This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

DoA Estimation and Capacity Analysis for 3D Millimeter Wave Massive-MIMO/FD-MIMO OFDM Systems

Rubayet Shafin, *Student Member, IEEE*, Lingjia Liu, *Senior Member, IEEE*, Jianzhong (Charlie) Zhang, *Fellow, IEEE*, and Yik-Chung Wu, *Senior Member, IEEE*

***Abstract*—With the promise of meeting future capacity demands, 3D massive-MIMO/Full Dimension MIMO (FD-MIMO) systems have gained much interest in recent years. Apart from the huge spectral efficiency gain, 3D massive-MIMO/FD-MIMO systems can also lead to significant reduction of latency, simplified multiple access layer, and robustness to interference. However, in order to completely extract the benefits of the system, accurate channel state information is critical.**

In this paper, a channel estimation method based on direction of arrival (DoA) estimation is presented for 3D millimeter wave massive-MIMO OFDM systems. To be specific, the DoA is estimated using Estimation of Signal Parameter via Rotational Invariance Technique (ESPRIT) method, and the root mean square error (RMSE) of the DoA estimation is analytically characterized for the corresponding MIMO-OFDM system. An ergodic capacity analysis of the system in the presence of DoA estimation error is also conducted, and an optimum power allocation algorithm is derived. Furthermore, it is shown that the DoA-based channel estimation achieves better performance than traditional linear minimum mean squared error estimation (LMMSE) in terms of ergodic throughput and minimum chordal distance between the subspaces of the downlink precoders obtained from the underlying channel and the estimated channel.

***Index Terms*—3D Massive MIMO OFDM Systems, Millimeter Wave Communication, Parametric Channel Estimation, 5G.**

I. Introduction

In recent years, wireless communications have experienced an unprecedented growth-rate of wireless data traffic due to the rapid introduction of various connected mobile devices and excessively data-hungry applications run on those devices. According to the International Telecommunication Union (ITU) [3], the total number of mobile cellular subscription is fast approaching the total number of people living on earth, and is expected to be more than 7 billion by the end of 2015, which corresponds to a penetration rate of 97%. Cisco Systems predicts that Global Mobile data traffic will increase nearly ten fold between 2014 and 2019 reaching 24.3 exabytes per month in 2019 [4]. In order to fulfill this huge capacity demand, multiple-input-multiple-output (MIMO) [5] systems have been introduced as the primary solution to increase the spectral-efficiency of the underlying communication system. However, the presently available MIMO transmissions, allowed by LTE-Advanced systems [6], do not support more than 8 antenna ports at the base station.

Recently, a new MIMO paradigm called massive-MIMO, also known as large-scale MIMO, has created much interest both in academia [7]–[10] and industry [11], with the promise of meeting future capacity demands by providing increased spectral-efficiency achieved through aggressive spatial multiplexing. Considering the form factor limitation at the base station (BS), instead of placing a large number of antennas horizontally, three dimensional (3D) massive-MIMO systems employ those antennas in a two dimensional (2D) antenna array enabling the exploitation of the degrees of freedom in elevation domain along with those in the azimuth domain [12]. Accordingly, 3D massive-MIMO is also called full-dimension MIMO (FD-MIMO) in 3GPP LTE-Advanced systems. Apart from the huge potential of providing excellent spatial resolution and array gains, massive-MIMO can also offer a significant reduction of latency, a simplified multiple access layer, and robustness to interference [13]. With the help of a large number of antennas, this system can concentrate more energy in a particular direction leading to a dramatic increase in energy efficiency. Furthermore, massive-MIMO is the key enabling technology for gigabit-per-second data transmission in millimeter wave (mmW) wireless communications with carrier frequency between 30 and 300 GHz. In mmW communications, it becomes feasible to pack a greater number of antennas at the base station. However, the benefits of Massive MIMO are limited by the accuracy of the channel state information (CSI) obtained at the transmitter. The CSI is critical for functionalities such as downlink beam-forming, transmit precoding, and user scheduling.

In general, there are two methods to estimate the MIMO channel. First is the traditional way where the channel transfer function is estimated. Channel state information can be obtained by sending some predefined pilot signals/reference signals and estimating the channel matrix from the received signal. Alternatively, channel estimation can be performed based on the parametric channel models where directions of arrival (DoA) and directions of departure (DoD) of the resolvable paths can be estimated [14], [15]. The time division duplex (TDD) 3D massive MIMO system can exploit the channel reciprocity in order to acquire the channel state information (CSI). To be specific, using the parametric model, the base station can estimate the DoAs in the uplink. Then with

R. Shafin and L. Liu are with the Department of Electrical Engineering and Computer Science, University of Kansas, USA. J. Zhang is with the Standards and Mobility Innovation Lab., Samsung Research America, USA. Y. Wu is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. Part of the contents in this paper has been presented in [1] and [2].

The corresponding author: L. Liu.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

the perfect calibration of the uplink-downlink signal chains, in the downlink the base station can generate the beamforming vectors based on the channel estimation (DoAs) conducted in the uplink phase.

MIMO technology, together with Orthogonal Frequency Division Multiplexing (OFDM), offers efficient ways of increasing the spectral efficiency of the cellular system. The high-data-rate wireless transmission schemes, such as OFDM, converts a frequency-selective MIMO channel into a parallel collection of frequency-flat subchannels, which is beneficial for detection and channel estimation. It is shown in [16] that unprecedented spectral efficiency and promising system throughput can be obtained by combining these two powerful technologies– MIMO and OFDM. Because of the sensitivity of MIMO algorithms with respect to the underlying channel matrix, channel state information is particularly critical in order to assess the performance of the MIMO-OFDM systems.

In general, channel matrix for a MIMO system can be modeled in different ways. The parametric channel model is adopted by performing virtual direction-of-arrival (DoA) and direction-of-departure (DoD) estimations of resolvable paths. It provides a simple geometric interpretation of the scattering environment in characterizing the two key MIMO channel metrics: ergodic capacity and diversity level [17]. Despite the advantage of reducing the number of estimation parameters, it is shown in [18] that channel estimation based on DoA and DoD provides the best performance in terms of error bound.

The contribution of the paper can be summarized as follows:

- First, based on a parametric channel model, we present a DoA estimation method for the single user 3D massive MIMO OFDM system. It has been shown that ESPRIT type algorithms can be elegantly deployed for estimating the azimuth as well as the elevation angles efficiently.
- Second, we derive an analytical expression for the root mean square error (RMSE) of the angle of arrival estimation for different channel taps. Our simplified results for the case of the 3D massive MIMO system show that if the angles are drawn independently from any continuous distribution, the RMSE depends heavily on the number of antennas, antenna orientation, number of snapshots, and correlation of the transmitted signal. Accordingly, valuable insights into pilot sequence design can be drawn for the massive MIMO OFDM systems.
- Third, we present a throughput analysis for the 3D massive MIMO OFDM system and derive an optimum downlink precoder for such a system. We show that, for the parametric channel model assumed in this work, the downlink precoder can be constituted from the DoAs estimated at the base station. Furthermore, we characterize the ergodic throughput under the DoA estimation error, and identify the optimal transmission and power allocation strategies. Our results also show that the DoA based channel estimation method achieves better performance than the traditional linear minimum mean squared estimation (LMMSE) in terms of system throughput and minimum chordal distance.
- Finally, we investigate the performance of the elevation and azimuth angle estimation under various antenna configurations and observe some interesting results. For example, for a system with the same 64 antennas, an 8×8 antenna pattern can yield better performance than a 4×16 array in azimuth angle estimation at low and medium signal-to-noise (SNR) regime. These kind of results can provide significant design intuition for developing practical massive MIMO OFDM system.

The remainder of the paper is organized as follows: the system model and an outline of the DoA estimation method is introduced in Section II. Analytical expression for the RMSE of the DoA estimation is presented in Section III followed by the capacity and achievable rate analysis in section IV. The simulation results are shown in Section V before we draw conclusion in Section VI.

II. System Model and DoA Estimation

Consider a MIMO-OFDM system with N_t transmit antennas and N_r receive antennas. At each transmitter, the high-rate information symbols to be transmitted are grouped into blocks of length N_c. The i-th such block at the j_t-th transmitter can be represented as

$$\mathbf{x}_{i,j_t} = [x_{i,j_t}(0), x_{i,j_t}(1), \ldots, x_{i,j_t}(N_c - 1)]^T, \quad (1)$$

where $x_{i,j_t}(k)$ denotes the k-th information symbol within the i-th block at the j_t-th transmitter. The IFFT operation yields the OFDM symbol consisting of the sequence $s_{i,j_t}(0), s_{i,j_t}(1), \ldots, s_{i,j_t}(N_c - 1)$, where $\{s_{i,j_t}(n)\}$ corresponds to the IFFT samples of the multi-carrier signal at the j_t-th transmitter, and can be written as

$$s_{i,j_t}(n) = \frac{1}{\sqrt{N_c}} \sum_{k=0}^{N_c-1} x_{i,j_t}(k) e^{j2\pi kn/N_c}, \quad 0 \le n \le N_c - 1. \quad (2)$$

The n-th samples at all the transmitters can be expressed as

$$\mathbf{s}_i(n) = [s_{i,j_0}(n), s_{i,j_1}(n), \ldots, s_{i,j_{(N_t-1)}}(n)]^T. \quad (3)$$

After appending CP, the resulting sequence $\tilde{s}_{i,j_t}(n)$ is first passed through a parallel-to-serial converter followed by a digital-to-analog converter, resulting in the baseband OFDM transmit signal at the j_t-th transmit antenna. The baseband signal is then upconverted and sent through a frequency selective fading channel, which is assumed to remain time-invariant during one OFDM symbol duration. We assume that the channel, which can be represented by an equivalent discrete-time linear channel impulse response (CIR), has a finite number (L) of non-zero taps.

In this paper, a clustered channel model is considered, where each scattering cluster is assumed to contribute a single propagation path. The channel impulse responses are then represented by a sequence of channel matrices, $\mathbf{C}(\ell)$ for $\ell = 0, 1, \ldots, (L-1)$. We assume there are a finite number of resolvable paths between the transmitter and receiver. For the broadband millimeter wave communication, we can consider that each channel tap corresponds to each path. The channel impulse response for the ℓ-th tap is given by

$$\mathbf{C}(\ell) = \alpha(\ell)\mathbf{e}_r(\ell)\mathbf{e}_t^H(\ell), \quad (4)$$

where $\alpha(\ell)$, $\mathbf{e}_r(\ell)$ and $\mathbf{e}_t(\ell)$ are, respectively, the channel gain, $N_r \times 1$ receive antenna array response and $N_t \times 1$

transmit antenna array response for the ℓ-th tap; $(.)^H$ denotes Hermitian transpose. It is obvious that the transmit and receive antenna array responses depend on DoD and DoA, respectively. For the transmitter equipped with a uniform linear array (ULA), the transmit antenna array response can be described using the Vandermonde structure: $\mathbf{e}_t(\ell) = \begin{bmatrix} 1 & e^{j\omega_\ell} & \dots & e^{j(N_t-1)\omega_\ell} \end{bmatrix}^T$, where $\omega_\ell = (2\pi\Delta/\lambda)\cos\Omega_\ell$, Δ is the spacing between the adjacent transmit antenna elements, Ω_ℓ is the transmit angle (DoD) for the ℓ-th tap, and λ is the carrier wavelength. The antenna array at the base station is a planar array placed in the X-Z plane, with M_1 and M_2 antenna elements in vertical and horizontal directions, respectively. Accordingly, the number of receive antenna elements at the base station is $N_r = M_1 M_2$. It is to be noted that we are focusing on ULA at the mobile station in this paper since ULA is the default mobile station antenna configuration of the FD-MIMO system in 3GPP Rel-14 LTE-Advanced Pro. However, most of our results can be readily extended for any planer arrays; i.e. most results can be generalized for uniform rectangular array (URA) or uniform circular array (UCA). The exact extension of the work to URA and UCA will be the future work of this paper.

Since the antenna elements at the base station are placed in a 2D plane, for each resolvable path, there will be an azimuth DoA and an elevation DoA. Therefore, the receive antenna array response can be expressed as $\mathbf{e}_r(\ell) = \mathbf{a}(v_\ell) \otimes \mathbf{a}(u_\ell)$, where $\otimes$ represents the Kronecker product. $\mathbf{a}(u_\ell) = \begin{bmatrix} 1 & e^{ju_\ell} & \dots & e^{j(M_1-1)u_\ell} \end{bmatrix}^T$ and $\mathbf{a}(v_\ell) = \begin{bmatrix} 1 & e^{jv_\ell} & \dots & e^{j(M_2-1)v_\ell} \end{bmatrix}^T$ can be viewed as the receive steering vectors of the elevation and azimuth angles, respectively. Here, $u_\ell = \frac{2\pi d}{\lambda}\cos\theta_\ell$ and $v_\ell = \frac{2\pi d}{\lambda}\sin\theta_\ell\cos\phi_\ell$ are the two receive spatial frequencies at the base station, d is the spacing between adjacent antenna elements in the receive antenna array, and θ_ℓ and ϕ_ℓ are the elevation and azimuth DoA, respectively. It is noteworthy here that if instead of using ULA, an URA is used at the transmitter side, the uplink transmit antenna array response can be represented as the kronecker product of steering vectors containing the azimuth and elevation DoD's, respectively.

The received base-band signal at the j_r-th receiver, $y_{i,j_r}(t)$ is passed through an analog-to-digital (A/D) converter. The cyclic prefix is then removed from the output of the A/D converter which yields the sequence $\mathbf{y}_i(n) = \mathbf{s}_i(n) \circledast \mathbf{C}(n) + \mathbf{w}_i(n)$, where $\circledast$ represents circular convolution, $\mathbf{y}_i(n) = [y_{i,j_0}(n), y_{i,j_1}(n), \dots, y_{i,j_{(N_r-1)}}(n)]^T$ is the vector containing n-th received sample at all the receiver, j_r represents the index for the receive antenna, and $\mathbf{w}_i(n)$ is the noise vector. It is to be noted here that $\mathbf{s}_i(n)$ has a sequence of length N_c, whereas, $\mathbf{C}(\ell)$ has sequence of only length L. Therefore, for facilitating the circular convolution in the DFT operation, we append $(N_c - L)$ zero matrices to the sequence of $\mathbf{C}(\ell)$, so that both the channel impulse response and transmitted IFFT sample vector are of the same sequence-length, N_c. We can arrange the transmitted time-samples of the OFDM symbol in a tall $N_t N_c \times 1$ vector, $\tilde{\mathbf{s}}_i = \begin{bmatrix} \mathbf{s}_i^T(0) & \mathbf{s}_i^T(1) & \dots & \mathbf{s}_i^T(N_c-1) \end{bmatrix}^T$. For clarity, $\tilde{\mathbf{s}}_i$ can be expressed as $\tilde{\mathbf{s}}_i = \mathbf{F}\tilde{\mathbf{x}}_i$, where $\tilde{\mathbf{x}}_i$ is the corresponding $N_t N_c \times 1$ vector containing frequency-domain transmit symbols for all the subcarriers at all the antennas, and $\mathbf{F}$ is the transformation matrix given by $\mathbf{F} = \mathcal{F}^H \otimes \mathbf{I}_{n_t}$, where $\mathcal{F}$ is a $N_c \times N_c$ DFT matrix. Accordingly, we can represent the corresponding received N_c time-samples as $\tilde{\mathbf{y}}_i = \begin{bmatrix} \mathbf{y}_i^T(0) & \mathbf{y}_i^T(1) & \dots & \mathbf{y}_i^T(N_c-1) \end{bmatrix}^T$. We can thereby express the $N_r N_c \times 1$ received sample vector as $\tilde{\mathbf{y}}_i = \mathbf{C_c}\tilde{\mathbf{s}}_i + \tilde{\mathbf{w}}_i$ where, $\tilde{\mathbf{w}}_i = \begin{bmatrix} \mathbf{w}_i^T(0) & \mathbf{w}_i^T(1) & \dots & \mathbf{w}_i^T(N_c-1) \end{bmatrix}^T$ is the $N_r N_c \times 1$ AWGN noise vector, and $\mathbf{C_c}$ is the $N_r N_c \times N_t N_c$ block-circulant matrix governing the circular convolution [1]. Now, after proper rearrangement of the transmit time sample vector, we can write the n-th received time sample vector by multiplying the transmit time-sample vector with any arbitrary row (say p-th row, where $1 \le p \le N_c$) of the matrix $\mathbf{C_c}$. To be specific, let $\tilde{\mathbf{s}}_i^n$ denote the vector $\begin{bmatrix} \mathbf{s}_i^T(N_c - p + n + 1), \dots, \mathbf{s}_i^T(N_c - p + n) \end{bmatrix}^T$. Notice that the vector $\tilde{\mathbf{s}}_i^n$ can be obtained by cyclically shifting the elements of the vector $\tilde{\mathbf{s}}_i$ defined earlier. Then, the n-th received time sample vector, $\mathbf{y}_i(n)$ can be obtained by multiplying $\tilde{\mathbf{s}}_i^n$ with the p-th row of the matrix $\mathbf{C_c}$. Therefore, the n-th received time-samples vector can be written as

$$\begin{aligned} \mathbf{y}(n) = \mathbf{C}(p-1)\mathbf{s}(N_c - p + n + 1) + \mathbf{C}(p-2) \\ \mathbf{s}(N_c - p + n + 2) + \dots + \mathbf{C}(p)\mathbf{s}(N_c - p + n) + \mathbf{w}(n) \end{aligned} \quad (5)$$

where $\mathbf{C}(\ell) = \mathbf{0}_{N_r \times N_t}$ for $L \le \ell \le (N_c - 1)$. It is to be noted here that we have dropped the index, i in (5) for notational convenience. It will be assumed from now on that all the samples correspond to i-th OFDM symbol. Recall here again that the index ℓ inside the bracket of $\mathbf{C}(\ell)$ represents the tap index, whereas l inside the bracket of $y(l), s(l)$ and $w(l)$ represents time-sample index. Using (4), and after a rearrangement, we can write (5) as

$$\mathbf{y}(n) = \mathbf{A}\mathrm{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\tilde{\mathbf{s}}^n + \mathbf{w}(n), \quad (6)$$

where $\mathbf{A} = \begin{bmatrix} \mathbf{e}_r(p-1) & \mathbf{e}_r(p-2) & \dots & \mathbf{e}_r(p) \end{bmatrix}$ can be viewed as the array steering matrix, $\tilde{\mathbf{E}}_t$ is a $N_c \times N_t N_c$ block diagonal matrix containing the transmit array responses,

$$\tilde{\mathbf{E}}_t = \mathrm{blkdiag}\{\mathbf{e}_t^H(p-1), \mathbf{e}_t^H(p-2), \dots, \mathbf{e}_t^H(p)\}, \quad (7)$$

where blkdiag$\{\mathbf{Q}\}$ represents block diagonalization of matrix $\mathbf{Q}$, and $\mathbf{b}$ is the $1 \times N_c$ vector containing the the complex fading envelopes. Now, taking $V = N_c$ snapshots of the received signal, i.e., one sample per subcarrier, we extend (6):

$$\mathbf{Y} = \mathbf{A}\mathrm{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\tilde{\mathbf{S}} + \mathbf{W}, \quad (8)$$

where $\mathbf{Y} = [\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(n), \dots, \mathbf{y}(N_c - 1)]$ is the $N_r \times V$ received signal matrix at the base-station, $\tilde{\mathbf{S}} = \begin{bmatrix} \tilde{\mathbf{s}}^0, \tilde{\mathbf{s}}^1, \dots, \tilde{\mathbf{s}}^n, \dots, \tilde{\mathbf{s}}^{N_c-1} \end{bmatrix}$ is the $N_t N_c \times V$ transmitted signal, and $\mathbf{W}$ is the $N_r \times V$ noise matrix.

We now introduce a low-complexity DoA estimation algorithm based on ESPRIT to jointly estimate the elevation and azimuth angles. We can write the received signal in (8) as:

$$\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{W}, \quad (9)$$

where $\mathbf{S} = \mathrm{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\tilde{\mathbf{S}}$ can be regarded as the equivalent transmit signal. The forward-backward averaged received signal can be expressed as:

$$\begin{aligned} \mathbf{Y}^{fba} &= \begin{bmatrix} \mathbf{Y} & \mathbf{\Pi}_{N_r}\mathbf{Y}^*\mathbf{\Pi}_V \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}\mathbf{S} & \mathbf{\Pi}_{N_r}\mathbf{A}^*\mathbf{S}^*\mathbf{\Pi}_V \end{bmatrix} + \begin{bmatrix} \mathbf{W} & \mathbf{\Pi}_{N_r}\mathbf{W}^*\mathbf{\Pi}_V \end{bmatrix}. \end{aligned} \quad (10)$$

Here, $\mathbf{\Pi}_p$ denotes the $p \times p$ exchange matrix with ones on its antidiagonal and zeros elsewhere. The dimension of $\mathbf{A}$ is $(N_r \times N_c)$, the dimension of diag$\{\mathbf{b}\}$ is $(N_c \times N_c)$, the dimension of $\tilde{\mathbf{E}}_t$ is $(N_c \times N_t N_c)$, and the dimension of $\tilde{\mathbf{S}}$ is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

$(N_tN_c \times V)$. Hence, the dimension of $\mathbf{Y}^{fba}$ is $(N_r \times 2V)$. The subspace decomposition of the signal space of the received signal through singular value decomposition (SVD) then can be written as:

$$\begin{bmatrix}\mathbf{AS} & \mathbf{\Pi}_{N_r}\mathbf{A}^*\mathbf{S}^*\mathbf{\Pi}_V\end{bmatrix} = \begin{bmatrix}\mathbf{U}_s & \mathbf{U}_n\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_s & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_s^H\\ \mathbf{V}_n^H\end{bmatrix}. \quad (11)$$

Following our line of work [19], we can apply ESPRIT based techniques on the received signal, and have the following shift-invariance relations:

$$\mathbf{K}_{x1}\mathbf{U}_s\mathbf{\Psi}_x = \mathbf{K}_{x2}\mathbf{U}_s \qquad \mathbf{K}_{y1}\mathbf{U}_s\mathbf{\Psi}_y = \mathbf{K}_{y2}\mathbf{U}_s, \quad (12)$$

where $\mathbf{\Psi}_x \triangleq \mathbf{T}^{-1}\mathbf{\Omega}_x\mathbf{T}$, $\mathbf{\Psi}_y \triangleq \mathbf{T}^{-1}\mathbf{\Omega}_y\mathbf{T}$, $\mathbf{T}$ is the non-singular transformation matrix; $\mathbf{K}_{x1}$, $\mathbf{K}_{x2}$, $\mathbf{K}_{y1}$, and $\mathbf{K}_{y2}$ are the selection matrices, and $\mathbf{\Omega}_x$ and $\mathbf{\Omega}_x$ are given by

$$\mathbf{\Omega}_x \triangleq \text{diag}\left\{\tan\left(\frac{u_n}{2}\right), \tan\left(\frac{u_{n-1}}{2}\right), \ldots, \tan\left(\frac{u_{n-N_c+1}}{2}\right)\right\}. \quad (13)$$

$$\mathbf{\Omega}_y \triangleq \text{diag}\left\{\tan\left(\frac{v_n}{2}\right), \tan\left(\frac{v_{n-1}}{2}\right), \ldots, \tan\left(\frac{v_{n-N_c+1}}{2}\right)\right\}. \quad (14)$$

From (12), we can solve for $\hat{\mathbf{\Psi}}_x$ and $\hat{\mathbf{\Psi}}_y$ based on estimated signal subspace using least square type of methods. Let the eigenvalues of the $N_c \times N_c$ complex matrix $\hat{\mathbf{\Psi}}_x + j\hat{\mathbf{\Psi}}_y$ be $\hat{\lambda}_\ell$, for $\ell = 1, 2, \ldots, N_c$. Then u_ℓ and v_ℓ can be calculated from:

$$\hat{u}_\ell = 2\tan^{-1}\left\{\text{Re}\left(\hat{\lambda}_\ell\right)\right\} \quad \hat{v}_\ell = 2\tan^{-1}\left\{\text{Im}\left(\hat{\lambda}_\ell\right)\right\}. \quad (15)$$

Accordingly, 2D DoAs of interest are obtained through simple parameter transformation.

III. RMSE Characterization

In this section, we present the theoretical analysis of root mean square error (RMSE) for DoA estimation using the standard ESPRIT method. Let $\hat{v}_\ell$ denote the estimated spatial frequency for ℓ-th tap; the estimation error is then given by $\Delta v_\ell = v_\ell - \hat{v}_\ell$. Similarly, $\Delta u_\ell = u_\ell - \hat{u}_\ell$. The first order approximation of the mean square estimation error of v_ℓ for the standard ESPRIT is given by [20]:

$$E\{(\Delta v_\ell)^2\} = \frac{1}{2}\left({\mathbf{r}_\ell^{(v)}}^H\mathbf{W}_{mat}^*\mathbf{R}_{nn}^T\mathbf{W}_{mat}^T\mathbf{r}_\ell^{(v)} - Re\left\{{\mathbf{r}_\ell^{(v)}}^T\mathbf{W}_{mat}\mathbf{C}_{nn}\mathbf{W}_{mat}^T\mathbf{r}_\ell^{(v)}\right\}\right), \quad (16)$$

where

$$\mathbf{r}_\ell^{(v)} = \mathbf{q}_\ell \otimes \left(\left[(\tilde{\mathbf{J}}_1^{(v)}\mathbf{U}_s)^+(\tilde{\mathbf{J}}_2^{(v)}/e^{jv_\ell} - \tilde{\mathbf{J}}_1^{(v)})\right]^T\mathbf{p}_\ell\right), \quad (17)$$

$$\mathbf{W}_{mat} = (\mathbf{\Sigma}_s^{-1}\mathbf{V}_s^T) \otimes (\mathbf{U}_n\mathbf{U}_n^H). \quad (18)$$

Here, $\tilde{\mathbf{J}}_1^{(v)}$ and $\tilde{\mathbf{J}}_2^{(v)}$ are the two effective selection matrices for the first and second subarrays, respectively, for the spatial frequency v_ℓ, $\mathbf{T}$ is the transformation matrix as described in Section II, $\mathbf{q}_\ell$ is the ℓ-th column of matrix $\mathbf{T}$, $\mathbf{p}_\ell^T$ is the ℓ-th row of matrix $\mathbf{T}^{-1}$; $\mathbf{R}_{nn}$ and $\mathbf{C}_{nn}$ are the covariance and complementary covariance matrices of the noise, respectively. Similarly, we have

$$E\{(\Delta u_\ell)^2\} = \frac{1}{2}\left({\mathbf{r}_\ell^{(u)}}^H\mathbf{W}_{mat}^*\mathbf{R}_{nn}^T\mathbf{W}_{mat}^T\mathbf{r}_\ell^{(u)} - Re\left\{{\mathbf{r}_\ell^{(u)}}^T\mathbf{W}_{mat}\mathbf{C}_{nn}\mathbf{W}_{mat}^T\mathbf{r}_\ell^{(u)}\right\}\right), \quad (19)$$

where

$$\mathbf{r}_\ell^{(u)} = \mathbf{q}_\ell \otimes \left(\left[(\tilde{\mathbf{J}}_1^{(u)}\mathbf{U}_s)^+(\tilde{\mathbf{J}}_2^{(u)}/e^{ju_\ell} - \tilde{\mathbf{J}}_1^{(u)})\right]^T\mathbf{p}_\ell\right). \quad (20)$$

Here, $\tilde{\mathbf{J}}_1^{(u)}$ and $\tilde{\mathbf{J}}_2^{(u)}$ are the two effective selection matrices for the first and second subarrays, respectively, for the spatial frequency u_ℓ. At this point, in order to facilitate the derivation of MSE expression, we consider the following Lemma:

Lemma 1. *If the elevation and azimuth angles are both drawn independently from a continuous distribution, the normalized array response vectors become orthogonal asymptotically, that is,* $\bar{\mathbf{e}}_r(k) \perp$ ***span*** $\{\bar{\mathbf{e}}_r(\ell) \mid \forall k \neq \ell\}$ *when the number of antennas at the base station goes large.*

Proof. See Appendix A. □

It is to be emphasized here that Lemma-1 holds for any continuous distribution. It can be seen that (16) depends on the singular value decomposition (SVD) of the noiseless received signal, which is not easy to obtain at the base station. In fact, it is extremely difficult to simplify such complicated result in the multiple path case. Fortunately, in the massive MIMO system, it can be significantly simplified because of the orthogonality of the steering vectors. The simplified result is only related to the real system parameters such as the number of antennas, number of snapshots, transmit power, and covariance matrix of transmit signal. Specifically, for the massive MIMO system, we have the following theorem:

Theorem 1. *For the case of 3D DoA estimation based on uniform rectangular array of $M_1 \times M_2$ elements, the root mean square errors of the elevation and azimuth angle estimations are given, respectively, by:*

$$\textit{RMSE}\{\theta_\ell\} = \frac{\sigma}{\pi\sin(\theta_\ell)(M_1-1)}\sqrt{\frac{\mathbf{R}_{SS}^{-1}(\ell,\ell)}{VM_2}} \quad (21)$$

$$\textit{RMSE}\{\phi_\ell\} = \frac{\sigma}{\pi\sin(\theta_\ell)}\sqrt{\frac{\mathbf{R}_{SS}^{-1}(\ell,\ell)}{V}}\sqrt{\left(\frac{\cot^2(\theta_\ell)\cot^2(\phi_\ell)}{(M_1-1)^2M_2} + \frac{1}{\sin^2(\phi_\ell)(M_2-1)^2M_1}\right)} \quad (22)$$

where σ^2 is the noise variance, and $\mathbf{R}_{SS}(\ell,\ell)$ is the ℓ-th diagonal element of the covariance matrix of the equivalent transmit signal.

Proof. See Appendix B. □

Now, $\mathbf{R}_{SS} = \mathbb{E}\{\mathbf{SS}^H\}/V$, where the expectation is with respect to the time-samples over different subcarriers on transmit antennas. $\mathbf{R}_{SS}$ can be expressed as:

$$\begin{aligned}\mathbf{R}_{SS} &= \text{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\mathbb{E}\{\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H\}\tilde{\mathbf{E}}_t^H(\text{diag}\{\mathbf{b}\})^H/V\\ &= \text{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\mathbf{F}\mathbb{E}\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H\}\mathbf{F}^H\tilde{\mathbf{E}}_t^H(\text{diag}\{\mathbf{b}\})^H/V \end{aligned} \quad (23)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

where $\tilde{\mathbf{X}} = \mathbf{F}^{-1}\tilde{\mathbf{S}}$ is the frequency domain transmit signal matrix. If we assume, $\mathbf{Z} = \text{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\mathbf{F}$, we can write, $\mathbf{R}_{SS} = \mathbf{Z}\mathbb{E}\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H\}\mathbf{Z}^H/V = \mathbf{Z}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}\mathbf{Z}^H/V$, where $\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}} = \mathbb{E}\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H\}$. Therefore, we can write, $\mathbf{R}_{SS}^{-1} = (V)\mathbf{Z}^{-H}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}\mathbf{Z}^{-1}$. Plugging the expression for $\mathbf{R}_{SS}^{-1}$ in (21) and (22), we can obtain:

$$\text{RMSE}\{\theta_\ell\} = \frac{\sigma}{\pi\sin(\theta_\ell)(M_1-1)}\sqrt{\frac{(\mathbf{Z}^{-H}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}\mathbf{Z}^{-1})(\ell,\ell)}{M_2}} \tag{24}$$

$$\text{RMSE}\{\phi_\ell\} = \frac{\sigma}{\pi\sin(\theta_\ell)}\sqrt{(\mathbf{Z}^{-H}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}\mathbf{Z}^{-1})(\ell,\ell)} \sqrt{\left(\frac{\cot^2(\theta_\ell)\cot^2(\phi_\ell)}{(M_1-1)^2M_2} + \frac{1}{\sin^2(\phi_\ell)(M_2-1)^2M_1}\right)} \tag{25}$$

where $(\mathbf{Z}^{-H}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}\mathbf{Z}^{-1})(\ell,\ell)$ represents the ℓ-th diagonal element of the matrix product $\mathbf{Z}^{-H}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}\mathbf{Z}^{-1}$. Based on the proof of **Theorem 1**, it is straight forward to obtain the RMSE of the spatial frequencies u_ℓ and v_ℓ as follows:

Corollary 1.1. *In massive MIMO system, RMSEs of the spatial frequencies u_ℓ and v_ℓ using standard ESPRIT are given by:*

$$RMSE\{u_\ell\} = \frac{\sigma}{(M_1-1)}\sqrt{\frac{(\mathbf{Z}^{-H}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}\mathbf{Z}^{-1})(\ell,\ell)}{M_2}}, \tag{26}$$

$$RMSE\{v_\ell\} = \frac{\sigma}{(M_2-1)}\sqrt{\frac{(\mathbf{Z}^{-H}\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}^{-1}\mathbf{Z}^{-1})(\ell,\ell)}{M_1}}. \tag{27}$$

It is clear from these equations that correlation between the data on different subcarriers will have adverse effect on the performance of the DoA estimation, i.e., root mean squared error increases as the correlation between the data increases. The pilot sequence design will also have a significant impact on the DoA estimation performance through the term $\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}$.

IV. Capacity Analysis

In this section, we investigate the downlink channel capacity of 3D massive MIMO OFDM system. First, we will analyze the system capacity assuming perfect DoA estimation. Then we will look into the system achievable rate under the presence of DoA estimation errors.

A. Ergodic Capacity with Perfect DoA Estimation:

The $N_r \times N_t$ uplink channel transfer function at the k-th subcarrier can be written as

$$\mathbf{H}(k) = \sum_{\ell=0}^{N_c-1}\mathbf{C}(\ell)e^{\frac{-j2\pi k\ell}{N_c}}. \tag{28}$$

Again, it is assumed there are N_c channel taps, but only L of them are non-zero. Using the expression of $\mathbf{C}(\ell)$ from (4),

$$\begin{aligned}\mathbf{H}(k) &= \sum_{\ell=0}^{N_c-1}\alpha(\ell)\mathbf{e}_r(\ell)\mathbf{e}_t^H(\ell)e^{\frac{-j2\pi k\ell}{N_c}} \\ &= \sum_{\ell=0}^{N_c-1}\alpha(\ell)\mathbf{e}_r(\ell,k)\mathbf{e}_t^H(\ell),\end{aligned} \tag{29}$$

where $\mathbf{e}_r(\ell,k) = e^{\frac{-j2\pi k\ell}{N_c}}\mathbf{e}_r(\ell)$. In a compact form:

$$\mathbf{H}(k) = \mathbf{A}(k)\mathbf{D}\mathbf{B}^H, \tag{30}$$

where $\mathbf{A}(k) = [\mathbf{e}_r(0,k), \mathbf{e}_r(1,k), \ldots, \mathbf{e}_r(N_c-1,k)]$, $\mathbf{D} = \text{diag}\{\mathbf{b}\} = \text{diag}\{\alpha(0), \alpha(1), \ldots, \alpha(N_c-1)\}$, and $\mathbf{B} = [\mathbf{e}_t(0), \mathbf{e}_t(1), \ldots, \mathbf{e}_t(N_c-1)]$, where $1/\sqrt{N_t}\mathbf{B}$ is a unitary matrix. Accordingly, using the channel reciprocity property, downlink channel at the k-th subcarrier can be represented as

$$\mathbf{H}^{dl}(k) = [\mathbf{H}(k)]^T = \mathbf{B}^*\mathbf{D}\mathbf{A}^T(k). \tag{31}$$

So, the $N_t \times 1$ downlink received signal at the k-th subcarrier can be written as

$$\begin{aligned}\mathbf{y}^{dl}(k) &= \mathbf{H}^{dl}(k)\mathbf{x}^{dl}(k) + \mathbf{n}^{dl}(k) \\ &= \mathbf{B}^*\mathbf{D}\mathbf{A}^T(k)\mathbf{x}^{dl}(k) + \mathbf{n}^{dl}(k),\end{aligned} \tag{32}$$

where $\mathbf{x}^{dl}(k)$ is the $N_r \times 1$ downlink transmitted signal, and $\mathbf{n}^{dl}(k)$ is the $N_t \times 1$ noise vector with $E\{\mathbf{n}^{dl}(i){\mathbf{n}^{dl}}^H(j)\} = \sigma^2\mathbf{I}_{N_t}\delta(i-j)$, and $E\{\cdot\}$ represents expectation. The mutual information at the k-th subcarrier:

$$\mathcal{I}_k = \log_2\det\left[\mathbf{I}_{N_t} + \frac{\mathbf{H}^{dl}(k)\mathbf{Q}_k{\mathbf{H}^{dl}}^H(k)}{\sigma^2}\right], \tag{33}$$

where $\mathbf{Q}_k = E\{\mathbf{x}^{dl}(k){\mathbf{x}^{dl}}^H(k)\}$ is the $N_r \times N_r$ covariance matrix of the downlink transmit signal at the k-th subcarrier. Now, the ergodic capacity of the MIMO OFDM system can be expressed as:

$$\begin{aligned}\mathcal{C} &= E\left\{\frac{1}{N_c}\sum_{k=0}^{N_c-1}\mathcal{I}_k\right\} \\ &= E\left\{\frac{1}{N_c}\sum_{k=0}^{N_c-1}\log_2\det\left[\mathbf{I}_{N_t} + \frac{\mathbf{H}^{dl}(k)\mathbf{Q}_k{\mathbf{H}^{dl}}^H(k)}{\sigma^2}\right]\right\},\end{aligned} \tag{34}$$

with the total power constraint, $Tr(\mathbf{Q}) \leq P$, where P is the total transmit power, and $\mathbf{Q}$ is the covariance matrix of the $N_rN_c \times 1$ downlink Gaussian input vector, and given by:

$$\mathbf{Q} = \text{diag}\{\mathbf{Q}_k\}_{k=0}^{N_c-1}. \tag{35}$$

The design of the taransmit covariance matrix, $\mathbf{Q}_k$, and hence, the design of the transmit signal will dictate the power allocation across the transmit antennas for the k-th tone. It is to be noted here that in the capacity expression, we have ignored the loss in spectral efficiency due to the presence of the cyclic prefix. Let us now consider the following Lemma:

Lemma 2. *For a uniform rectangular array (URA) with azimuth and elevation DoAs drawn independently from a continuous distribution, the normalized frequency-domain array response vectors becomes orthogonal asympototically, that is, $\bar{\mathbf{e}}_r(i,k) \perp$ **span**$\{\bar{\mathbf{e}}_r(j,k)|\forall i \neq j\}$, when the number of base-station antennas, $N_r = M_1M_2$ goes large, where $\bar{\mathbf{e}}_r(l,k) = \frac{1}{\sqrt{N_r}}\mathbf{e}_r(l,k)$.*

Proof. See Appendix C. □

From **Lemma 2**, the matrix $\sqrt{1/N_r}\mathbf{A}(k)$ can be regarded as unitary for the large antenna system. Therefore, downlink channel representation in (31) can essentially be thought of as singular value decomposition for the massive MIMO system. In order to simplify the analysis, we can assume that the mobile station has perfect knowledge about the DoD so that optimal received processing can be used at the mobile station.

Observing (32), the optimum precoding matrix, therefore, can be expressed as

$$\begin{aligned}\mathbf{V}^{opt}(k) &= \frac{1}{N_r}\mathbf{A}^*(k)\\ &= \frac{1}{N_r}\left[\mathbf{e}_r^*(0,k), \mathbf{e}_r^*(1,k), \ldots, \mathbf{e}_r^*(N_c-1,k)\right]\\ &= \frac{1}{N_r}\left[e^{\frac{j2\pi k(0)}{N_c}}\mathbf{e}_r^*(0), e^{\frac{j2\pi k(1)}{N_c}}\mathbf{e}_r^*(1), \ldots, e^{\frac{j2\pi k(N_c-1)}{N_c}}\mathbf{e}_r^*(N_c-1)\right]\\ &= \frac{1}{N_r}\left[e^{\frac{j2\pi k(0)}{N_c}}\mathbf{a}^*(v_0)\otimes\mathbf{a}^*(u_0), e^{\frac{j2\pi k(1)}{N_c}}\mathbf{a}^*(v_1)\otimes\mathbf{a}^*(u_1),\right.\\ &\qquad\left.\ldots, e^{\frac{j2\pi k(N_c-1)}{N_c}}\mathbf{a}^*(v_{N_c-1})\otimes\mathbf{a}^*(u_{N_c-1})\right]. \end{aligned} \tag{36}$$

From (36), it is apparent that the optimum precoding matrix is essentially composed of the array steering vectors, which again depend on the DoA estimation at the base station. Let $\mathbf{s}^{dl}(k)$ be the $N_c \times 1$ information signal vector for the k-th subcarrier. Therefore, using the optimum precoding matrix, we have $\mathbf{x}^{dl}(k) = (1/N_r)\mathbf{A}^*(k)\mathbf{s}^{dl}(k)$. Plugging this into (33) , for the perfect DoA estimation case, we can obtain (37). It is not difficult to show that $\mathbf{D}^*\mathbf{B}^T\mathbf{B}^*\mathbf{D} = N_t\text{diag}[|\alpha(0)|^2, |\alpha(1)|^2, \ldots, |\alpha(N_c-1)|^2]$. Using **Lemma 2** , we have $\mathbf{A}^T(k)\mathbf{A}^*(k)/N_r = \mathbf{I}_{N_c}$. Assuming Gaussian input information signal, $\mathbf{s}^{dl}(k)\mathbf{s}^{dl^H}(k)$ will also be diagonal, i.e., $\mathbf{s}^{dl}(k)\mathbf{s}^{dl^H}(k) = \text{diag}[|s_0^{dl}(k)|^2, \ldots, |s_{N_c-1}^{dl}(k)|^2]$ $= \text{diag}[p_0(k), \ldots, p_{N_c-1}(k)]$, where $s_\ell^{dl}(k)$ is the ℓ-th information symbol on the k-th subcarrier, and $p_\ell(k) = |s_\ell^{dl}(k)|^2$ is the power to be allocated on the corresponding channel. Finally, using Hadamard inequality, we can write (37) as

$$\begin{aligned}\mathcal{I}_k &= \log_2\prod_\ell\left(1 + \frac{N_t|\alpha(\ell)|^2 p_\ell(k)}{\sigma^2}\right)\\ &= \sum_{\ell=0}^{N_c-1}\log_2\left(1 + \gamma_\ell p_\ell(k)\right),\end{aligned} \tag{38}$$

where $\gamma_\ell = N_t|\alpha(\ell)|^2/\sigma^2$. It is to be noted here that the matrix inside the $\log_2\det$ is diagonal, and hence Hadamard inequality follows with equality. Accordingly, the optimal power allocation is the well-known water-filling solution which can be expressed as $p_\ell(k) = [\mu_\ell(k) - 1/\gamma_\ell]^\diamondsuit$, where $[x]^\diamondsuit$ is the function with $[x]^\diamondsuit = 0$ when $x < 0$, and $[x]^\diamondsuit = x$ when $x \geq 0$, and $\mu_\ell(k)$ is the corresponding Lagrange multiplier.

We now briefly discuss about the optimum receive processing for the perfect DoA estimation scenario. From (32), the $N_t \times 1$ downlink received signal can be expressed as

$$\begin{aligned}\mathbf{y}^{dl}(k) &= \mathbf{B}^*\mathbf{D}\mathbf{A}^T(k)\mathbf{V}^{opt}(k)\mathbf{s}^{dl}(k) + \mathbf{n}^{dl}(k)\\ &= \frac{1}{N_r}\mathbf{B}^*\mathbf{D}\mathbf{A}^T(k)\mathbf{A}^*(k)\mathbf{s}^{dl}(k) + \mathbf{n}^{dl}(k)\\ &= \mathbf{B}^*\mathbf{D}\mathbf{s}^{dl}(k) + \mathbf{n}^{dl}(k)\\ &= [\mathbf{e}_t^*(0), \mathbf{e}_t^*(1), \ldots, \mathbf{e}_t^*(N_c-1)]\\ &\quad \text{diag}\{\alpha(0), \alpha(1), \ldots, \alpha(N_c-1)\}\mathbf{s}^{dl}(k) + \mathbf{n}^{dl}(k).\end{aligned} \tag{39}$$

Let the received signal vector corresponding to i-th information symbol on the k-th subcarrier be represented as $\mathbf{y}_i^{dl}(k)$, and the corresponding noise vector as $\mathbf{n}_i^{dl}(k)$. Then,

$$\mathbf{y}_i^{dl}(k) = \mathbf{e}_t^*(i)\alpha(i)s_i^{dl}(k) + \mathbf{n}_i^{dl}(k) = \mathbf{g}_i^{'}s_i^{dl}(k) + \mathbf{n}_i^{dl}(k), \tag{40}$$

where $\mathbf{g}_i^{'} = \mathbf{e}_t^*(i)\alpha(i)$. Accordingly, for the receive processing, the maximum ratio combining (MRC) vector, $\mathbf{w}_i^{'}$ can be written as $\mathbf{w}_i^{'} = \mathbf{g}_i^{'}/|\mathbf{g}_i^{'}|$. Multiplying both sides of (40) by $\mathbf{w}_i^{'H}$, we obtain the MRC detected signal $\tilde{y}_i^{dl}(k)$ corresponding to the i-th information symbol:

$$\begin{aligned}\tilde{y}_i^{dl}(k) &= \mathbf{w}_i^{'H}\mathbf{y}_i^{dl}(k) = \frac{\mathbf{g}_i^{'H}}{|\mathbf{g}_i^{'}|}\mathbf{g}_i^{'}s_i^{dl}(k) + \tilde{n}_i^{dl}(k)\\ &= |\mathbf{g}_i^{'}|s_i^{dl}(k) + \tilde{n}_i^{dl}(k) = |\alpha(i)||\mathbf{e}_t^*(i)|s_i^{dl}(k) + \tilde{n}_i^{dl}(k)\\ &= \sqrt{N_t}|\alpha(i)|s_i^{dl}(k) + \tilde{n}_i^{dl}(k),\end{aligned} \tag{41}$$

where $\tilde{n}_i^{dl}(k) = \mathbf{w}_i^{'H}\mathbf{n}_i^{dl}(k) = \frac{\mathbf{g}_i^{'H}}{|\mathbf{g}_i^{'}|}\mathbf{n}_i^{dl}(k)$ is the noise element corresponding to i-th information symbol on k-th subcarrier.

B. System Achievable Rate under the presence of DoA Estimation Errors:

In this section, we study the effects of DoA estimation error on the achievable rate of the massive-MIMO OFDM system. Since base station does not have perfect DoA estimation, the precoding vectors, in the presence of DoA estimation error, will be in the form $(1/N_r)\hat{\mathbf{e}}_r(\ell,k)$, where $\hat{\mathbf{e}}_r(\ell,k) = e^{\frac{-j2\pi k(\ell)}{N_c}}\mathbf{a}(v_\ell + \Delta v_\ell)\otimes\mathbf{a}(u_\ell + \Delta u_\ell)$, and Δu_ℓ and Δv_ℓ represent the DoA estimation errors for the ℓ-th tap. For this imperfect channel estimation case, let $\hat{\mathbf{s}}^{dl}(k)$ denote the $N_c \times 1$ information symbol vector for the k-th subcarrier. The optimum precoding matrix, in the presence of DoA estimation error, can be written as $\hat{\mathbf{V}}^{opt}(k) = (1/N_r)\hat{\mathbf{A}}^*(k)$, where $\hat{\mathbf{A}}^*(k) = [\hat{\mathbf{e}}_r(0,k), \hat{\mathbf{e}}_r(1,k), \ldots, \hat{\mathbf{e}}_r(N_c-1,k)]$. Accordingly, similar to (33), the expected mutual information or achievable rate of the k-th subcarrier, $\hat{\mathcal{I}}_k$, can be represented as (42), where the expectation is with respect to the DoA estimation error. Here, we are treating the terms inside the expectation as random variable, where, for each channel realization, the DoA estimation error will be random. Hence, from (42), we can have (43), where the term $\mathbf{A}^T(k)\hat{\mathbf{A}}^*(k)$ can be written as

$$\begin{aligned}&\mathbf{A}^T(k)\hat{\mathbf{A}}^*(k)\\ &= \begin{bmatrix}\mathbf{e}_r^T(0,k)\hat{\mathbf{e}}_r^*(0,k) & \ldots & \mathbf{e}_r^T(0,k)\hat{\mathbf{e}}_r^*(N_c-1,k)\\ \mathbf{e}_r^T(1,k)\hat{\mathbf{e}}_r^*(0,k) & \ldots & \mathbf{e}_r^T(1,k)\hat{\mathbf{e}}_r^*(N_c-1,k)\\ \vdots & \ddots & \vdots\\ \mathbf{e}_r^T(N_c-1,k)\hat{\mathbf{e}}_r^*(0,k) & \ldots & \mathbf{e}_r^T(N_c-1,k)\hat{\mathbf{e}}_r^*(N_c-1,k)\end{bmatrix}.\end{aligned} \tag{44}$$

Now, let us consider the following lemma:

Lemma 3. *For the 3D massive MIMO OFDM systems, the normalized vectors $\bar{\mathbf{e}}_r(i,k) = 1/\sqrt{N_r}\mathbf{e}_r(i,k)$ and $\bar{\hat{\mathbf{e}}}_r(m,k) = 1/\sqrt{N_r}\hat{\mathbf{e}}_r(m,k)$, $\forall i \neq m$ becomes orthonormal asymptotically as the number of antenna, $N_r \to \infty$.*

Proof. See Appendix C. □

Using Lemma 3, we can write $\mathbf{A}^T(k)\hat{\mathbf{A}}^*(k) = \text{diag}[\mathbf{e}_r^T(0,k)\hat{\mathbf{e}}_r^*(0,k), \mathbf{e}_r^T(1,k)\hat{\mathbf{e}}_r^*(1,k), \ldots, \mathbf{e}_r^T(N_c-1,k)\hat{\mathbf{e}}_r^*(N_c-1,k)]$. With Gaussian information signal, $\hat{\mathbf{s}}^{dl}(k)(\hat{\mathbf{s}}^{dl}(k))^H = \text{diag}[\hat{p}_0(k), \hat{p}_1(k), \ldots, \hat{p}_{N_c-1}(k)]$, where $\hat{p}_\ell(k)$ is the power to be allocated for the ℓ-th information symbol on the k-th subcarrier. We can clearly see that $\mathbf{A}^T(k)\hat{\mathbf{A}}^*(k)\hat{\mathbf{s}}^{dl}(k)(\hat{\mathbf{s}}^{dl}(k))^H\mathbf{A}^T(k)\hat{\mathbf{A}}^*(k)\mathbf{D}^*\mathbf{B}^T\mathbf{B}^*\mathbf{D}$ is a diagonal matrix with ℓ-th diagonal element being

$$\begin{aligned}\mathcal{I}_k &= \log_2 \det \left[\mathbf{I}_{N_t} + \frac{\mathbf{B}^* \mathbf{D} \mathbf{A}^T(k) \mathbf{A}^*(k) \mathbf{s}^{dl}(k) {\mathbf{s}^{dl}}^H(k) \mathbf{A}^T(k) \mathbf{A}^*(k) \mathbf{D}^* \mathbf{B}^T}{N_r^2 \sigma^2} \right] \\ &= \log_2 \det \left[\mathbf{I}_{N_c} + \frac{\mathbf{A}^T(k) \mathbf{A}^*(k) \mathbf{s}^{dl}(k) {\mathbf{s}^{dl}}^H(k) \mathbf{A}^T(k) \mathbf{A}^*(k) \mathbf{D}^* \mathbf{B}^T \mathbf{B}^* \mathbf{D}}{N_r^2 \sigma^2} \right] \\ &= \log_2 \det \left[\mathbf{I}_{N_c} + \frac{\mathbf{A}^T(k) \mathbf{A}^*(k) \mathbf{s}^{dl}(k) {\mathbf{s}^{dl}}^H(k) \mathbf{A}^T(k) \mathbf{A}^*(k) \mathbf{D}^* \mathbf{B}^T \mathbf{B}^* \mathbf{D}}{N_r^2 \sigma^2} \right]. \end{aligned} \quad (37)$$

$$\hat{\mathcal{I}}_k = E\left\{ \log_2 \det \left[\mathbf{I}_{N_t} + \frac{\mathbf{B}^* \mathbf{D} \mathbf{A}^T(k) \hat{\mathbf{V}}^{opt}(k) \hat{\mathbf{s}}^{dl}(k) (\hat{\mathbf{s}}^{dl}(k))^H {\hat{\mathbf{V}}^{opt}}^H(k) \mathbf{A}^*(k) \mathbf{D}^* \mathbf{B}^T}{\sigma^2} \right] \right\}. \quad (42)$$

$$\begin{aligned}\hat{\mathcal{I}}_k &= E\left\{ \log_2 \det \left[\mathbf{I}_{N_t} + \frac{\mathbf{B}^* \mathbf{D} \mathbf{A}^T(k) \hat{\mathbf{A}}^*(k) \hat{\mathbf{s}}^{dl}(k) (\hat{\mathbf{s}}^{dl}(k))^H \hat{\mathbf{A}}^T(k) \mathbf{A}^*(k) \mathbf{D}^* \mathbf{B}^T}{N_r^2 \sigma^2} \right] \right\} \\ &= E\left\{ \log_2 \det \left[\mathbf{I}_{N_c} + \frac{\mathbf{A}^T(k) \hat{\mathbf{A}}^*(k) \hat{\mathbf{s}}^{dl}(k) (\hat{\mathbf{s}}^{dl}(k))^H \hat{\mathbf{A}}^T(k) \mathbf{A}^*(k) \mathbf{D}^* \mathbf{B}^T \mathbf{B}^* \mathbf{D}}{N_r^2 \sigma^2} \right] \right\}. \end{aligned} \quad (43)$$

$N_t |\mathbf{e}_r^T(\ell,k) \hat{\mathbf{e}}_r^*(\ell,k)|^2 |\alpha(\ell,k)|^2 \hat{p}_\ell(k)$. Now, invoking Hadamard's Inequality, we can write (43) as

$$\begin{aligned}\hat{\mathcal{I}}_k &= E\left\{ \log_2 \prod_\ell \left(1 + \frac{N_t |\mathbf{e}_r^T(\ell,k) \hat{\mathbf{e}}_r^*(\ell,k)|^2 |\alpha(\ell)|^2 \hat{p}_\ell(k)}{N_r^2 \sigma^2} \right) \right\} \\ &= E\left\{ \sum_{\ell=0}^{N_c-1} \log_2 \left(1 + \frac{N_t |\mathbf{e}_r^T(\ell,k) \hat{\mathbf{e}}_r^*(\ell,k)|^2 |\alpha(\ell)|^2 \hat{p}_\ell(k)}{N_r^2 \sigma^2} \right) \right\} \\ &= E\left\{ \sum_{\ell=0}^{N_c-1} \log_2 \left(1 + \hat{\gamma}_\ell |\mathbf{e}_r^T(\ell,k) \hat{\mathbf{e}}_r^*(\ell,k)|^2 \hat{p}_\ell(k) \right) \right\}, \end{aligned} \quad (45)$$

where $\hat{\gamma}_\ell = N_t |\alpha(\ell)|^2 / (N_r^2 \sigma^2)$. Using method of Lagrangian multiplier, the optimal power allocation for the k-th subcarrier:

$$\hat{p}_\ell(k) = \left[\hat{\mu}_\ell(k) - \frac{1}{\hat{\gamma}_\ell |\mathbf{e}_r^T(\ell,k) \hat{\mathbf{e}}_r^*(\ell,k)|^2} \right]^{\diamond}, \quad (46)$$

where $\hat{\mu}_\ell(k)$ is the corresponding Lagrange multiplier, and $\diamond$ indicates the water-filling algorithms. The expected transmit power for the ℓ-th tap can be expressed as

$$E\{\hat{p}_\ell(k)\} = \left[\hat{\mu}_\ell(k) - \frac{1}{\hat{\gamma}_\ell E\{|\mathbf{e}_r^T(\ell,k) \hat{\mathbf{e}}_r^*(\ell,k)|^2\}} \right]^{\diamond}. \quad (47)$$

Now,

$$\begin{aligned} & E[|\mathbf{e}_r^T(\ell,k) \hat{\mathbf{e}}_r^*(\ell,k)|^2] = E[|\hat{\mathbf{e}}_r^H(\ell,k) \mathbf{e}_r(\ell,k)|^2] \\ &= E\left[|[\mathbf{a}(v_\ell + \Delta v_\ell) \otimes \mathbf{a}(u_\ell + \Delta u_\ell)]^H [\mathbf{a}(v_\ell) \otimes \mathbf{a}(u_\ell)]|^2 \right] \\ &= E\left[|[\mathbf{a}^H(v_\ell + \Delta v_\ell) \otimes \mathbf{a}^H(u_\ell + \Delta u_\ell)][\mathbf{a}(v_\ell) \otimes \mathbf{a}(u_\ell)]|^2 \right] \\ &= E\left[|[\mathbf{a}^H(v_\ell + \Delta v_\ell) \mathbf{a}(v_\ell)] \otimes [\mathbf{a}^H(u_\ell + \Delta u_\ell) \mathbf{a}(u_\ell)]|^2 \right] \\ &= E\left[|[\mathbf{a}^H(v_\ell + \Delta v_\ell) \mathbf{a}(v_\ell)][\mathbf{a}^H(u_\ell + \Delta u_\ell) \mathbf{a}(u_\ell)]|^2 \right] \\ &= E\left[|\mathbf{a}^H(v_\ell + \Delta v_\ell) \mathbf{a}(v_\ell)|^2 \right] E[|\mathbf{a}^H(u_\ell + \Delta u_\ell) \mathbf{a}(u_\ell)|^2], \end{aligned} \quad (48)$$

where the last equation follows based on the assumption that Δv_ℓ and Δu_ℓ are independent. Now, we can write

$$\begin{aligned} & \left| \mathbf{a}^H(v_\ell + \Delta v_\ell) \mathbf{a}(v_\ell) \right|^2 \\ &= \left| 1 + e^{jv_\ell} e^{-j(v_\ell + \Delta v_\ell)} + \ldots + e^{j(M_1-1)v_\ell} e^{-j(M_1-1)(v_\ell + \Delta v_\ell)} \right|^2 \\ &= \left| 1 + e^{-j\Delta v_\ell} + (e^{-j\Delta v_\ell})^2 + \ldots + (e^{-j\Delta v_\ell})^{(M_1-1)} \right|^2 \\ &= \left| \frac{1 - e^{-jM_1 \Delta v_\ell}}{1 - e^{-j\Delta v_\ell}} \right|^2 = \left| \frac{1 - \cos(M_1 \Delta v_\ell) + j \sin(M_1 \Delta v_\ell)}{1 - \cos(\Delta v_\ell) + j \sin(\Delta v_\ell)} \right|^2 \\ &= \frac{1 - \cos(M_1 \Delta v_\ell)}{1 - \cos(\Delta v_\ell)} \\ &= \frac{1 - \left(1 - \frac{(M_1 \Delta v_\ell)^2}{2!} + \frac{(M_1 \Delta v_\ell)^4}{4!} + \ldots + \infty \right)}{1 - \left(1 - \frac{(\Delta v_\ell)^2}{2!} + \frac{(\Delta v_\ell)^4}{4!} + \ldots + \infty \right)} \\ &\overset{(\mathbf{a})}{\approx} \frac{\frac{M_1^2 (\Delta v_\ell)^2}{2} - \frac{M_1^4 (\Delta v_\ell)^4}{24}}{\frac{(\Delta v_\ell)^2}{2} - \frac{(\Delta v_\ell)^4}{24}} = \frac{M_1^2 \left(1 - \frac{M_1^2 (\Delta v_\ell)^2}{12} \right)}{1 - \frac{(\Delta v_\ell)^2}{12}} \\ &\overset{(\mathbf{b})}{\approx} M_1^2 \left(1 - \frac{M_1^2 (\Delta v_\ell)^2}{12} \right), \end{aligned} \quad (49)$$

where $(\mathbf{a})$ results from truncating the higher order terms, and $(\mathbf{b})$ results by noticing that $(\Delta v_\ell)^2/12$ is negligible compared to unity. Therefore,

$$E[|\mathbf{a}^H(v_\ell + \Delta v_\ell) \mathbf{a}(v_\ell)|^2] = M_1^2 \left(1 - \frac{M_1^2 E\left[(\Delta v_\ell)^2 \right]}{12} \right). \quad (50)$$

Following exactly same procedure, we can also show that

$$E[|\mathbf{a}^H(u_\ell + \Delta u_\ell) \mathbf{a}(u_\ell)|^2] = M_2^2 \left(1 - \frac{M_2^2 E\left[(\Delta u_\ell)^2 \right]}{12} \right). \quad (51)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

Therefore, from (48), we have

$$\begin{aligned}&\frac{1}{E[|\mathbf{e}_r^T(\ell,k)\hat{\mathbf{e}}_r^*(\ell,k)|^2}\\&=\frac{1}{M_1^2M_2^2}\left(1-\frac{M_1^2E\left[(\Delta v_\ell)^2\right]}{12}\right)^{-1}\left(1-\frac{M_2^2E\left[(\Delta u_\ell)^2\right]}{12}\right)^{-1}\\&\approx\frac{1}{M_1^2M_2^2}\left(1+\frac{M_1^2E\left[(\Delta v_\ell)^2\right]}{12}\right)\left(1+\frac{M_2^2E\left[(\Delta u_\ell)^2\right]}{12}\right).\end{aligned} \tag{52}$$

Hence, from (47):

$$E\{\hat{p}_\ell(k)\}=\left[\hat{\mu}_\ell(k)-\frac{1}{\hat{\gamma}_\ell M_1^2M_2^2}\left(1+\frac{M_1^2E\left[(\Delta v_\ell)^2\right]}{12}\right)\left(1+\frac{M_2^2E\left[(\Delta u_\ell)^2\right]}{12}\right)\right]^\diamond. \tag{53}$$

From (53), it can be observed that in the absence of DoA estimation error, the proposed power allocation algorithm becomes identical to traditional water-filling solution.

V. Simulation Results

In this section, we evaluate the RMSE of the ESPRIT-based DoA estimation for 3D millimeter wave massive MIMO systems. Furthermore, we will evaluate the system achievable rate under the presence of DoA estimation errors and present a performance comparison between our proposed power allocation algorithm and the traditional water-filling solution. To evaluate the performance of the DoA estimation, we assume there are 4 resolvable paths, which is a typical number for the outdoor millimeter-wave communication systems at both 28GHz and 73GHz [21]. Number of subcarriers of the OFDM system is 32, and the length of cyclic prefix is 4. The antenna spacing for both the received and transmit antennas is assumed to be 0.5λ. The number of transmit antennas is set to be 8. The elevation and azimuth DoAs are chosen randomly from either the uniform distribution: $U[-180°, 180°]$, the exponential distribution: $Exp(1/50)$, or the Gaussian distribution: $\mathcal{N}(50, 20)$.

In our work, we invoke the far field assumption, and the wavefront impinging on the antenna array was assumed to be planer. However, using models such as Costa model, it would not be difficult to extend our results to incorporate the spherical wavefront. It is also noteworthy here that the transmission medium assumed in this paper is isotropic and linear. Normally distributed random numbers were used for generating frequency domain data for simulation. The number of snapshots is taken to be 32. The success of ESPRIT type high resolution DoA estimation method depends on the full rank condition of the data covariance matrix. However, if appropriate preprocessing schemes such as forward-backward averaging or spatial smoothing can be applied, the data covariance matrix can be ensured to be of full rank and non-singular even when all the data signals are correlated. This is the main reason why we performed the forward-backward averaging in our DoA estimation process in Section II to deal with the potential issue of not having enough independent realizations of the random wavefield. Therefore, even though we used independent data signals in our simulation evaluation, our method/simulation procedure is expected to handle the non-singular covariance matrix case smoothly. The main requirement for forward-backward averaging to be valid is that the properties of the process under consideration be approximately the same independent of the orientation of space axis and that the samples be taken in a geometry that is also reversible. The antenna array that we are using at the base station fulfills this condition since it has a centro-symmetric structure. For the forward-backward averaging, the data are first taken from the 2D rectangular array, and then these data are vectorized. Finally, an extended data matrix is used for forward-backward averaging which is defined in Equation (10). Through forward-backward averaging, the data is effectively doubled to improve the estimation performance. Finally, the total available transmit power is assumed to be unity, and the SNR is defined as the ratio of the received signal power to the noise power, i.e. $\text{SNR} = 10\log_{10}\left(1/\sigma^2\right)$.

The performance of the estimation of elevation and azimuth angles for an 8×8 antenna array with uniform DoA distribution is shown in Figure 1 and Figure 2, respectively, where the analytical results for RMSEs are compared with the empirical ones. It can be observed that as SNR increases, the empirical results match the analytical results asymptotically. We can

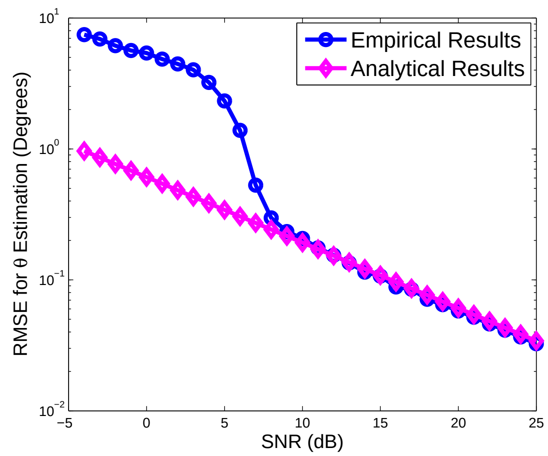

Figure 1: Elevation angle estimation for 8×8 array (Uniform DoA distribution).

investigate the impact of various antenna configurations on the estimation performance and obtain some interesting design intuitions. For example, for a 3D massive MIMO system with the same total 64 antenna elements, the RMSE for 4×16 antenna array is shown in Figure 3 and Figure 4, while the elevation and azimuth angle estimation for a 16×4 array are shown in Figure 5 and Figure 6, respectively. Comparing these figures with Figure 1 and 2, we observe that as the number of antennas in the elevation domain increases, the elevation angle estimation performance improves. However, the RMSE of azimuth estimation of an 8×8 array shown in Figure 2 is even better than that of the 4×16 array shown in Figure 4, especially at low and medium SNR regime. This is somewhat surprising because 4×16 array has more antenna elements in the azimuth domain compared with the 8×8 array. The reason behind this is that the azimuth estimation is actually

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

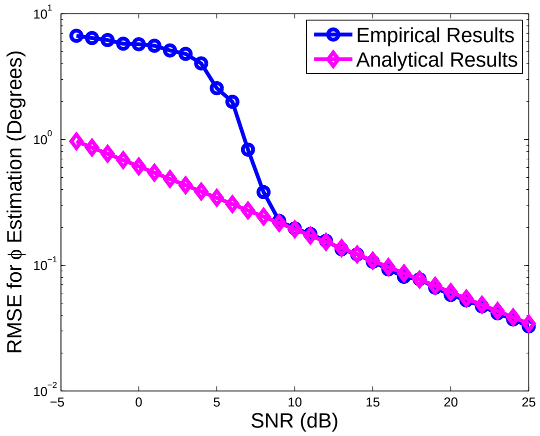

Figure 2: Azimuth angle estimation for 8×8 array (Uniform DoA distribution).

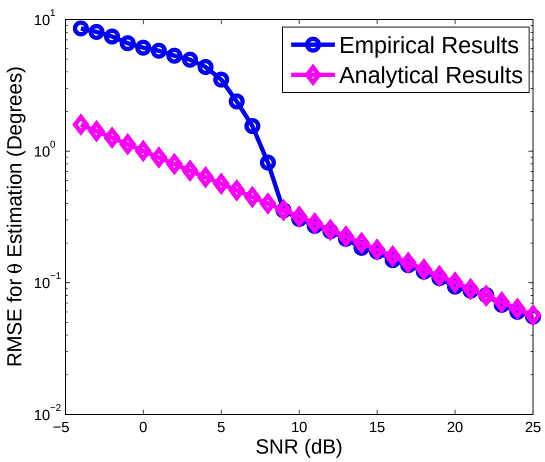

Figure 3: Elevation angle estimation for 4×16 array (Uniform DoA distribution).

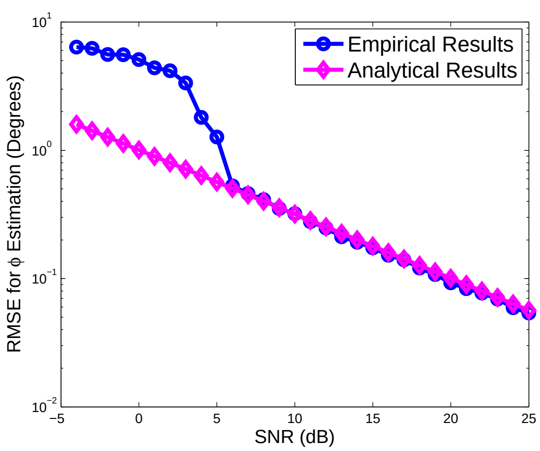

Figure 4: Azimuth angle estimation for 4×16 array (Uniform DoA distribution).

coupled with the elevation estimation performance, and when the elevation estimation performance decreases, azimuth estimation performance also deteriorates even though the number of horizontal antennas is increased. On the other hand, the elevation angle estimation performance does not depend on the azimuth angle estimation performance. For the case of

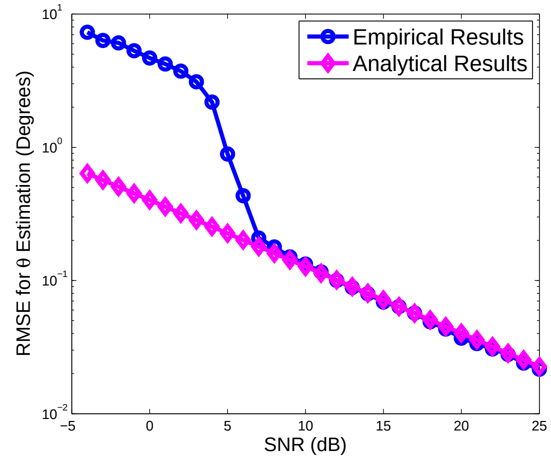

Figure 5: Elevation angle estimation for 16×4 array (Uniform DoA distribution).

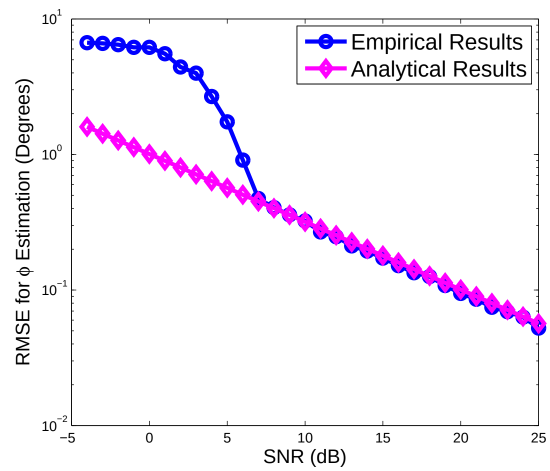

Figure 6: Azimuth angle estimation for 16×4 array (Uniform DoA distribution).

16×16 antenna array, the RMSE estimation performance is shown in Figure 7 and 8. As expected, the 16×16 array outperforms the 8×8, 4×16 and 16×4 arrays in both elevation and azimuth angle estimation. It is to be emphasized here that these DoA estimation results hold for any continuous distribution, and is not specific to uniform distribution only. The results for 8×8 and 16×16 antenna arrays where the DoA's are drawn from Gaussian $(\mathcal{N}(50, 20))$ and exponential distribution $(Exp(1/50))$, are shown in Figures 9 to 12. From these results, and also comparing with Figures 1,2,7, and 8, we can clearly observe that for the same antenna configuration, the DoA estimation performance is very similar irrespective of the underlying distribution from which the DoA's are drawn. This, in fact, validates the results in Lemmas 1 to 3.

Figures that plot the RMSE of elevation and azimuth angles estimation as a function of the number of antennas are shown

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

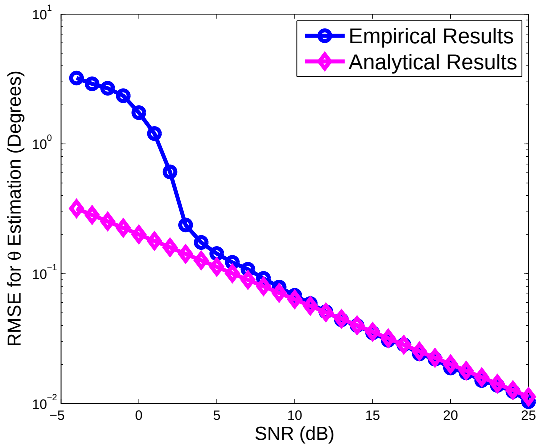

Figure 7: Elevation angle estimation for 16×16 array (Uniform DoA distribution).

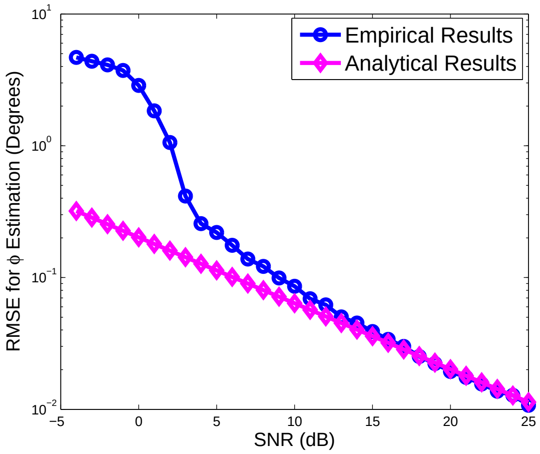

Figure 8: Azimuth angle estimation for 16×16 array (Uniform DoA distribution).

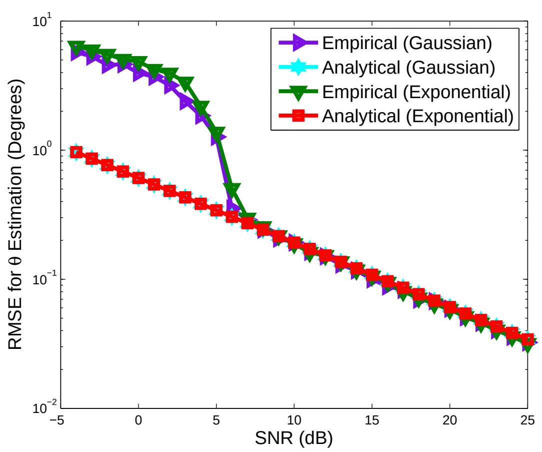

Figure 9: Elevation angle estimation for 8×8 array.

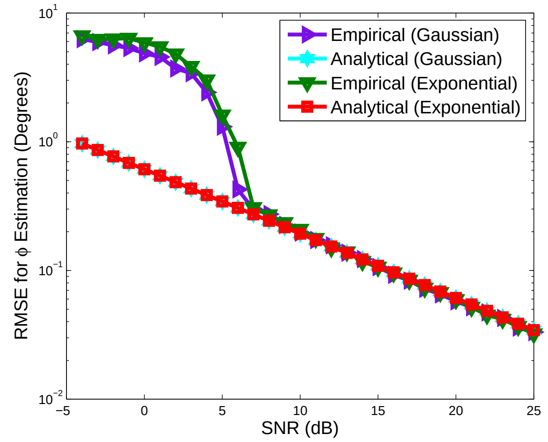

Figure 10: Azimuth angle estimation for 8×8 array.

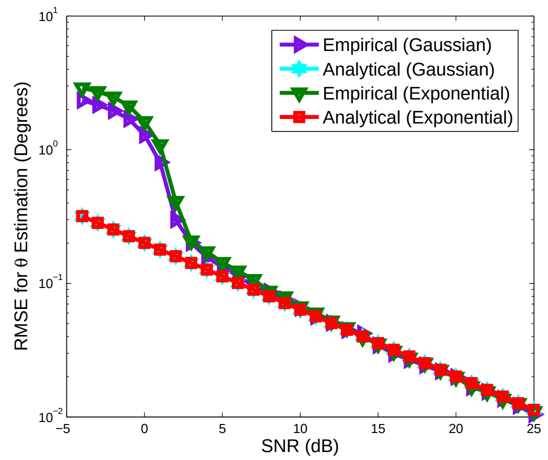

Figure 11: Elevation angle estimation for 16×16 array.

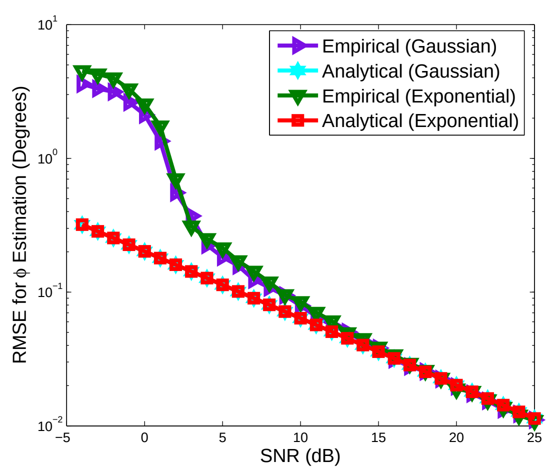

Figure 12: Azimuth angle estimation for 16×16 array.

in Figures 13 and 14 for uniform, Gaussian, and exponential DoA distributions, and SNR = 15 and 20 dB. Dashed lines represent the empirical results while solid lines represent the analytical results. The square array, where $M_1 = M_2$, is assumed for the evaluation. It can be observed that the DoA estimation performance improves as the number of antennas increases. Furthermore, in all cases, the empirical results are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

very close to the analytical results irrespective of the DoA distribution.

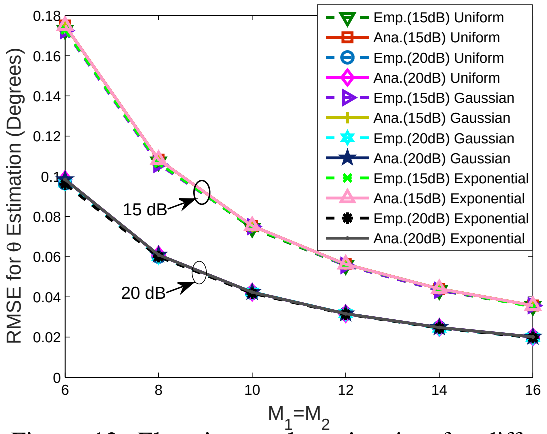

Figure 13: Elevation angle estimation for different number of antennas.

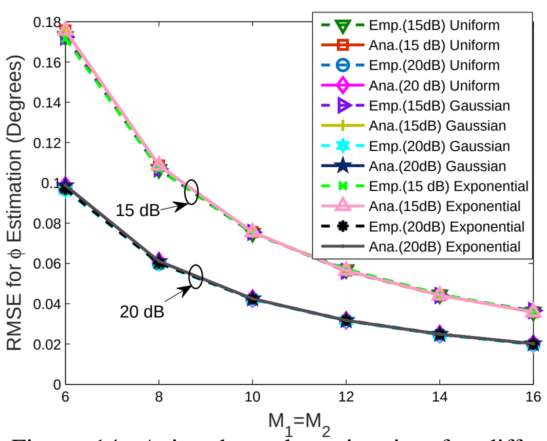

Figure 14: Azimuth angle estimation for different number of antennas.

Figure 15 provides a throughput comparison of the system in the presence of the DoA estimation error for uniform DoA distribution. To be specific, three schemes are compared in Figure 15: the first one utilizes the introduced DoA estimation for channel acquisition and conducts the corresponding water-filling power allocation; the second one utilizes the introduced DoA estimation for channel acquisition and conducts the introduced power allocation strategy in Equation (53); the third one employs the linear mean squared error (LMMSE) channel estimation to acquire the channel transfer functions and conducts the corresponding water-filling power allocation. In performing LMMSE channel estimation, no apriori knowledge of second order statistics, i.e. channel correlation matrix, is assumed to be available at the base station. Instead, each column of the $N_r \times N_t$ channel matrix for each subcarrier is estimated by sending the $N_t \times 1$ transmit signal vector of the form $[0, \ldots, 0, 1, 0, \ldots, 0]$ with the l-th element of the vector being 1 which is used to estimate the l-th column of the channel matrix. Therefore, in LMMSE channel estimation it takes at least N_t OFDM symbols to estimate the $N_r \times N_t$ channel matrix. The comparison in Figure 15 is performed for an 8×8 antenna array, and for uniform DoA distribution. It can be observed that in the low and medium SNR regime, the proposed power allocation algorithm outperforms the traditional water-filling scheme. However, in the high-SNR regime, both algorithms perform almost identically. This is because in the high-SNR regime, the DoA estimation is very accurate with very small estimation errors. As shown in Section IV-B, our introduced power allocation strategy will converge to the traditional water-filling one as DoA estimation error decreases. Furthermore, from Figure 15, it is obvious that both of the parametric-based methods outperform the LMMSE-based channel estimation in terms of the throughput of the underlying 3D massive MIMO/FD-MIMO OFDM system. This is mainly because in parametric-based estimation methods we exploit the channel structure for the underlying 3D MIMO channel, which significantly improves the estimation performance. Chordal distance between the subspace spanned

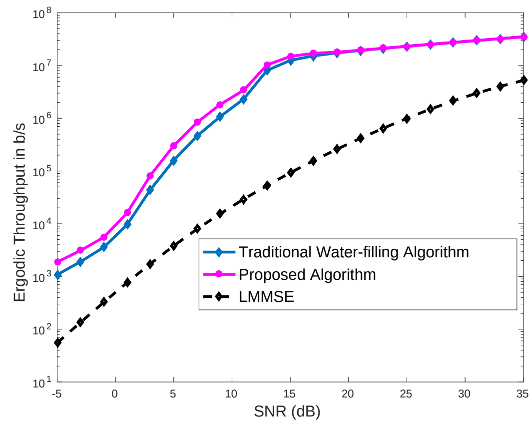

Figure 15: Ergodic Throughput under Traditional and Proposed Power Allocation Algorithms.

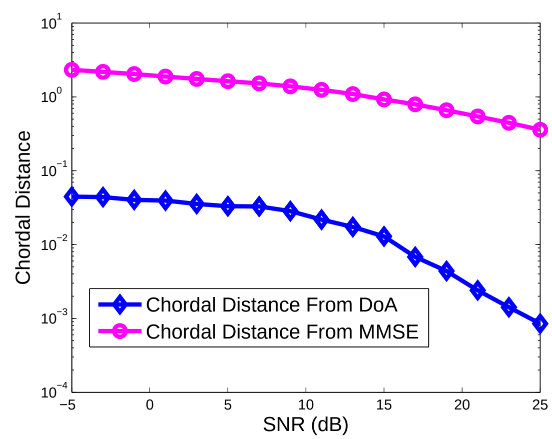

Figure 16: Chordal Distance Comparison.

by different sets of precoding vectors is another metric that is widely used for evaluating the underlying MIMO precoding strategies. In Figure 16, we present a chordal distance comparison between the Rank-4 precoding matrix obtained from the DoA estimation and that obtained from the LMMSE-based method. To be specific, let $\mathbf{U}_H$ denote the matrix containing four left-singular vectors corresponding to the four nonzero singular values obtained from SVD of the underlying uplink channel, $\mathbf{H}$. Note that $\mathbf{U}_H$ essentially contains optimal downlink MIMO precoders. Let $\mathbf{U}_{DoA}$ denote the estimated steering matrix; and $\mathbf{U}_{MMSE}$ denote the matrix attained by taking the four dominant singular vectors obtained from the SVD of the channel estimated using the LMMSE method. Now the chordal distance between the subspace spanned by the optimal precoding matrix $\mathbf{U}_H$, and $\mathbf{U}_{DoA}$ is defined as $\gamma_{DoA} = 1/\sqrt{2}E\left\{||\mathbf{U}_H\mathbf{U}_H^H - \mathbf{U}_{DoA}\mathbf{U}_{DoA}^H||_F\right\}$. Similarly, $\gamma_{MMSE} = 1/\sqrt{2}E\left\{||\mathbf{U}_H\mathbf{U}_H^H - \mathbf{U}_{MMSE}\mathbf{U}_{MMSE}^H||_F\right\}$. It is important to note that chordal distance is a measure of closeness of subspaces spanned by the corresponding basis vectors. As we can see from Figure 16, the chordal distance between the precoders obtained from DoA-based channel estimation and that from the ideal channel is much smaller. This means that DoA-based precoding is much more effective than the MIMO precoding based on LMMSE channel estimation.

VI. Conclusion

In this paper, we presented DoA-based channel estimation strategies as well as novel power allocation algorithms for 3D massive-MIMO/FD-MIMO OFDM systems. To be specific, we derived the analytical expression for the RMSE of ESPRIT-type DoA estimation algorithms. The performance of elevation and azimuth angle estimation of various antenna configurations have been investigated. The capacity and achievable rate of 3D massive MIMO/FD-MIMO downlink systems are characterized, and the optimal precoding matrix is related to DoA vectors. The optimal power allocation under the presence of DoA estimation error is also introduced.

DoA estimation evaluation showed that the empirical performance of DoA estimation matches very well with that of the analytical performance. Furthermore, the antenna array configuration plays a vital role in determining the performance of the 3D massive-MIMO/FD-MIMO OFDM systems: the performance of azimuth angle estimation is tightly coupled with that of elevation angle estimation. The correlation among data on different subcarriers has an adverse effect on the DoA estimation performance. An ergodic throughput evaluation suggested that the DoA-based channel estimation achieves much higher throughput than traditional non-parametric channel estimations such as LMMSE, while our introduced power allocation strategy outperforms the traditional water-filling scheme. These observations may provide significant insights and intuitions toward designing realistic 3D massive-MIMO/FD-MIMO OFDM systems for 5G networks.

Appendix A
Proof of Lemma 1

Proof. For any $i \neq m$,

$$[\bar{\mathbf{e}}_r(i)]^H[\bar{\mathbf{e}}_r(m)] = \frac{1}{N_r}\left[\mathbf{a}(v_i)\otimes\mathbf{a}(u_i)\right]^H\left[\mathbf{a}(v_m)\otimes\mathbf{a}(u_m)\right]$$

$$\begin{aligned}
&= \frac{1}{N_r}[1, e^{-ju_i}, \ldots, e^{-j(M_1-1)u_i}, \ldots, e^{-j(M_2-1)v_i},\\
&\qquad e^{-j(M_2-1)v_i}e^{-ju_i}, \ldots, e^{-j(M_2-1)v_i}e^{-j(M_1-1)u_i}]\\
&\quad [1, e^{ju_m}, \ldots, e^{j(M_1-1)u_m}, \ldots, e^{j(M_2-1)v_m},\\
&\qquad e^{j(M_2-1)v_m}e^{ju_m}, \ldots, e^{j(M_2-1)v_m}e^{j(M_1-1)u_m}]^T\\
&= \frac{1}{N_r}[1 + e^{-j(u_i-u_m)}+, \ldots, +e^{-j(M_1-1)(u_i-u_m)}+, \ldots,\\
&\quad + e^{-j(M_2-1)(v_i-v_m)} + e^{-j(M_2-1)(v_i-v_m)}e^{-j(u_i-u_m)}+, \ldots,\\
&\quad \ldots, +e^{-j(M_2-1)(v_i-v_m)}e^{-j(M_1-1)(u_i-u_m)}]\\
&= \frac{1}{N_r}\sum_{n=0}^{M_2-1}\sum_{p=0}^{M_1-1}[e^{-j(v_i-v_m)}]^n[e^{-j(u_i-u_m)}]^p
\end{aligned} \tag{54}$$

Since both azimuth and elevation DoA's are drawn independently from a continuous distribution, $(v_i - v_m \neq 0)$ with probability = 1; therefore $e^{-j(v_i-v_m)} \neq 1$. Similarly, $e^{-j(u_i-u_m)} \neq 1$. We can write,

$$\begin{aligned}
&\lim_{N_r\to\infty}\left|[\bar{\mathbf{e}}_r(i,k)]^H[\bar{\mathbf{e}}_r(j,k)]\right|\\
&= \lim_{N_r\to\infty}\frac{1}{N_r}\left|\sum_{n=0}^{M_2-1}\sum_{p=0}^{M_1-1}[e^{-j(v_i-v_m)}]^n[e^{-j(u_i-u_m)}]^p\right|\\
&\overset{a}{=} \lim_{N_r\to\infty}\frac{1}{N_r}\left|\frac{1-e^{-j(u_i-u_m)M_1}}{1-e^{-j(u_i-u_m)}}\right|\left|\frac{1-e^{-j(v_i-v_m)M_2}}{1-e^{-j(v_i-v_m)}}\right|\\
&\leq \lim_{N_r\to\infty}\frac{1}{N_r}\left|\frac{2}{1-e^{-j(u_i-u_m)}}\right|\left|\frac{2}{1-e^{-j(v_i-v_m)}}\right| = 0
\end{aligned} \tag{55}$$

where, (a) follows by noticing that both $\sum_{n=0}^{M_2-1}[e^{-j(v_i-v_m)}]^n$ and $\sum_{p=0}^{M_1-1}[e^{-j(u_i-u_m)}]^p$ form geometric series with the ratios $e^{-j(v_i-v_m)}$ and $e^{-j(u_i-u_m)}$, respectively. □

Appendix B
Proof of Theorem 1

Proof. In the case of standard ESPRIT and for circularly symmetric white noise, $\mathbf{C}_{nn} = \mathbf{0}$ [20]. So, we can write (16) and (19) as

$$E\{(\Delta v_\ell)^2\} = \frac{1}{2}({\mathbf{r}_\ell^{(v)}}^H\mathbf{W}_{mat}^*\mathbf{R}_{nn}^T\mathbf{W}_{mat}^T\mathbf{r}_\ell^{(v)}), \tag{56}$$

$$E\{(\Delta u_\ell)^2\} = \frac{1}{2}({\mathbf{r}_\ell^{(u)}}^H\mathbf{W}_{mat}^*\mathbf{R}_{nn}^T\mathbf{W}_{mat}^T\mathbf{r}_\ell^{(u)}). \tag{57}$$

Let us now denote

$$\boldsymbol{\beta}_\ell = \mathbf{V}_s\boldsymbol{\Sigma}_s^{-1}\mathbf{q}_\ell, \tag{58}$$

$$\boldsymbol{\alpha}_{v,\ell} = (\mathbf{p}_\ell^T(\tilde{\mathbf{J}}_1^{(v)}\mathbf{U}_s)^+(\tilde{\mathbf{J}}_2^{(v)}/e^{jv_\ell} - \tilde{\mathbf{J}}_1^{(v)})(\mathbf{U}_n\mathbf{U}_n^H))^T, \tag{59}$$

$$\boldsymbol{\alpha}_{u,\ell} = (\mathbf{p}_\ell^T(\tilde{\mathbf{J}}_1^{(u)}\mathbf{U}_s)^+(\tilde{\mathbf{J}}_2^{(u)}/e^{ju_\ell} - \tilde{\mathbf{J}}_1^{(u)})(\mathbf{U}_n\mathbf{U}_n^H))^T. \tag{60}$$

So, we can have

$$\begin{aligned}
\mathbf{W}_{mat}^T\mathbf{r}_\ell^{(v)} &= \left((\boldsymbol{\Sigma}_s^{-1}\mathbf{V}_s^T)\otimes(\mathbf{U}_n\mathbf{U}_n^H)\right)^T\\
&\quad\left(\mathbf{q}_\ell\otimes\left(\left[(\tilde{\mathbf{J}}_1^{(v)}\mathbf{U}_s)^+(\tilde{\mathbf{J}}_2^{(v)}/e^{jv_\ell} - \mathbf{J}_1^{(v)})\right]^T\mathbf{p}_\ell\right)\right)\\
&= \left(\mathbf{V}_s\boldsymbol{\Sigma}_s^{-1}\mathbf{q}_\ell\right)\\
&\quad\otimes\left(\mathbf{p}_\ell^T(\tilde{\mathbf{J}}_1^{(v)}\mathbf{U}_s)^+\left(\tilde{\mathbf{J}}_2^{(v)}/e^{jv_\ell} - \mathbf{J}_1^{(v)}\right)\left(\mathbf{U}_n\mathbf{U}_n^H\right)\right)^T\\
&= \boldsymbol{\beta}_{\boldsymbol{\ell}}\otimes\boldsymbol{\alpha}_{\boldsymbol{v},\boldsymbol{\ell}}.
\end{aligned} \tag{61}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

Similarly, $\mathbf{W}_{mat}^T \mathbf{r}_\ell^{(u)} = \boldsymbol{\beta}_\ell \otimes \boldsymbol{\alpha}_{u,\ell}$. The MSE in (56) and (57) then can be expressed as

$$E\{(\Delta v_\ell)^2\} = \frac{1}{2}((\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,\ell}})^H \mathbf{R}_{nn}^T (\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,l}})), \quad (62)$$

$$E\{(\Delta u_\ell)^2\} = \frac{1}{2}((\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{u,\ell}})^H \mathbf{R}_{nn}^T (\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{u,\ell}})). \quad (63)$$

It can be easily verified that $\boldsymbol{\alpha}_{v,l}$ can be written as $\boldsymbol{\alpha}_{v,l}^T = \mathbf{c}_\ell^T \left(\left(\tilde{\mathbf{J}}_{v,2}\mathbf{A}\right)^+ \tilde{\mathbf{J}}_{v,2} - \left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^+ \tilde{\mathbf{J}}_{v,1} \right)$; where $\mathbf{c}_\ell = [0,\ldots,1,\ldots 0]^T$ is the column selection vector with ℓ-th element being one, and other elements being zero, $\tilde{\mathbf{J}}_{v,1} = \mathbf{I}_{M_1} \otimes [\mathbf{I}_{M_2-1} \quad \mathbf{0}]$ and $\tilde{\mathbf{J}}_{v,2} = \mathbf{I}_{M_1} \otimes [\mathbf{0} \quad \mathbf{I}_{M_2-1}]$ are the selection matrices. The pseudo inverse of the selected signal can be significantly simplified as follows:

$$\begin{aligned}\left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^+ &= \left(\left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H (\mathbf{J}_{v,1}\mathbf{A}) \right)^{-1} \left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H \\ &= \frac{1}{(M_2-1)M_1} \left(\frac{\left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H \left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)}{(M_2-1)M_1} \right)^{-1} \left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H \quad (64) \\ &\stackrel{(a)}{=} \frac{1}{(M_2-1)M_1} \left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H,\end{aligned}$$

where (a) holds due to **Lemma** 1. Similarly, we have

$$\left(\tilde{\mathbf{J}}_{v,2}\mathbf{A}\right)^+ = \frac{1}{(M_2-1)M_1} \left(\tilde{\mathbf{J}}_{v,2}\mathbf{A}\right)^H. \quad (65)$$

In the same way, $\boldsymbol{\alpha}_{u,l}$ can be written as $\boldsymbol{\alpha}_{u,l}^T = \mathbf{c}_\ell^T \left(\left(\tilde{\mathbf{J}}_{u,2}\mathbf{A}\right)^+ \tilde{\mathbf{J}}_{u,2} - \left(\tilde{\mathbf{J}}_{u,1}\mathbf{A}\right)^+ \tilde{\mathbf{J}}_{u,1} \right)$; where $\tilde{\mathbf{J}}_{u,1} = [\mathbf{I}_{M_1-1} \quad \mathbf{0}] \otimes \mathbf{I}_{M_2}$ and $\tilde{\mathbf{J}}_{u,2} = [\mathbf{0} \quad \mathbf{I}_{M_1-1}] \otimes \mathbf{I}_{M_2}$ are the selection matrices. Therefore, we can have

$$\left(\tilde{\mathbf{J}}_{u,1}\mathbf{A}\right)^+ = \frac{1}{(M_1-1)M_2} \left(\tilde{\mathbf{J}}_{u,1}\mathbf{A}\right)^H, \quad (66)$$

$$\left(\tilde{\mathbf{J}}_{u,2}\mathbf{A}\right)^+ = \frac{1}{(M_1-1)M_2} \left(\tilde{\mathbf{J}}_{u,2}\mathbf{A}\right)^H. \quad (67)$$

Now,

$$\tilde{\mathbf{J}}_{v,1} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}_{M_1 \times M_1} \otimes \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}_{M_2-1 \times M_2} \quad (68)$$

The ℓ-th row of $\left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H$ thus can be written as

$$\left[1, e^{-ju_\ell}, \ldots, e^{-j(M_2-2)u_\ell}, \ldots, e^{-j(M_2-1)v_\ell} e^{-j(M_1-1)u_\ell}\right].$$

Accordingly, ℓ-th row of $\left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H \tilde{\mathbf{J}}_{v,1}$ can be written as

$$\begin{aligned}&\mathbf{c}_\ell^T \left(\left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H \tilde{\mathbf{J}}_{v,1} \right) \\ &= \left[1, e^{-ju_\ell}, \ldots, e^{-j(M_2-2)u_\ell}, \ldots, e^{-j(M_2-1)v_\ell} e^{-j(M_1-2)u_\ell}, 0\right]. \quad (69)\end{aligned}$$

Moreover, $\tilde{\mathbf{J}}_{v,2}$ can be represented as

$$\tilde{\mathbf{J}}_{v,2} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}_{M_1 \times M_1} \otimes \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}_{M_2-1 \times M_2} \quad (70)$$

The ℓ-th row of $\left(\tilde{\mathbf{J}}_{v,2}\mathbf{A}\right)^H$ then can be written as

$$\left[e^{-ju_\ell}, e^{-j2u_\ell}, \ldots, e^{-j(M_2-1)u_\ell}, \ldots, e^{-j(M_2-1)v_\ell} e^{-j(M_1-1)u_\ell}\right],$$

and the ℓ-th row of $\left(\tilde{\mathbf{J}}_{v,2}\mathbf{A}\right)^H \tilde{\mathbf{J}}_{v,1}$ can be expressed as

$$\begin{aligned}&\mathbf{c}_\ell^T \left(\left(\tilde{\mathbf{J}}_{v,2}\mathbf{A}\right)^H \tilde{\mathbf{J}}_{v,2} \right) \\ &= \left[0, e^{-ju_\ell}, \ldots, e^{-j(M_2-1)u_\ell}, \ldots, e^{-j(M_2-1)v_\ell} e^{-j(M_1-1)u_\ell}, 0\right]. \quad (71)\end{aligned}$$

Combining (69) and (71), we can obtain

$$\begin{aligned}&\mathbf{c}_\ell^T \left(\left(\tilde{\mathbf{J}}_{v,2}\mathbf{A}\right)^H \tilde{\mathbf{J}}_{v,2} - \left(\tilde{\mathbf{J}}_{v,1}\mathbf{A}\right)^H \tilde{\mathbf{J}}_{v,1} \right) \\ &= \left[-1, 0, \ldots, 0, e^{-j((M_2-1)u_\ell)}, \ldots, e^{-j((M_1-1)u_\ell)}, 0, \right. \\ &\quad \left. , \ldots, e^{-j((M_2-1)v_\ell + (M_1-1)u_\ell)}\right]. \quad (72)\end{aligned}$$

Accordingly, we obtain $||\boldsymbol{\alpha}_{v,\ell}||^2 = 2/(M_2-1)^2 M_1$. Following the same procedure, we can also have $||\boldsymbol{\alpha}_{u,\ell}||^2 = 2/(M_1-1)^2 M_2$. With a view to deriving the expression for $\boldsymbol{\beta_\ell}$, we first rewrite the singular value decomposition of the forward-backward averaged noiseless received signal as:

$$\begin{aligned}&\left[\mathbf{A}\mathrm{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\tilde{\mathbf{S}} \quad \boldsymbol{\Pi}_{N_r}\mathbf{A}^*\mathrm{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t^*\tilde{\mathbf{S}}^*\boldsymbol{\Pi}_V\right] \\ &= \left[\mathbf{A}\mathrm{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t\tilde{\mathbf{S}} \quad \mathbf{A}\boldsymbol{\Lambda}\mathrm{diag}\{\mathbf{b}\}\tilde{\mathbf{E}}_t^*\tilde{\mathbf{S}}^*\boldsymbol{\Pi}_V\right] \\ &= \mathbf{A}\mathrm{diag}\{\mathbf{b}\}\left[\tilde{\mathbf{E}}_t\tilde{\mathbf{S}} \quad \boldsymbol{\Lambda}\tilde{\mathbf{E}}_t^*\tilde{\mathbf{S}}^*\boldsymbol{\Pi}_V\right], \quad (73)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\Lambda} =& \mathrm{diag}\{[e^{-j((M_1-1)u_n+(M_2-1)v_n)}, \ldots, \\ &, \ldots, e^{-j((M_1-1)u_{(n-N_c+1)}+(M_2-1)v_{(n-N_c+1)})}]\}. \quad (74)\end{aligned}$$

Comparing with (11), and based on **Lemma** 1, we have $\mathbf{U}_s = 1/\sqrt{N_r}\mathbf{A}$, $\boldsymbol{\Sigma}_s = \sqrt{2N_rV}\mathrm{diag}\{\mathbf{b}\}$, and $\mathbf{V}_s^H = 1/\sqrt{2V}\left[\tilde{\mathbf{E}}_t\tilde{\mathbf{S}} \quad \boldsymbol{\Lambda}\tilde{\mathbf{E}}_t^*\tilde{\mathbf{S}}^*\boldsymbol{\Pi}_V\right]$. The vector, $\boldsymbol{\beta_\ell}$ is given by $\boldsymbol{\beta_\ell} = \mathbf{V}_s\Sigma_s^{-1}\mathbf{U}_s^H\mathbf{A}\mathbf{c}_\ell$ [22]. It can be shown that the unitary trransformation does not affect the MSE of the ESPRIT method [23]. However, the statistics of the noise and signal subspace are changed when preprocessing like forward-backward averaging is applied. Now, the noise covariance matrix can be written as:

$$\mathbf{R}_{nn} = \mathbf{E}\{\mathrm{vec}\{\mathbf{W}\}\mathrm{vec}\{\mathbf{W}\}^H\}. \quad (75)$$

However, if the noise is assumed to be circularly symmetric and white Gaussian, we have $\mathbf{R}_{nn} = \sigma^2\mathbf{I}_{N_rV}$. Therefore

$$\begin{aligned}(\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,\ell}})^H \mathbf{R}_{nn}^T (\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,\ell}}) &= \sigma^{\mathbf{2}} (\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,\ell}})^H (\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,\ell}}) \\ &= \sigma^{\mathbf{2}} (\boldsymbol{\beta_\ell}^H \boldsymbol{\beta_\ell}) \otimes (\boldsymbol{\alpha}_{v,\ell}^H \boldsymbol{\alpha}_{v,\ell}). \quad (76)\end{aligned}$$

Since $\boldsymbol{\beta_\ell}$ is the ℓ-th column of $\mathbf{V}_s\Sigma_s^{-1}\mathbf{U}_s^H\mathbf{A}$, $||\boldsymbol{\beta_\ell}||^2$ is the ℓ-th diagonal element of $\mathbf{A}^H\mathbf{U}_s\boldsymbol{\Sigma}_s^{-2}\mathbf{U}_s^H\mathbf{A}$, which we can write as $||\boldsymbol{\beta}_\ell||^2 = \mathbf{R}_{SS}^{-1}(\ell,\ell)/V$ [22], [24], where $\mathbf{R}_{SS}(\ell,\ell)$ is the ℓ-th

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

diagonal element of the equivalent transmit signal covariance matrix. Plugging the value of $||\boldsymbol{\beta}_\ell||^2$ and $||\boldsymbol{\alpha}_{v,\ell}||^2$ in (76):

$$(\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,\ell}})^H \mathbf{R}_{nn}^T (\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{v,\ell}}) = \sigma^{\mathbf{2}} \frac{\mathbf{R}_{SS}^{-1}(\ell,\ell)}{V} \frac{2}{(M_2-1)^2 M_1}. \tag{77}$$

Similarly, we also have

$$(\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{u,\ell}})^H \mathbf{R}_{nn}^T (\boldsymbol{\beta_\ell} \otimes \boldsymbol{\alpha_{u,\ell}}) = \sigma^{\mathbf{2}} \frac{\mathbf{R}_{SS}^{-1}(\ell,\ell)}{V} \frac{2}{(M_1-1)^2 M_2}. \tag{78}$$

Plugging the expression from (77) into (62) we get

$$E\{(\Delta v_\ell)^2\} = \frac{\mathbf{R}_{SS}^{-1}(\ell,\ell)}{V} \frac{\sigma^{\mathbf{2}}}{(M_2-1)^2 M_1}. \tag{79}$$

Similarly, by plugging the expression from (78) into (63):

$$E\{(\Delta u_\ell)^2\} = \frac{\mathbf{R}_{SS}^{-1}(\ell,\ell)}{V} \frac{\sigma^{\mathbf{2}}}{(M_1-1)^2 M_2}. \tag{80}$$

Based on Jacobian matrix, we have:

$$\mathbb{E}\left\{(\triangle\theta_\ell)^2\right\} = \mathbb{E}\left\{(\triangle u_\ell)^2\right\} \frac{1}{\pi^2 \sin^2(\theta_\ell)}, \tag{81}$$

$$\begin{aligned}\mathbb{E}\left\{(\triangle\phi_\ell)^2\right\} = & \frac{\mathbb{E}\left\{(\triangle u_\ell)^2\right\} \cot^2(\theta_\ell)\cot^2(\phi_\ell)}{\pi^2 \sin^2(\theta_\ell)} \\ & + \frac{\mathbb{E}\left\{(\triangle v_\ell)^2\right\}}{\pi^2 \sin^2(\theta_\ell)\sin^2(\phi_\ell)}.\end{aligned} \tag{82}$$

Recognizing that $\mathrm{RMSE}\{\theta_\ell\} = \sqrt{\mathbb{E}\{(\triangle\theta_\ell)^2\}}$ and $\mathrm{RMSE}\{\phi_\ell\} = \sqrt{\mathbb{E}\{(\triangle\phi_\ell)^2\}}$, by substituting (79) and (80) into (81) and (82), respectively, we obtain the desired results. □

Appendix C
Proof of Lemma 2

Proof. For any $i \neq m$,

$$\begin{aligned}
&[\bar{\mathbf{e}}_r(i,k)]^H[\bar{\mathbf{e}}_r(m,k)] = \frac{1}{N_r} e^{j\frac{2\pi k i}{N_c}} \mathbf{e}_r^H(i) e^{-j\frac{2\pi k m}{N_c}} \mathbf{e}_r(m) \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} \left[\mathbf{a}(v_i)\otimes\mathbf{a}(u_i)\right]^H \left[\mathbf{a}(v_m)\otimes\mathbf{a}(u_m)\right] \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} [1, e^{-ju_i}, \ldots, e^{-j(M_1-1)u_i}, \ldots, e^{-j(M_2-1)v_i}, \\
&\quad e^{-j(M_2-1)v_i} e^{-ju_i}, \ldots, e^{-j(M_2-1)v_i} e^{-j(M_1-1)u_i}] \\
&\quad [1, e^{ju_m}, \ldots, e^{j(M_1-1)u_m}, \ldots, e^{j(M_2-1)v_m}, \\
&\quad e^{j(M_2-1)v_m} e^{ju_m}, \ldots, e^{j(M_2-1)v_m} e^{j(M_1-1)u_m}]^T \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} [1 + e^{-j(u_i-u_m)} +, \ldots, + e^{-j(M_1-1)(u_i-u_m)} \\
&\quad +, \ldots, + e^{-j(M_2-1)(v_i-v_m)} + e^{-j(M_2-1)(v_i-v_m)} e^{-j(u_i-u_m)} \\
&\quad +, \ldots, + e^{-j(M_2-1)(v_i-v_m)} e^{-j(M_1-1)(u_i-u_m)}] \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} \sum_{n=0}^{M_2-1}\sum_{p=0}^{M_1-1} [e^{-j(v_i-v_m)}]^n [e^{-j(u_i-u_m)}]^p
\end{aligned} \tag{83}$$

Since both azimuth and elevation DoA's are drawn independently from a continuous distribution, $(v_i - v_m \neq 0)$ with probability = 1; therefore $e^{-j(v_i-v_m)} \neq 1$. Similarly, $e^{-j(u_i-u_m)} \neq 1$ with probability 1. We can write,

$$\begin{aligned}
&\lim_{N_r\to\infty} \left|[\bar{\mathbf{e}}_r(i,k)]^H[\bar{\mathbf{e}}_r(j,k)]\right| \\
&= \left|e^{j\frac{2\pi k(i-m)}{N_c}}\right| \lim_{N_r\to\infty} \frac{1}{N_r} \left|\sum_{n=0}^{M_2-1}\sum_{p=0}^{M_1-1} [e^{-j(v_i-v_m)}]^n [e^{-j(u_i-u_m)}]^p\right| \\
&\overset{a}{=} \left|e^{j\frac{2\pi k(i-m)}{N_c}}\right| \lim_{N_r\to\infty} \frac{1}{N_r} \left|\frac{1-e^{-j(u_i-u_m)M_1}}{1-e^{-j(u_i-u_m)}}\right| \left|\frac{1-e^{-j(v_i-v_m)M_2}}{1-e^{-j(v_i-v_m)}}\right| \\
&\leq \left|e^{j\frac{2\pi k(i-m)}{N_c}}\right| \lim_{N_r\to\infty} \frac{1}{N_r} \left|\frac{2}{1-e^{-j(u_i-u_m)}}\right| \left|\frac{2}{1-e^{-j(v_i-v_m)}}\right| = 0
\end{aligned} \tag{84}$$

where, (a) follows by noticing that both $\sum_{n=0}^{M_2-1}[e^{-j(v_i-v_m)}]^n$ and $\sum_{p=0}^{M_1-1}[e^{-j(u_i-u_m)}]^p$ form geometric series with the ratios $e^{-j(v_i-v_m)}$ and $e^{-j(u_i-u_m)}$, respectively. □

Appendix D
Proof of Lemma 3

Proof. For any $i \neq m$,

$$\begin{aligned}
&[\hat{\bar{\mathbf{e}}}_r(i,k)]^H[\bar{\mathbf{e}}_r(m,k)] = \frac{1}{N_r}\hat{\mathbf{e}}_r^H(i,k)\mathbf{e}_r(m,k) \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} \left[\mathbf{a}(v_i+\Delta v_i)\otimes\mathbf{a}(u_i+\Delta u_i)\right]^H \left[\mathbf{a}(v_m)\otimes\mathbf{a}(u_m)\right] \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} \Big[1, e^{-j(u_i+\Delta u_i)}, \ldots, e^{-j(M_1-1)(u_i+\Delta u_i)}, \ldots, \\
&\quad e^{-j(M_2-1)(v_i+\Delta v_i)}, e^{-j(M_2-1)(v_i+\Delta v_i)} e^{-j(u_i+\Delta u_i)}, \ldots, \\
&\quad e^{-j(M_2-1)(v_i+\Delta v_i)} e^{-j(M_1-1)(u_i+\Delta u_i)}\Big] \\
&\quad \Big[1, e^{ju_m}, \ldots, e^{j(M_1-1)u_m}, \ldots, e^{j(M_2-1)v_m}, \\
&\quad e^{j(M_2-1)v_m} e^{ju_m}, \ldots, e^{j(M_2-1)v_m} e^{j(M_1-1)u_m}\Big]^T \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} \Big[1 + e^{-j(u_i+\Delta u_i-u_m)} +, \ldots, \\
&\quad + e^{-j(M_1-1)(u_i+\Delta u_i-u_m)} +, \ldots, + e^{-j(M_2-1)(v_i+\Delta v_i-v_m)} \\
&\quad + e^{-j(M_2-1)(v_i+\Delta v_i-v_m)} e^{-j(u_i+\Delta u_i-u_m)} +, \ldots, \\
&\quad + e^{-j(M_2-1)(v_i+\Delta v_i-v_m)} e^{-j(M_1-1)(u_i+\Delta u_i-u_m)}\Big] \\
&= \frac{1}{N_r} e^{j\frac{2\pi k(i-m)}{N_c}} \sum_{n=0}^{M_2-1}\sum_{p=0}^{M_1-1} [e^{-j(v_i+\Delta v_i-v_m)}]^n [e^{-j(u_i+\Delta u_i-u_m)}]^p
\end{aligned} \tag{85}$$

Now, using the same procedure of the proof of Lemma 2, we can show that

$$\lim_{N_r\to\infty} \left|[\hat{\bar{\mathbf{e}}}_r(i,k)]^H[\bar{\mathbf{e}}_r(m,k)]\right| = 0. \tag{86}$$

Hence, $\hat{\bar{\mathbf{e}}}_r(i,k)$ and $\bar{\mathbf{e}}_r(m,k)$, $\forall i \neq m$, becomes orthonormal as N_r goes large. □

References

[1] R. Shafin, L. Liu, and J. Zhang, "DoA Estimation and RMSE Characterization for 3D Massive-MIMO/FD-MIMO OFDM System," in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[2] R. Shafin, L. Liu, and J. C. Zhang, "DoA Estimation and Capacity Analysis for 3D Massive-MIMO/FD-MIMO OFDM System," in *IEEE GlobalSIP*, Dec. 2015, pp. 181–184.

[3] *World Telecommunication/ICT Indicators Database 2015* International Telecommunications Union (ITU), Jun. 2015.

[4] Cisco Systems, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014-2019*, Feb. 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2594173, IEEE Transactions on Wireless Communications

[5] L. Liu, R. Chen, S. Geirhofer, K. Sayana, Z. Shi, and Y. Zhou, "Downlink MIMO in LTE-Advanced: SU-MIMO vs. MU-MIMO," *IEEE Commun. Mag.*, 50(2): 140–147, Feb. 2012.

[6] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced, 1st ed.* Wiley, Oct. 2012.

[7] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, 9(11): 3590–3600, Nov. 2010.

[8] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, 30(1): 40–60, Jan. 2013.

[9] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 myths and one grand question," *CoRR*, vol. abs/1503.06854, 2015.

[10] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, "Achieving "massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, 11(9): 3226–3239, Sept. 2012.

[11] Y. Kim, H. Ji, J. Lee, Y. Nam, B. Ng, I. Tzanidis, Y. Li, and J. Zhang, "Full dimension MIMO (FD-MIMO): The next evolution of MIMO in LTE systems," *IEEE Wireless Commun.*, 21(3): 92–100, Jun. 2014.

[12] B. Ng, Y. Kim, J. Lee, Y. Li, Y. Nam, J. Zhang, and K. Sayana, "Fulfilling the Promise of Massive MIMO with 2D Active Antenna Array," in *IEEE GLOBECOM Workshop*, Dec. 2012, pp. 691–696.

[13] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, 52(2): 186–195, Feb. 2014.

[14] B. Yang, K. Letaief, R. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, 49(3): 467–479, Mar. 2001.

[15] L. Cheng, Y.-C. Wu, J. Zhang, and L. Liu, "Subspace Identification for DOA Estimation in Massive/Full-Dimension MIMO Systems: Bad Data Mitigation and Automatic Source Enumeration," *IEEE Trans. Signal Process.*, 63(22): 5897–5909, Nov. 2015.

[16] G. Stuber, J. Barry, S. McLaughlin, Y. Li, M. Ingram, and T. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. of the IEEE*, 92(2): 271–294, Feb. 2004.

[17] A. Sayeed and T. Sivanadyan, *Handbook on Array Processing the Sensor Network*, Wiley, Jan. 2010, Ch. Wireless communication and sensing in multipath environments using multi-antenna transceivers.

[18] M. Larsen, A. Swindlehurst, and T. Svantesson, "Performance bounds for MIMO-OFDM channel estimation," *IEEE Trans. Signal Process.*, 57(5): 1901–1916, May 2009.

[19] A. Wang, L. Liu, and J. Zhang, "Low complexity direction of arrival (DoA) estimation for 2D massive MIMO systems," in *IEEE Globecom Workshops*, Dec. 2012, pp. 703–707.

[20] F. Roemer, M. Haardt, and G. Del Galdo, "Analytical performance assessment of multi-dimensional matrix- and tensor-based ESPRIT-type algorithms," *IEEE Trans. Signal Process.*, 62(10): 2611–2625, May 2014.

[21] S. Rangan, T. Rappaport, and E. Erkip, "Millimeter Wave Cellular Wireless Networks: Potentials and Challenges," *Proc. of IEEE*, 102(3): 366–385, Nov. 2014.

[22] F. Li, H. Liu, and R. Vaccaro, "Performance analysis for DoA estimation algorithms: unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst.*, 29(4): 1170–1184, Oct. 1993.

[23] F. Roemer, "Advanced algebraic concept for efficient multi-channel signal processing," 2012, doctoral Thesis.

[24] L. Liu, Y. Li, and J. Zhang, "DoA estimation and achievable rate analysis for 3D millimeter wave massive MIMO systems," in *IEEE SPAWC*, Jun. 2014, pp. 6–10.

Rubayet Shafin received his B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2013. Currently, he is working toward the Ph.D. degree in Electrical Engineering at the University of Kansas, USA. His research interests include MIMO systems, cellular networks, and mmWave Communications.

Lingjia Liu received his B.S. degree in Electronic Engineering from Shanghai Jiao Tong University and Ph.D. degree in Electrical and Computer Engineering from Texas A&M University. Prior to joining the faculty in the EECS at KU, he spent more than three years at Samsung Research America where he received the Global Samsung Best Paper Award in 2008 and 2010. He was leading Samsung's efforts on downlink multiuser MIMO, coordinated multipoint (CoMP), and heterogeneous networks (HetNet) in LTE/LTE-Advanced standards.

Dr. Liu is currently an Editor for the *IEEE Trans. Wireless Commun.*, an Editor for the *IEEE Trans. Commun.* and Associate Editor for the *EURASIP J. on Wireless Commun. and Netw.* and Wiley's *Intl J. on Commun. Systems*. He is co-editor of several special issues. His general research interests mainly lie in emerging technologies for 5G cellular networks including massive MIMO, massive MTC communications, and mmWave communications.

Jianzhong (Charlie) Zhang is a Vice President (VP) and the head of Standards and Mobility Innovation Lab with Samsung Research America, where he leads research and standards for 5G cellular systems and next generation multimedia networks. He received his Ph.D. degree from University of Wisconsin, Madison. From August 2009 to August 2013, he served as the Vice Chairman of the 3GPP RAN1 working group and led development of LTE and LTE-Advanced technologies such as 3D channel modeling, UL-MIMO and CoMP, Carrier Aggregation for TD-LTE, etc. Before joining Samsung, he was with Motorola from 2006 to 2007 working on 3GPP HSPA standards, and with Nokia Research Center from 2001 to 2006 working on *IEEE* 802.16e (WiMAX) standard and EDGE/CDMA receiver algorithms. Charlie is a Fellow of *IEEE*.

Yik-Chung Wu received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from the University of Hong Kong (HKU). He received the Croucher Foundation scholarship in 2002 to study Ph.D. degree at Texas A&M University, College Station, and graduated in 2005. From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with HKU, currently as an Associate Professor. He was a visiting scholar at Princeton University, in summers of 2011 and 2015. His research interests are in general areas of signal processing, machine learning and communication systems, and in particular distributed signal processing and robust optimization theories with applications to communication systems and smart grid. Dr. Wu served as an Editor for *IEEE* Communications Letters, is currently an Editor for *IEEE* Transactions on Communications and Journal of Communications and Networks.