2002 IEEE Signal Processing Society Young Author Best Paper Award

and the 2005 IEEE Signal Processing Society Best Paper Award. He was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE SIGNAL PROCESSING LETTERS, and the *EURASIP Journal on Advances in Signal Processing*. He is currently a Member-at-Large of the Board of Governors of the IEEE Signal Processing Society and a member of the IEEE Sensor Array and Multichannel Technical Committee. He serves as the Editor in Chief of the *EURASIP Journal on Advances in Signal Processing*.



Robert W. Heath, Jr. (S'96–M'01–SM'06–F'11) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, USA, in 1996 and 1997 respectively, and the Ph.D. degree from Stanford University, Stanford, CA, USA, in 2002, all in electrical engineering. From 1998 to 2001, he was a Senior Member of the Technical Staff, then a Senior Consultant with Iospan Wireless Inc., San Jose, CA, USA, where he was involved in the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication

system. Since 2002, he has been with the Department of Electrical and Computer Engineering, The University of Texas at Austin, where he has been a Cullen Trust for Higher Education Endowed Professor, and a member of the Wireless Networking and Communications Group. He is also the President and CEO of MIMO Wireless Inc. He authored *Introduction to Wireless Digital Communication* (Prentice Hall, 2017) and *Digital Wireless Communication: Physical Layer Exploration Laboratory Using the NI USRP* (National Technology and Science Press, 2012), and co-authored *Millimeter Wave Wireless Communications* (Prentice Hall, 2014).

Dr. Heath has been a co-author of fifteen award winning conference and journal papers, including the 2010 and 2013 *EURASIP Journal on Wireless Communications and Networking* Best Paper Awards, the 2012 *Signal Processing Magazine* Best Paper Award, the 2013 Signal Processing Society Best Paper Award, the 2014 *EURASIP Journal on Advances in Signal Processing* Best Paper Award, the 2014 *Journal of Communications and Networks* Best Paper Award, the 2016 IEEE Communications Society Fred W. Ellersick Prize, the 2016 IEEE Communications and Information Theory Societies Joint Paper Award, the 2017 IEEE Marconi Prize Paper Award in Wireless Communications, and the 2017 EURASIP Technical Achievement Award. He was a Distinguished Lecturer in the IEEE Signal Processing Society and is an ISI Highly Cited Researcher. He is also an Elected Member of the Board of Governors for the IEEE Signal Processing Society, a licensed amateur radio operator, a private pilot, and a registered Professional Engineer in Texas.