Angle Domain Channel Estimation in Hybrid MmWave Massive MIMO Systems

Dian Fan, Feifei Gao, *Senior Member, IEEE*, Yuanwei Liu, *Member, IEEE*, Yansha Deng *Member, IEEE*, Gongpu Wang, Zhangdui Zhong, *Senior Member, IEEE*, Arumugam Nallanathan, *Fellow, IEEE*

***Abstract*—This paper proposes a novel direction of arrival (DOA)-aided channel estimation for hybrid millimeter wave (mmWave) massive MIMO system with the uniform planar array (UPA) at base station (BS). To explore the physical characteristics of antenna array in mmWave systems, the parameters of each channel path are decomposed into the DOA information and the channel gain information. We first estimate the initial DOAs of each uplink path through the two dimensional discrete Fourier transform (2D-DFT), and enhance the estimation accuracy via the angle rotation technique. We then estimate the channel gain information using small amount of training resources, which significantly reduces the training overhead and the feedback cost. More importantly, to examine the estimation performance, we derive the theoretical bounds of the mean squared errors (MSEs) and the Cramér-Rao Lower bounds (CRLBs) of the joint DOA and channel gain estimation. Simulation results show that the performances of the proposed methods are close to the theoretical MSEs analysis. Furthermore, the theoretical MSEs are also close to the corresponding CRLBs.**

***Index Terms*—Millimeter Wave, Massive MIMO, DOA Estimation, Channel Estimation, CRLB.**

I. Introduction

As an important candidate in the fifth generation (5G) mobile communications, the millimeter-wave (mmWave) communication that explores large amount of bandwidth resources at frequencies 30GHz to 300GHz has been proposed for outdoor cellular systems [1]–[4]. The mmWave communication requires massive antennas to overcome high propagation path loss and to provide beamforming power gain. Meanwhile, for a given size of antenna array, it is possible to equip hundreds or thousands of antennas at the transceiver due to the small carrier wavelengths at millimeter wave frequencies.

Full digital baseband precoding introduces extremely high hardware cost and energy con- sumption in massive multiple-input multiple-output (MIMO) system, due to the requirement for the same number of radio frequency (RF) chains [5]. Alternatively, the hybrid precoding which divides the precoding operations between the analog and digital domains can be low-cost solution to reduce the number of RF chains. In this architecture, the digital beamforming is conducted by controlling the digital weights associated with each RF chain. The analog beamforming is realized by controlling the phase of the signal transmitted at each antenna via a network of analog phase shifters. By doing so, hybrid analog–digital beamforming facilitates the hardware-constrained mmWave massive MIMO communication system to exploit both spatial diversity and multiplexing gain [6]–[12].

It is recognized that the full benefits of massive MIMO techniques in mmWave communication systems, such as high energy efficiency and high spectrum efficiency, heavily rely on the accurate channel state information (CSI) estimation, which is also regarded as one of the main challenges for massive MIMO system as well as mmWave systems. To the best of our knowledge, traditional channel estimation techniques developed for lower-frequency MIMO system are no longer applicable for mmWave massive MIMO system due to the implementation of large antenna arrays, hybrid precoding, and the sparsity of mmWave channel [13]. Thus, specific channel estimation techniques for mmWave and massive MIMO system have been proposed in [13]–[30].

In [14]–[16], the eigen-decomposition based algorithms exploiting the availability of low rank channel covariance matrices were developed for channel estimation. Unfortunately, the complexity of these channel estimation algorithms are extremely high and requires large overhead to obtain reliable channel covariance matrices. To solve this problem, [17]–[24] proposed the compressive sensing (CS) based channel estimation schemes by utilizing the sparsity of mmWave channel in the angle domain and incorporating the hybrid architecture. However, the complexity is still high due to the non-linear optimization, and its effectiveness highly depends on the restricted isometry property (RIP).

The authors in [25], [26] proposed an open-loop channel estimation strategy, which is in- dependent of the hardware constraints. In [27], a grid-of-beams (GoB) based approach

Manuscript received March 31, 2018; revised June 12, 2018 and August 28, 2018; accepted September 24, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant {61771274, 61831013, 61571037}, by EPSRC research grant EP/M016145/2, by Key Laboratory of Universal Wireless Communications (BUPT), Ministry of Education, P.R.China (No. KFKT-2018104), by NFSC Outstanding Youth under Grant 61725101, by National key research and development program under Grant 2016YFE0200900, and by Major projects of Beijing Municipal Science and Technology Commission under Grant No. Z181100003218010.

Dian Fan, Gongpu Wang and Zhangdui Zhong are with School of Computer and Information Technology, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, P. R. China. Dian Fan is currently a visiting research scholar in Queen Mary University of London, United Kingdom, (e-mail: fandian@bjtu.edu.cn, gpwang@bjtu.edu.cn, and zhdzhong@bjtu.edu.cn).

F. Gao is with Institute for Artificial Intelligence, Tsinghua University (THUAI), State Key Lab of Intelligent Technologies and SystemsTsinghua University, Beijing National Research Center for Information Science and TechnologyBNRist), Department of Automation, Tsinghua University, Beijing, P.R.China (email: feifeigao@ieee.org).

Yuanwei Liu and Arumugam Nallanathan are with School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom (e-mail: yuanwei.liu@qmul.ac.uk and a.nallanathan@qmul.ac.uk).

Yansha Deng is with Department of Informatics, King's College London, United Kingdom (e-mail: yansha.deng@kcl.ac.uk).

was proposed to obtain the angle-of-departure (AOD) and the direction-of-arrival (DOA). This approach requires large amount of training with high overhead because the best combinations of analog transmit and receive beams were achieved via exhaustive sequential search. In [28], compressed measurements on the mmWave channel were applied to estimate the second order statistics of the channel enabled adaptive hybrid precoding. It is noted that only quantized angle estimation with limited resolution can be achieved via the CS and grid-of-beams based methods. Recently, an array signal processing aided channel estimation scheme has been proposed in [29], [30], where the angle information of the user is exploited to simplify the channel estimation.

Angle information plays a very important role in the mmWave massive systems. Hence, there is an urgent requirement for a fast and accurate estimation approach that could efficiently estimate the angle information, especially for mmWave massive MIMO with hybrid precoding. Many high resolution subspace based angle estimation algorithms, such as multiple signal classification (MUSIC), estimation of signal parameters via rotational invariance technique (ESPRIT) and their variants have attracted enormous interests inside the array processing community due to their high resolution angle estimation [31]–[33]. Their applications in massive MIMO systems and full-dimension MIMO systems for two-dimensional angles estimation have been extensively studied in [34]–[38]. However, the conventional MUSIC and ESPRIT are not suitable for the mmWave communications due to the following main reasons: 1) They are of very high computational complexity during singular value decomposition (SVD) operation due to the massive number of antennas; 2) They belong to blind estimation category, which is originally designed for Radar application, and did not make full use of the training sequence in wireless communication systems.

In this work, we focus on the DOA estimation and channel estimation for the mmWave massive MIMO system with hybrid precoding. We first formulate the uplink channel model in hybrid precoding system, where the base station (BS) is equipped with an MN-antenna uniform planar array (UPA) while all users have single antenna. The parameters of each path in the channel matrix are decomposed into the corresponding channel gain and the DOA information. Using the two dimensional discrete Fourier transform (2D-DFT), the initial DOAs can be estimated through two dimensional fast Fourier transform (2D-FFT) with a resolution inversely proportional to MN, and its resolution can be further enhanced via angle rotation technique. Our proposed 2D-DFT based estimation method is of quite low complexity and is easy for practical implementation. Moreover, the obtained DOA estimation results are used for the subsequent channel gain estimation. Most importantly, we further derive a simple expression for the theoretical bounds of mean squared errors (MSEs) performance in high signal-to-noise ratio (SNR) region, as well as the corresponding Cramér-Rao lower bounds (CRLBs). Both theoretical and numerical results are provided to corroborate the effectiveness of the proposed method.

The rest of the paper is organized as follows. In section II, the system model of mmWave massive MIMO system with

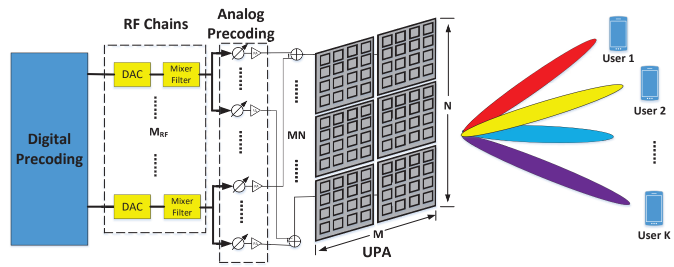


Fig. 1. Simplified system model of multiuser mmWave massive MIMO with hybrid precoding.

hybrid precoding and the channel characteristics are described. In section III, we present a two-stage 2D-DFT aided DOA estimation algorithm. The MSE and CRLB performance are analyzed in the section IV. Simulation results are then provided in Section V and conclusions are drawn in Section VI.

Notations: Small and upper bold-face letters donate column vectors and matrices, respectively; the superscripts $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^{-1}$, $(\cdot)^\dagger$ stand for the conjugate-transpose, transpose, conjugate, inverse, pseudo-inverse of a matrix, respectively; $tr(\mathbf{A})$ donates the trace of $\mathbf{A}$; $[\mathbf{A}]_{ij}$ is the (i,j)th entry of $\mathbf{A}$; Diag$\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal element constructed from $\mathbf{a}$, while Diag$\{\mathbf{A}\}$ denotes a vector whose elements are extracted from the diagonal components of $\mathbf{A}$; vec$(\mathbf{A})$ denotes the vectorization of $\mathbf{A}$; $\mathcal{R}\{\mathbf{A}\}$ denotes the real part of $\mathbf{A}$; $\mathcal{S}\{\mathbf{A}\}$ denotes the Imaginary part of $\mathbf{A}$; $[\mathbf{a}]_{i:j}$ denotes the subvector of $\mathbf{a}$ that starts with $[\mathbf{a}]_i$ and ends at $[\mathbf{a}]_j$; $[\mathbf{A}]_{i:j}$ denotes the submatrix of A that starts with row $[\mathbf{a}]_{i,:}$ and ends at row $[\mathbf{a}]_{j,:}$; $\mathbb{E}\{\cdot\}$ denotes the statistical expectation, and $\|\mathbf{h}\|$ is the Euclidean norm of $\mathbf{h}$.

II. System Model

Let us consider a multiuser mmWave massive MIMO time division duplex (TDD) systems with a hybrid precoding structure as shown in Fig. 1. The BS is equipped with MN antennas in the form of UPA where M represents the number of antennas in the horizon and N represents the number of antennas in the vertical. The BS has $M_{RF} \leq M \times N$ RF chains transmitting data streams to $K \leq M_{RF}$ mobile users, each with a single antenna [19]. We denote the distance between the neighboring antenna elements in both horizon and vertical as d. The BS is assumed to apply an $M_{RF} \times K$ complex valued based-band digital beamformer $\mathbf{F}_{BB}(\mathbf{F}_{BB} \in \mathbb{C}^{M_{RF}\times K})$, followed by an analog beamformer $\mathbf{F}_{RF}$ $(\mathbf{F}_{RF} \in \mathbb{C}^{MN \times M_{RF}})$. To simplify the hardware implementation, each element of $\mathbf{F}_{RF}$ has unitary magnitude with arbitrary phase. As $\mathbf{F}_{RF}$ is implemented using analog phase shifters, its elements are constrained to satisfy $[[\mathbf{F}_{RF}]_{:,j}[\mathbf{F}_{RF}]^*_{:,j}]_{i,i} = \frac{1}{MN}$, $(j = 1, 2, \cdots M_{RF}$, $i = 1, 2, \cdots MN)$, where all elements of $\mathbf{F}_{RF}$ have equal norm. The total transmit power constraint is enforced by normalizing $\mathbf{F}_{BB}$ to satisfy $\|\mathbf{F}_{RF}[\mathbf{F}_{BB}]_{:,k}\|_F^2 = 1$, $k = 1, 2, \cdots, K$.

A. Transmitter Model

Denote $\mathbf{H}_k$ as the $M \times N$ channel matrix between the BS and the kth mobile user. In the uplink transmission stage, the

$$\mathbf{A}(\phi_{l,k},\theta_{l,k}) = \frac{1}{\sqrt{MN}} \begin{bmatrix} 1 & \cdots & e^{j2\pi d/\lambda(N-1)\sin\phi_{l,k}\cos\theta_{l,k}} \\ \vdots & \ddots & \vdots \\ e^{j2\pi d/\lambda(M-1)\cos\phi_{l,k}} & \cdots & e^{j2\pi d/\lambda((M-1)\cos\phi_{l,k}+(N-1)\sin\phi_{l,k}\cos\theta_{l,k})} \end{bmatrix}_{M\times N}, \tag{4}$$

received signal at the BS can be expressed as

$$\mathbf{y}(t) = \mathbf{F}_{BB}^H\mathbf{F}_{RF}^H\sum_{k=1}^{K}\text{vec}\{\mathbf{H}_k\}s_k(t) + \mathbf{N}, t = 1,\cdots,T, \tag{1}$$

where $s_k(t)$ is the transmitted signal at time t, $\mathbf{N} \sim \mathcal{CN}(\mathbf{0},\sigma_n^2\mathbf{I})$ is the complex Gaussian noise matrix, and σ_n^2 is the noise covariance.

Since the uplink and downlink channels are reciprocal in a TDD system, the received signal in the downlink transmission at the kth mobile user is given by

$$y_k = \text{vec}^H\{\mathbf{H}_k\}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + n_k, \tag{2}$$

where $\mathbf{s} = [s_1, s_2, \cdots, s_K]^T$ is the transmitted signal vector for all K mobile users. Thus, we can express the received SNR at the kth mobile user as

$$\Gamma_k = |\text{vec}^H\{\mathbf{H}_k\}\mathbf{F}_{RF}\mathbf{F}_{BB}|^2\frac{\sigma_s^2}{\sigma_n^2}, \tag{3}$$

where $\text{E}\{|\mathbf{s}_k|^2\} = \sigma_s^2$ denotes the power of $\mathbf{s}_k$.

B. Channel Model

Due to the limited scattering characteristics in the mmWave environment [39]–[42], we assume the channel representation based on the extended Saleh-Valenzuela (SV) model in [25]–[27]. Let us define $\phi_{l,k} \in [-90^\circ, 90^\circ)$ and $\theta_{l,k} \in [-180^\circ, 180^\circ)$ as the signal elevation angle and the azimuth of the lth ($l = 1, 2, \cdots, L$) path of the kth user. The corresponding steering matrix can be expressed as (4) shown on top of this page, where λ is the wavelength of the carrier signal. Denoting $w_{1,l,k} = \frac{2\pi d}{\lambda}\cos\phi_{l,k}$ and $w_{2,l,k} = \frac{2\pi d}{\lambda}\sin\phi_{l,k}\cos\theta_{l,k}$, we can express (4) as

$$\mathbf{A}(\phi_{l,k},\theta_{l,k}) = \mathbf{A}(w_{1,l,k}, w_{2,l,k}) = \mathbf{a}(w_{1,l,k})\mathbf{a}^T(w_{2,l,k}), \tag{5}$$

where $\mathbf{a}(w_{1,l,k}) = \frac{1}{\sqrt{M}}[1,\cdots,e^{j(M-1)w_{1,l,k}}]^T$, and $\mathbf{a}(w_{2,l,k}) = \frac{1}{\sqrt{N}}[1,\cdots,e^{j(N-1)w_{2,l,k}}]^T$.

Using the geometric channel model with L scatters in mmWave channel, where each scatter contributes to single propagation path between the BS and the mobile user, we can write the channel matrix as

$$\begin{aligned}\mathbf{H}_k =& \frac{1}{\sqrt{L}}\sum_{l=1}^{L}a_{l,k}\mathbf{A}(\phi_{l,k},\theta_{l,k}) \\ =& \frac{1}{\sqrt{L}}\sum_{l=1}^{L}a_{l,k}\mathbf{a}(w_{1,l,k})\mathbf{a}^T(w_{2,l,k}) \\ =& \frac{1}{\sqrt{L}}\mathbf{A}_{w_1,k}\mathbf{H}_{a,k}\mathbf{A}_{w_2,k}^T,\end{aligned} \tag{6}$$

where

$$\begin{aligned}\mathbf{A}_{w_1,k} =& \; [\mathbf{a}(w_{1,1,k}), \mathbf{a}(w_{1,2,k}), \cdots, \mathbf{a}(w_{1,L,k})], \\ \mathbf{A}_{w_2,k} =& \; [\mathbf{a}(w_{2,1,k}), \mathbf{a}(w_{2,2,k}), \cdots, \mathbf{a}(w_{2,L,k})], \\ \mathbf{H}_{a,k} =& \; \text{diag}(a_{1,k}, a_{2,k}, \cdots, a_{L,k}),\end{aligned} \tag{7}$$

and $a_{l,k}$ is the channel gain along the lth path of the kth user ($l = 0$ for the line-of-sight (LOS) path and $l > 1$ for the non-line-of-sight (NLOS) paths).

The (m, n)th element of the channel matrix $\mathbf{H}_k$ can be written as

$$[\mathbf{H}_k]_{m,n} = \frac{1}{\sqrt{L}}\sum_{l=1}^{L}a_{l,k}e^{j(mw_{1,l,k}+nw_{2,l,k})}, \tag{8}$$

with $m = 0, 1, \cdots, M-1$, and $n = 0, 1, \cdots, N-1$. It is worth noting that at mmWave frequencies, the amplitude of channel gain $|a_{1,k}|$ of LOS components are typically 5 to 10dB stronger than the $\{|a_{l,k}|\}_{l=2}^{L}$ of the NLOS component [40].

Obviously, (6) is a sparse channel model that represents the low rank property and the spatial correlation characteristics of mmWave massive MIMO system. Importantly, the parameters of $\mathbf{H}_k$ have only L complex channel gains and $2L$ real phases $(\phi_{l,k},\theta_{l,k})$, where the number of paths is usually much less than the number of antennas, i.e., $L \ll MN$. Instead of directly estimating the channel $\mathbf{H}_k$, one could first estimate the DOA information $(\phi_{l,k},\theta_{l,k})$, and then estimate the corresponding path gain $a_{l,k}$ via the conventional estimation theory, such as least square (LS), maximum-likelihood (ML) algorithms. By doing this, the number of the parameters to be estimated is greatly reduced [43]. It is noted that the beamspace method and CS method from channel model (6) is not the real physical angle, but only provide an approximation of the quantized angle with limited resolution.

III. DOA Estimation from Array Signal Processing

In this section, we propose a new DOA estimation algorithm for the hybrid antenna array. To facilitate the understanding, we start with the uplink transmission.

A. Preamble

In the uplink transmission, the preamble will only be sent once at the beginning of the transmission. The received signal at the BS can be written as

$$\mathbf{Y}_{BB} = \mathbf{F}_{RF}^H\sum_{k=1}^{K}\text{vec}\{\mathbf{H}_k\}\mathbf{x}_k^T + \mathbf{N}, \tag{9}$$

where $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \cdots, x_{k,\tau}]^T$ is the preamble of the kth user, $\tau \geq K$ is the length of preamble, and $\mathbf{Y}_{BB}$ is the baseband signal before the digital precoding in the BS. Since the received signal has only a few observations, the whole CSI of the kth user cannot be extracted from the received signal directly.

To simplify the illustration, we assume that MN is divisible by M_{RF}, and denote $D = MN/M_{RF}$ as a suitable integer parameter that is related with the length of the RF chains and the number of antennas at the BS. Stacking the $M_{RF} \times 1$ subvectors into the channel vector vec$\{\tilde{\mathbf{H}}_k\}$ as

$$\text{vec}\{\mathbf{H}_k\} = [\text{vec}^H\{\mathbf{H}_k\}^{(0)}, \cdots, \text{vec}^H\{\mathbf{H}_k\}^{(D-1)}]^H, \quad (10)$$

where

$$\text{vec}\{\mathbf{H}_k\}^{(p)} = \text{vec}\{\mathbf{H}_k\}_{(pM_{RF}):((p+1)M_{RF}-1)}. \quad (11)$$

We design the analog receive beamformer $\mathbf{F}_{RF}(p)$ to switch on only the $(pM_{RF})th$, $\cdots$, $((p+1)M_{RF}-1)th$ receive antennas, as

$$\mathbf{F}_{RF}(p) = \begin{bmatrix} \mathbf{0}_{(pM_{RF})\times M_{RF}} \\ \mathbf{U}_{RF} \\ \mathbf{0}_{(MN-(p+1)M_{RF})\times M_{RF}} \end{bmatrix}_{MN\times M_{RF}}, \quad (12)$$

where $\mathbf{U}_{RF}$ is an $M_{RF} \times M_{RF}$ Hadamard matrix, and its elements are of unitary magnitude. Based on (9), the baseband signal of the kth user before digital precoding at the pth position and the tth time is written as

$$\begin{aligned} \mathbf{y}_{BB,k}(p,t) =& (\mathbf{F}_{RF}(p,t))^H \text{vec}\{\mathbf{H}_k\} x_k(t) + \mathbf{n}(p,t) \\ =& \mathbf{U}_{RF}^H \text{vec}\{\mathbf{H}_k\}^{(p)} x_k(t) + \mathbf{n}_k(p,t), \end{aligned} \quad (13)$$

where $\mathbf{n}_k(p,t) = [\mathbf{n}_k(t)]_{(pM_{RF}):((p+1)M_{RF}-1)}$, $t = 1, 2, \cdots, T$, and $p = 0, 1, \cdots D-1$.

We can then estimate vec$\{\mathbf{H}_k\}^{(p)}$ from (13) via

$$\begin{aligned} \text{vec}\{\hat{\mathbf{H}}_k\}^{(p)} =& (\mathbf{U}_{RF}^H)^{-1} \mathbf{y}_{BB,k}(p,t) \frac{x_k(t)}{\|x_k(t)\|^2} \\ =& \text{vec}\{\mathbf{H}_k\}^{(p)} + \mathbf{n}_k(p,t) \frac{x_k(t)}{\|x_k(t)\|^2}. \end{aligned} \quad (14)$$

If we probe all values of p in (14), we can obtain an estimate of $\mathbf{H}_k$. Similar to (10) and (12), we define the equivalent analog beamforming matrix as

$$\tilde{\mathbf{F}}_{RF}^H = \begin{bmatrix} \mathbf{F^1}_{RF}^H & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F^2}_{RF}^H & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F^D}_{RF}^H \end{bmatrix}_{MN\times MN}, \quad (15)$$

where $\mathbf{F^q}_{RF}$ is also an $M_{RF} \times M_{RF}$ Hadamard matrix such that it can guarantee the full rank of matrix $\tilde{\mathbf{F}}_{RF}$. By doing so, we can rewrite the equivalent baseband signal of the kth user before digital precoding at the BS as

$$\tilde{\mathbf{Y}}_{BB,k} = \tilde{\mathbf{F}}_{RF}^H \text{vec}\{\mathbf{H}_k\} \tilde{\mathbf{x}}_k^T + \tilde{\mathbf{N}}_k, \quad (16)$$

where $\tilde{\mathbf{x}}_k = [x_k(1), x_k(2), \cdots, x_k(D)]^H$, and

$$\tilde{\mathbf{N}}_k = \begin{bmatrix} \mathbf{n}_k(1,1) & \mathbf{n}_k(1,2) & \cdots & \mathbf{n}_k(1,D) \\ \mathbf{n}_k(2,1) & \mathbf{n}_k(2,2) & \cdots & \mathbf{n}_k(2,D) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{n}_k(D,1) & \mathbf{n}_k(D,2) & \cdots & \mathbf{n}_k(D,D) \end{bmatrix}. \quad (17)$$

Hence, the LS estimation of the channel can be expressed as

$$\text{vec}\{\hat{\mathbf{H}}_k\} = (\tilde{\mathbf{F}}_{RF}^H)^{-1} \mathbf{Y}_{BB,k} \frac{\tilde{\mathbf{x}}_k^T}{\|\tilde{\mathbf{x}}_k^T\|^2} + (\tilde{\mathbf{F}}_{RF}^H)^{-1} \tilde{\mathbf{N}}_k \frac{\tilde{\mathbf{x}}_k^T}{\|\tilde{\mathbf{x}}_k^T\|^2}. \quad (18)$$

During practical transmission, the signals of K users are overlay together. Therefore, we must use orthogonal pilot sequences to distinguish each user. In order to reduce the pilot overhead, we assume that only $\tau = K$ orthogonal pilots can be used to estimate vec$\{\mathbf{H}_k\}$. However, it is not enough to estimate the complete channel. Define $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K]^T$ as the orthogonal training matrix, and $\|\mathbf{x}_k\|^2 = 1$. We can send the same orthogonal training matrix for D times to estimate the whole channel matrix. Similar to (16), we have

$$\tilde{\mathbf{Y}}_{BB} = \tilde{\mathbf{F}}_{RF}^H \mathbf{H} \tilde{\mathbf{X}} + \tilde{\mathbf{N}}, \quad (19)$$

where

$$\begin{aligned} \mathbf{H} =& [\text{vec}\{\mathbf{H}_1\}, \text{vec}\{\mathbf{H}_2\}, \cdots, \text{vec}\{\mathbf{H}_K\}], \\ \tilde{\mathbf{X}} =& [\underbrace{\mathbf{X}, \mathbf{X}, \cdots, \mathbf{X}}_{D}], \quad \tilde{\mathbf{N}} = \sum_{k=1}^{K} [\underbrace{\tilde{\mathbf{N}}_k, \tilde{\mathbf{N}}_k, \cdots, \tilde{\mathbf{N}}_k}_{D}]. \end{aligned} \quad (20)$$

Thus, the channel for all users can be estimated from

$$\hat{\mathbf{H}} = (\tilde{\mathbf{F}}_{RF}^H)^{-1} \tilde{\mathbf{Y}}_{BB} \tilde{\mathbf{X}}^T + (\tilde{\mathbf{F}}_{RF}^H)^{-1} \tilde{\mathbf{N}} \tilde{\mathbf{X}}^T. \quad (21)$$

Note that the preamble process seems time consuming but will only be performed once at the start of the transmission. Usually, the transmitter and the receiver may not physically change its position in a relatively longer time, thus we can treat the DOA component of the channel as unchanged within several or even tens of the channel coherence times [29], while the remaining channel gain component could be re-estimated via much simplified approach.

After obtaining the initial channel estimation for all users, the next step is to extract the angular information $(\phi_{l,k}, \theta_{l,k})$ via the 2D-DFT and angular rotation approaches for each user, which will be described in the next subsection, and we omit k for simplicity.

B. DOA Estimation Algorithm

Thanks to the massive number of antennas at the BS as well as the UPA structure, we propose an efficient 2D-DFT approach for DOA estimation in the following.

1) Initial DOA Estimation: We first define two normalized DFT matrix $\mathbf{F}_M$ and $\mathbf{F}_N$, whose elements are given by $[\mathbf{F}_M]_{pp'} = \frac{1}{\sqrt{M}} e^{-j\frac{2\pi}{M}pp'}$, $p, p' = 0, 1, \cdots, M-1$ and $[\mathbf{F}_N]_{qq'} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}qq'}$, $q, q' = 0, 1, \cdots, N-1$, respectively. Meanwhile, we define the normalized 2D-DFT of the channel matrix $\mathbf{H}$ as $\mathbf{H}_{DFT} = \mathbf{F}_M \mathbf{H} \mathbf{F}_N$, whose (p,q)th element

$(p = 0, 1, ..., M-1; q = 0, 1, ..., N-1)$ is computed as

$$\begin{aligned}[\mathbf{H}_{DFT}]_{pq} &= \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [\mathbf{H}]_{pq} e^{-j2\pi\left(\frac{pm}{M}+\frac{qn}{N}\right)} \\ &= \frac{1}{\sqrt{LMN}} \sum_{l=1}^{L} a_l e^{-j\frac{M-1}{2}\left(\frac{2\pi}{M}p - w_{1,l}\right)} e^{-j\frac{N-1}{2}\left(\frac{2\pi}{N}q - w_{2,l}\right)} \\ &\times \frac{\sin\left(\pi p - \frac{Mw_{1,l}}{2}\right)}{\sin((\pi p - \frac{Mw_{1,l}}{2})/M)} \cdot \frac{\sin\left(\pi q - \frac{Nw_{2,l}}{2}\right)}{\sin((\pi q - \frac{Nw_{2,l}}{2})/N)}. \end{aligned} \tag{22}$$

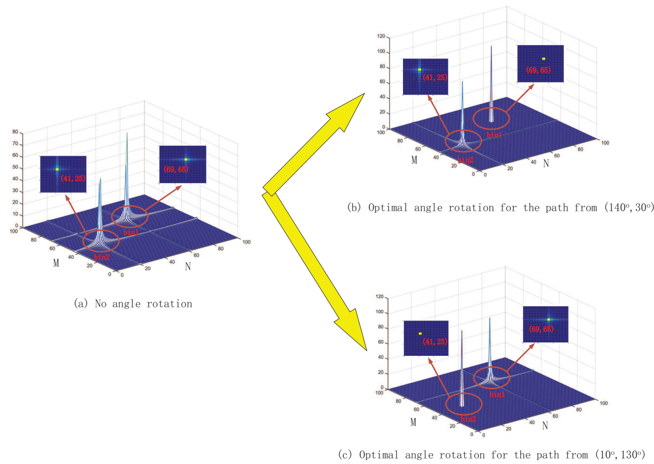
It is noted that with infinite number of antennas in the array, i.e., $M \to \infty$, $N \to \infty$, there always exists some integers $p_l = \frac{Mw_{1,l}}{2\pi}$, $q_l = \frac{Nw_{2,l}}{2\pi}$ such that $[\mathbf{H}_{DFT}]_{p_l q_l} = \frac{a_l}{\sqrt{LMN}}$, while the other elements of $\mathbf{H}_{DFT}$ are all zero, In other words, all power is concentrated on the (p_l, q_l)th elements and the elements of $\mathbf{H}_{DFT}$ possess sparse property, such that the elevation and the azimuth DOA of the lth path (ϕ_l, θ_l) can be immediately estimated from the non-zero positions (p_l, q_l) of $\mathbf{H}_{DFT}$ using

$$\begin{aligned}\phi_l &= \cos^{-1}\left(\frac{\lambda p_l}{Md}\right), \\ \theta_l &= \cos^{-1}\left(\frac{\lambda q_l}{Nd}\Big/\sqrt{1-\left(\frac{\lambda p_l}{Md}\right)^2}\right). \end{aligned} \tag{23}$$

Unfortunately, in practice, the array aperture cannot be infinitely large, even if MN could be as greater as hundreds or thousands in hybrid mmWave massive MIMO systems. In special case, if some specific angles satisfy that $Mw_{1,l}/(2\pi)$ is integer and $Nw_{2,l}/(2\pi)$ is integer, all power of channel will concentrate on some separated single points. We call these as on-grid angles. In more general case, $Mw_{1,l}/(2\pi)$ and $Nw_{2,l}/(2\pi)$ will not be integers, and the channel power of $\mathbf{H}_{DFT}$ will leak from the $(\lfloor Mw_{1,l}/(2\pi)\rceil, \lfloor Nw_{2,l}/(2\pi)\rceil)$th element to its nearby elements. In fact, the leakage of channel power is positively proportional to the deviation $(Mw_{1,l}/(2\pi) - \lfloor Mw_{1,l}/(2\pi)\rceil)$ and $(Nw_{2,l}/(2\pi) - \lfloor Nw_{2,l}/(2\pi)\rceil)$, but is inversely proportional to M and N as shown in (22). However, $\mathbf{H}_{DFT}$ can still be approximated as a sparse matrix with most of power concentrated around the $(\lfloor Mw_{1,l}/(2\pi)\rceil, \lfloor Nw_{2,l}/(2\pi)\rceil)$th element. Hence, the peak power position of $\mathbf{H}_{DFT}$ is still useful for extracting initial DOA information.

An example of a two-path channel from $(30°, 140°)$ and $(-50°, 10°)$ with $M = 100$, $N = 100$ as shown in Fig. 2(a), whose channel sparse characteristics after 2D-DFT is depicted. For clear illustration, we demonstrate only for a noise-free scenario. It can be seen that each path corresponds to one bin and each bin has a central point that contains the largest power. Each bin encounters the power leakage and the points around the central point also contain considerable power but the power of other points are ignorable. In Fig. 2(a), the central point of the channel after initial 2D-DFT are $(69, 65)$ and $(41, 25)$. Hence, we can use these two peak power positions as the initial DOA estimation.

Based on the above discussion, we can formulate the 2D-DFT of the estimated channel matrix $\hat{\mathbf{H}}$, with its (p,q)th element being



(a) No angle rotation

(b) Optimal angle rotation for the path from (140°, 30°)

(c) Optimal angle rotation for the path from (10°, 130°)

Fig. 2. An example of a two paths channel sparse characteristics after 2D-DFT and optimal angle rotation, where BS array has 100×100 antennas.

$$[\hat{\mathbf{H}}_{DFT}]_{pq} = [\mathbf{H}_{DFT}]_{pq} + [\mathbf{N}_{DFT}]_{pq}, \tag{24}$$

where $\mathbf{N}_{DFT} \sim \mathcal{CN}\left(0, \frac{\sigma_n^2}{\sqrt{MN}}\mathbf{I}\right)$. Denote the L largest peaks in L bins of $\hat{\mathbf{H}}_{DFT}$ as (p_l^{ini}, q_l^{ini}). We can express the initial DOA estimates as

$$\begin{aligned}\hat{\phi}_l^{ini} &= \cos^{-1}\left(\frac{\lambda p_l^{ini}}{Md}\right), \\ \hat{\theta}_l^{ini} &= \cos^{-1}\left(\frac{\lambda q_l^{ini}}{Nd}\Big/\sqrt{1-\left(\frac{\lambda p_l^{ini}}{Md}\right)^2}\right). \end{aligned} \tag{25}$$

2) Fine DOA Estimation: The resolution of $(\hat{\phi}_l^{ini}, \hat{\theta}_l^{ini})$ via directly applying 2D-DFT is limited by half of the DFT interval, i.e., $1/(2M)$ and $1/(2N)$. For example, for $M = 100$ and $N = 100$, the worst MSE of the $(\hat{\phi}_l^{ini}, \hat{\theta}_l^{ini})$ is in the order of 10^{-4}. To improve the DOA estimation accuracy, we next show how this mismatch could be compensated via an angle rotation operation.

The angle rotation of the original channel matrix is defined as

$$\mathbf{H}^{ro} = \mathbf{\Phi}_M(\Delta\phi_l)\mathbf{H}\mathbf{\Phi}_N(\Delta\theta_l), \tag{26}$$

where the diagonal matrices $\mathbf{\Phi}_M(\Delta\phi_l)$ and $\mathbf{\Phi}_N(\Delta\theta_l)$ are given by

$$\begin{aligned}\mathbf{\Phi}_M(\Delta\phi_l) &= \mathrm{diag}\{1, e^{j\Delta\phi_l}, \cdots, e^{j(M-1)\Delta\phi_l}\}, \\ \mathbf{\Phi}_N(\Delta\theta_l) &= \mathrm{diag}\{1, e^{j\Delta\theta_l}, \cdots, e^{j(N-1)\Delta\theta_l}\}. \end{aligned} \tag{27}$$

In (27), $\Delta\phi_l \in [-\frac{\pi}{M}, \frac{\pi}{M}]$ and $\Delta\theta_l \in [-\frac{\pi}{N}, \frac{\pi}{N}]$ are the angle rotation parameters. After the angle rotation operation, the 2D-DFT of the rotated channel matrix $\mathbf{H}_{DFT}^{ro}$ can be calculated as (28) shown on top of next page.

It can be readily found that the entries of $\mathbf{H}_{DFT}^{ro}$ have only L non-zero elements when the angle shifts satisfying

$$\Delta\phi_l = 2\pi p_l/M - w_{1,l}, \qquad \Delta\theta_l = 2\pi q_l/N - w_{2,l}, \tag{29}$$

where $(\Delta\phi_l, \Delta\theta_l)$ in (29) are the optimal angle shifts.

$$[\mathbf{H}^{ro}_{DFT}]_{pq} = \frac{1}{\sqrt{LMN}} \sum_{l=1}^{L} a_l e^{-j\frac{M-1}{2}(\frac{2\pi}{M}p - w_{1,l} - \Delta\phi_l)} e^{-j\frac{N-1}{2}(\frac{2\pi}{N}q - w_{2,l} - \Delta\theta_l)} \times \frac{\sin\left(\pi p - \frac{Mw_{1,l}}{2} - \frac{M\Delta\phi_l}{2}\right)}{\sin((\pi p - \frac{Mw_{1,l}}{2} - \frac{M\Delta\phi_l}{2})/M)} \cdot \frac{\sin\left(\pi q - \frac{Nw_{2,l}}{2} - \frac{N\Delta\theta_l}{2}\right)}{\sin((\pi q - \frac{Nw_{2,l}}{2} - \frac{N\Delta\theta_l}{2})/N)}. \tag{28}$$

Based on the derived optimal angle shifts, the elevation angle and the azimuth angle of the lth path (ϕ_l, θ_l) can be estimated as

$$\hat{\phi}_l = \cos^{-1}\left(\frac{\lambda p_l}{Md} - \frac{\lambda\Delta\phi_l}{2\pi d}\right), \tag{30}$$
$$\hat{\theta}_l = \cos^{-1}\left(\left(\frac{\lambda q_l}{Nd} - \frac{\lambda\Delta\theta_l}{2\pi d}\right)\Big/\sqrt{1 - \left(\frac{\lambda p_l}{Md} - \frac{\lambda\Delta\phi_l}{2\pi d}\right)^2}\right).$$

For finding the optimal angle shifts $(\Delta\phi_l, \Delta\theta_l)$ from channel matrix $\mathbf{H}$, one way is the simple two-dimensional searching of $\Delta\phi$ and $\Delta\theta$ over the very small region $\Delta\phi \in [-\frac{\pi}{M}, \frac{\pi}{M}]$ and $\Delta\theta \in [-\frac{\pi}{N}, \frac{\pi}{N}]$. Then we can extract the corresponding $(\Delta\phi_l, \Delta\theta_l)$ when the (p_l^{ini}, q_l^{ini})th element of $\mathbf{F}_M \mathbf{\Phi}_M(\Delta\phi_l)\mathbf{H}\mathbf{\Phi}_N(\Delta\theta_l)\mathbf{F}_N$ shrink into their highest form. Mathematically, there is

$$(\Delta\phi_l, \Delta\theta_l) = \arg \max_{\Delta\phi \in [-\frac{\pi}{M}, \frac{\pi}{M}], \Delta\theta \in [-\frac{\pi}{N}, \frac{\pi}{N}]} \|\mathbf{f}^H_{M p_l^{ini}} \mathbf{\Phi}_M(\Delta\phi_l)\mathbf{H}\mathbf{\Phi}_N(\Delta\theta_l)\mathbf{f}_{N q_l^{ini}}\|^2, \tag{31}$$

where $\mathbf{f}_{M p_l^{ini}}$ is the p_l^{ini}th column of $\mathbf{F}_M$ and $\mathbf{f}_{N q_l^{ini}}$ is the q_l^{ini}th column of $\mathbf{F}_N$.

To demonstrate the effect of the angle rotation, we consider a two paths channel as an example shown in Fig. 2(b) and Fig. 2(c). Fig. 2(b) and Fig. 2(c) show the 2D-DFT spectrum of the channel matrix with the optimal angle rotation for the two paths respectively. Through angle rotation, the 2D-DFT spectrum becomes highly concentrated around the DOAs of two paths, which could improve the accuracy of the DOA estimation. After searching all $\Delta\phi$ and $\Delta\theta$, we can obtain the position of the maximal power and the optimal angle rotation $(\Delta\phi_l, \Delta\theta_l)$ for each path.

The DOA information of different paths can be estimated using the method outlined in **Algorithm 1**. Note that the number of search grids $G = G_M G_N$, where G_M means the search grids within $[-\frac{\pi}{M}, \frac{\pi}{M}]$ and G_N means the search grids within $[-\frac{\pi}{N}, \frac{\pi}{N}]$, determines the complexity and accuracy of the whole DOA estimation algorithm. It is easy to find that the accuracy of the DOA estimation is directly proportional to the number of searched grids, but the complexity of the algorithm is inversely proportional to the number of searched grids. Since the complexity of (31) is proportional to $O(MN)$ for the given $\Delta\phi_l$ and $\Delta\theta_l$, the complexity of the whole algorithm is about $O(MNlogMN + MN + GKMN)$. The complexity of different algorithms is shown in the Table I. Since $K \ll MN$ and $G \ll MN$[1], the complexity of the proposed algorithm is much less than $O(M^2N^2)$.

[1]For mmWave massive MIMO with very large MN, a small value of G is already good enough to provide very high accuracy and low complexity.

Algorithm 1 2D-DFT and angle rotation based DOA estimation

Input: $\mathbf{H}$. **Output**: $\phi_l, \theta_l, l = 1, 2, \cdots, L$.

1. Find the central point (p_l^{ini}, q_l^{ini}) of each bin in $\mathbf{H}_{DFT} = \mathbf{F}_M\mathbf{H}\mathbf{F}_N$, where $(p_l^{ini}, q_l^{ini}) = \arg\max_{(p,q)\in bin(l)} \|[\mathbf{H}_{DFT}]_{pq}\|^2$, $l = 1, 2, \cdots, L$.
2. **For** $\Delta\phi = -\frac{\pi}{M} : \frac{\pi}{M}$, $\Delta\theta_l = -\frac{\pi}{N} : \frac{\pi}{N}$
3. **For** $l = 1 : L$
4. $(\Delta\hat{\phi}_l, \Delta\hat{\theta}_l) = \arg\max_{\Delta\phi\in[-\frac{\pi}{M}, \frac{\pi}{M}], \Delta\theta\in[-\frac{\pi}{N}, \frac{\pi}{N}]} \|\mathbf{f}^H_{M p_l^{ini}} \mathbf{\Phi}_M(\Delta\phi_l)\mathbf{H}\mathbf{\Phi}_N(\Delta\theta_l)\mathbf{f}_{N q_l^{ini}}\|^2$,
5. **End For**
6. **End For**
7. Obtain $\hat{\phi}_l, \hat{\theta}_l$ from the equation (30).

Remark 1. *The DOA can be estimated with a resolution proportional to the number of antenna at the BS, and this resolution can be enhanced via angle rotation technique with fast Fourier transform (FFT). Note that different users with very close DOA can be recognized by our DOA estimation method, but it cannot be achieved by blind algorithms.*

C. Channel Gain Estimation and Hybrid Precoding

To estimate the uplink channel gains, the BS needs to know the estimated DOA parameters $(\Delta\hat{\phi}_{l,k}, \Delta\hat{\theta}_{l,k})$, $l = 1, 2, \cdots, L$, $k = 1, 2, \cdots, K$. The received signal of the BS can be written as

$$\begin{aligned}\mathbf{Y} &= \mathbf{F}^H_{BB}\mathbf{F}^H_{RF}\sum_{k=1}^{K} \text{vec}\{\mathbf{H}_k\}\mathbf{x}_k^T + \mathbf{N} \\ &= \frac{1}{\sqrt{L}}\mathbf{F}^H_{BB}\mathbf{F}^H_{RF}\sum_{k=1}^{K}\sum_{l=1}^{L} a_{l,k}\text{vec}\{\mathbf{A}(\hat{\phi}_{l,k}, \hat{\theta}_{l,k})\}\mathbf{x}_k^T + \mathbf{N} \\ &= \frac{1}{\sqrt{L}}\mathbf{F}^H_{BB}\mathbf{F}^H_{RF}\sum_{k=1}^{K}\mathbf{A}_k\mathbf{a}_k\mathbf{x}_k^T + \mathbf{N},\end{aligned} \tag{32}$$

where $\mathbf{A}_k = [\text{vec}\{\mathbf{A}(\hat{\phi}_{1,k}, \hat{\theta}_{1,k})\}, \cdots, \text{vec}\{\mathbf{A}(\hat{\phi}_{L,k}, \hat{\theta}_{L,k})\}]$, and $\mathbf{a}_k = [a_{1,k}, a_{2,k}, \cdots, a_{L,k}]^T$. Note that the digital beamforming matrix $\mathbf{F}_{BB}$, analog beamforming matrix $\mathbf{F}_{RF}$ and the steering matrix $\mathbf{A}_k$ are known at the BS. The BS can refine the channel gains via LS estimation as

$$\begin{aligned}\hat{\mathbf{a}}_k &= \sqrt{L}(\mathbf{F}^H_{BB}\mathbf{F}^H_{RF}\mathbf{A}_k)^\dagger\mathbf{Y}\mathbf{x}_k \\ &= \mathbf{a}_k + \sqrt{L}(\mathbf{F}^H_{BB}\mathbf{F}^H_{RF}\mathbf{A}_k)^\dagger\mathbf{N}\mathbf{x}_k.\end{aligned} \tag{33}$$

Thus, with the DOAs information from (30) and gains information from (33), we can obtain the uplink channel estimation

TABLE I
COMPLEXITY OF DIFFERENT ESTIMATE ALGORITHMS

Algorithm	Complexity
Proposed 2D-DFT and angle rotation estimate algorithm	$O(MNlogMN + MN + GKMN)$
Beamspace estimate algorithm	$O(MNlogMN + MN)$
CS estimate algorithm	$O(M^2N^2 + KMN)$

for all users as

$$\hat{\mathbf{H}}_k = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \hat{a}_{l,k} \mathrm{vec}\{\mathbf{A}(\hat{\phi}_{l,k}, \hat{\theta}_{l,k})\}. \tag{34}$$

From (2), the downlink received signals can be expressed as

$$\mathbf{y}^d = \mathbf{H}^d \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{s} + \mathbf{n}, \tag{35}$$

where $\mathbf{H}^d = [\mathrm{vec}\{\mathbf{H}_1\}, \mathrm{vec}\{\mathbf{H}_2\}, \cdots, \mathrm{vec}\{\mathbf{H}_K\}]^H$, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_K)$ is additive white Gaussian noise vector. We assume $M_{RF} = KL$. From the previous discussion, the analog precoding matrix can be immediately obtained from

$$\begin{aligned}\mathbf{F}_{RF} = [&\mathrm{vec}\{\mathbf{\Phi}_M(\Delta\phi_{1,1})\mathbf{f}_{M_{p_{1,1}}} \mathbf{f}_{N_{q_{1,1}}}^H \mathbf{\Phi}_N(\Delta\theta_{1,1})\}, \cdots, \\ &\mathrm{vec}\{\mathbf{\Phi}_M(\Delta\phi_{L,K})\mathbf{f}_{M_{p_{L,K}}} \mathbf{f}_{N_{q_{L,K}}}^H \mathbf{\Phi}_N(\Delta\theta_{L,K})\}],\end{aligned} \tag{36}$$

where $(p_{l,k}, q_{l,k})$ denotes the position that contains the largest power after 2D-DFT of the lth path of the kth user, and each column of $\mathbf{F}_{RF}$ represents the spatial angle (after angle rotation) of each path for all K users. Note that the analog precoding indicates that each path of each user is transmitting exactly towards its signal direction and is thus named as anglespace transmission[2].

Similar to the conventional digital precoding approach, $\mathbf{F}_{BB}$ can be obtained via the zero-forcing (ZF) beamforming algorithm, as

$$\mathbf{F}_{BB} = \frac{1}{\sqrt{P}} (\mathbf{H}^d \mathbf{F}_{RF})^H ((\mathbf{H}^d \mathbf{F}_{RF})(\mathbf{H}^d \mathbf{F}_{RF})^H)^{-1}, \tag{37}$$

where P is the power constraint.

Remark 2. *The reciprocity of the channel cannot be applied to the frequency division duplex (FDD) system due to the different transmission frequencies of the uplink and downlink channels. Nonetheless, the uplink and downlink channels share a common propagation space between the BS and the user. The spatial directions or angles in the uplink channel are almost the same as those in the downlink channel. For example, the DOA information of both uplink and downlink are the same. Therefore, our DOA estimation algorithm can also be applied to FDD system and the channel gain component of the downlink can be estimated using small training overhead.*

IV. PERFORMANCE ANALYSIS

In this section, we derive the theoretical MSE of the joint DOA and channel gain estimation for hybrid mmWave massive MIMO system. Generally, a closed-form MSE analysis for multiple DOA estimations is hard to obtain. An alternatively acceptable approach is to consider single user and single propagation path and derive corresponding MSE of ϕ, θ as a benchmark [44]. We first show that the MSE of the proposed estimation algorithm is the same as the ML estimator in single propagation path scenario and derive the closed-form expressions of the DOA information and channel gain using the ML estimator in the high SNR region. Next, the CRLB analysis of the DOA information and channel gain are carried on.

A. Theoretical MSE of The Proposed Estimator

Limiting to single propagation path, the received signal can be rewritten as

$$\mathbf{y} = \mathbf{F}_{RF}^H \mathrm{vec}(\mathbf{H}) s + \mathbf{n} = \mathbf{F}_{RF}^H \mathrm{vec}(\mathbf{A}) \alpha s + \mathbf{n}, \tag{38}$$

where $\mathbf{A} \triangleq \mathbf{A}(w_1, w_2)$ is the $M \times N$ steering matrix with its (p,q)th entry given by

$$[\mathbf{A}]_{pq} = e^{j((p-1)w_1 + (q-1)w_2)}, \tag{39}$$

and $\mathbf{n}$ is the $M_{RF} \times 1$ vector representing the white Gaussian noise with zero mean and variance σ_n^2.

The proposed estimator can be rewritten as

$$\begin{aligned}[\hat{w_1}, \hat{w_2}] &= \operatorname*{argmax}_{w_1, w_2} \|\mathrm{vec}^H(\mathbf{A})(\mathbf{F}_{RF}^H)^\dagger \mathbf{y}\|^2 \\ &= \operatorname*{argmax}_{w_1, w_2} \mathbf{y}^H \mathbf{F}_{RF}^H \mathrm{vec}(\mathbf{A}) \mathrm{vec}^H(\mathbf{A})(\mathbf{F}_{RF}^H)^\dagger \mathbf{y},\end{aligned} \tag{40}$$

where $\mathrm{vec}(\mathbf{A}) = \mathrm{vec}\{\mathbf{\Phi}_M(\Delta\phi)\mathbf{f}_{M_p} \mathbf{f}_{N_q}^H \mathbf{\Phi}_N(\Delta\theta)\}$.

For given w_1, w_2 and α, the probability density function (PDF) of $\mathbf{y}$ can be expressed as

$$\begin{aligned}&f(\mathbf{y}|w_1, w_2, \alpha) \\ =&\frac{1}{(\pi\sigma_n^2)^{M_{RF}}} \exp\left\{-\frac{\|\mathbf{y} - \mathbf{F}_{RF}^H \mathrm{vec}(\mathbf{A})\alpha s\|^2}{\sigma_n^2}\right\}.\end{aligned} \tag{41}$$

The joint ML estimates of w_1, w_2 and α can be obtained via

$$[\hat{w_1}, \hat{w_2}, \hat{\alpha}] = \operatorname*{argmax}_{w_1, w_2, \alpha} f(\mathbf{y}|w_1, w_2, \alpha), \tag{42}$$

or equivalently

$$[\hat{w_1}, \hat{w_2}, \hat{\alpha}] = \operatorname*{argmin}_{w_1, w_2, \alpha} \|\mathbf{y} - \mathbf{F}_{RF}^H \mathrm{vec}(\mathbf{A})\alpha s\|^2. \tag{43}$$

In the next analysis, we first estimate w_1 and w_2, and then estimate channel gain a, which is a two-step optimization rather than joint optimization.

For given w_1 and w_2, the ML estimate of α is obtained from (43) as

$$\hat{\alpha} = \mathrm{vec}^H(\mathbf{A})(\mathbf{F}_{RF}^H)^\dagger s^* \mathbf{y}. \tag{44}$$

[2] This is a key difference from beamspace transmission.

Substituting (44) into (43), the ML estimates of w_1, w_2 can be written as

$$\begin{aligned}[\hat{w_1}, \hat{w_2}] =& \operatorname*{argmin}_{w_1,w_2} \|\mathbf{y} - \mathbf{F}_{RF}^H \text{vec}(\mathbf{A})\text{vec}^H(\mathbf{A})(\mathbf{F}_{RF}^H)^\dagger s^* \mathbf{y} s\|^2 \\ =& \operatorname*{argmin}_{w_1,w_2} \|\mathbf{y} - \sigma_s^2 \mathbf{F}_{RF}^H \text{vec}(\mathbf{A})\text{vec}^H(\mathbf{A})(\mathbf{F}_{RF}^H)^\dagger \mathbf{y}\|^2 \\ =& \operatorname*{argmax}_{w_1,w_2} \mathbf{y}^H \mathbf{F}_{RF}^H \mathbf{P}_\mathbf{A} (\mathbf{F}_{RF}^H)^\dagger \mathbf{y} \\ =& \operatorname*{argmax}_{w_1,w_2} g(w_1, w_2),\end{aligned} \tag{45}$$

where $g(w_1, w_2)$ denotes the cost function of w_1, w_2, $s^* s = \sigma_s^2$ means the power of signal, and $\mathbf{P}_\mathbf{A} = \text{vec}(\mathbf{A})\text{vec}^H(\mathbf{A})$ represents the projection matrix onto the subspace spanned by $\text{vec}(\mathbf{A})$. Interestingly, the MSE (40) of the proposed estimator coincides with the ML estimator (45). Till now, we have (44) and (45) as the ML estimates of α, w_1, and w_2.

Lemma 1. *Under high SNR, the perturbations of the estimation of w_1 and w_2 from* (45) *are given by*

$$E\{\Delta w_i^2\} = \frac{\sigma_n^2}{2\sigma_s^2 |\alpha|^2 vec^H(\mathbf{A})\mathbf{W}_i \mathbf{P}_a^\perp \mathbf{W}_i vec(\mathbf{A})}, \tag{46}$$

where $i = 1, 2$, $\mathbf{P}_a^\perp = \mathbf{I} - \mathbf{P}_a$ is the projection matrix onto the orthogonal space spanned by $\mathbf{A}$, and $\mathbf{W}_1$, $\mathbf{W}_2$ are the diagonal matrices as

$$\mathbf{W}_1 = \mathit{Diag}\{\underbrace{0.\cdots,0}_{N}, \cdots, \underbrace{(M-1),\cdots,(M-1)}_{N}\}, \tag{47}$$

$$\mathbf{W}_2 = \mathit{Diag}\{\underbrace{0.\cdots,0}_{M}, \cdots, \underbrace{(N-1),\cdots,(N-1)}_{M}\}. \tag{48}$$

Proof. See Appendix. □

In the channel estimation process, we need to further examine the MSE of the azimuth angle ϕ and the elevation angle θ. Based on the fact that $w_1 = \frac{2\pi d}{\lambda}\cos\phi$, $w_2 = \frac{2\pi d}{\lambda}\sin\phi\cos\theta$, we have

$$\phi = \cos^{-1}\left(\frac{\lambda w_1}{2\pi d}\right), \quad \text{and} \quad \theta = \cos^{-1}\left(\frac{\lambda w_2}{2\pi d \sin\phi}\right). \tag{49}$$

From (46) and (49), we can derive the mean and the MSE of the azimuth angle and the elevation angle, namely

$$\begin{aligned}&\text{E}\{\Delta\phi\} = \text{E}\{\Delta\theta\} = 0, \\ &\text{E}\{\Delta\phi^2\} = \text{E}\{(\hat{\phi}-\phi)(\hat{\phi}-\phi)^H\} = \frac{\partial\phi}{\partial w_1}\text{E}\{w_1^2\}(\frac{\partial\phi}{\partial w_1})^H \\ &= \frac{(\frac{\lambda}{2\pi d})^2}{1-(\frac{\lambda w_1}{2\pi d})^2} \times \frac{\sigma_n^2/\sigma_s^2}{2|\alpha|^2 \text{vec}^H(\mathbf{A})\mathbf{W}_1\mathbf{P}_a^\perp\mathbf{W}_1\text{vec}(\mathbf{A})}, \\ &\text{E}\{\Delta\theta^2\} = \text{E}\{(\hat{\theta}-\theta)(\hat{\theta}-\theta)^H\} = \frac{\partial\theta}{\partial w_2}\text{E}\{w_2^2\}(\frac{\partial\theta}{\partial w_2})^H \\ &= \frac{(\frac{\lambda}{2\pi d\sin\phi})^2}{1-(\frac{\lambda w_2}{2\pi d\sin\phi})^2} \times \frac{\sigma_n^2/\sigma_s^2}{2|\alpha|^2 \text{vec}^H(\mathbf{A})\mathbf{W}_2\mathbf{P}_a^\perp\mathbf{W}_2\text{vec}(\mathbf{A})}.\end{aligned} \tag{50}$$

Based on (44), we write $\hat{\alpha}$ as

$$\begin{aligned}\hat{\alpha} =& \text{vec}^H(\hat{\mathbf{A}})(\mathbf{F}_{RF}^H)^\dagger s^*(\mathbf{F}_{RF}^H \text{vec}(\mathbf{A})\alpha s^* + \mathbf{n}) \\ =& \sigma_s^2 \text{vec}^H(\hat{\mathbf{A}})\text{vec}(\mathbf{A})\alpha + \text{vec}^H(\hat{\mathbf{A}})(\mathbf{F}_{RF}^H)^\dagger s^* \mathbf{n},\end{aligned} \tag{51}$$

where $\text{vec}^H(\hat{\mathbf{A}})$ is constructed from the estimate $(\hat{w_1}, \hat{w_2})$. With the help of Taylor's expansion, $\text{vec}^H(\hat{\mathbf{A}})$ can be approximated by

$$\text{vec}^H(\hat{\mathbf{A}}) \approx \text{vec}^H(\mathbf{A}) + j\text{vec}^H(\mathbf{A})\mathbf{W}_i\Delta w_i, \quad i = 1, 2. \tag{52}$$

Substituting (52) into (51), we rewrite $\hat{\alpha}$ as

$$\begin{aligned}\hat{\alpha} =& \sigma_s^2(\text{vec}^H(\mathbf{A}) + j\text{vec}^H(\mathbf{A})\mathbf{W}_1\Delta w_1)\text{vec}(\mathbf{A})\alpha \\ &+ \text{vec}^H(\hat{\mathbf{A}})(\mathbf{F}_{RF}^H)^\dagger s^* \mathbf{n} \\ =& \alpha + j\text{vec}^H(\mathbf{A})\mathbf{W}_1\text{vec}(\mathbf{A})\Delta w_1 \alpha + \text{vec}^H(\hat{\mathbf{A}})(\mathbf{F}_{RF}^H)^\dagger s^* \mathbf{n}.\end{aligned} \tag{53}$$

With the help of (46), we can derive the mean and the MSE of the channel gain estimation as

$$\begin{aligned}\text{E}\{\Delta\alpha\} =& \text{E}\left\{j\text{vec}^H(\mathbf{A})\mathbf{W}_1\text{vec}(\mathbf{A})\Delta w_1\alpha \right. \\ &\left. + \text{vec}^H(\hat{\mathbf{A}})(\mathbf{F}_{RF}^H)^\dagger s^* \mathbf{n}\right\} = 0, \\ \text{E}\{\Delta\alpha^2\} =& \text{E}\{(\hat{\alpha}-\alpha)(\hat{\alpha}-\alpha)^H\} \\ =& \alpha\text{E}\{(\Delta w_1)^2\}\alpha^H|\text{vec}^H(\mathbf{A})\mathbf{W}_1\text{vec}(\mathbf{A})|^2 \\ &+ \sigma_s^2\text{vec}^H(\hat{\mathbf{A}})(\mathbf{F}_{RF}^H)^\dagger\text{E}\{\mathbf{n}\mathbf{n}^H\}((\mathbf{F}_{RF}^H)^\dagger)^H\text{vec}(\hat{\mathbf{A}}) \\ =& \frac{\sigma_n^2|\text{vec}^H(\mathbf{A})\mathbf{W}_1\text{vec}(\mathbf{A})|^2}{2\sigma_s^2\text{vec}^H(\mathbf{A})\mathbf{W}_1\mathbf{P}_a^\perp\mathbf{W}_1\text{vec}(\mathbf{A})} + \sigma_n^2.\end{aligned} \tag{54}$$

In (54), the first term is caused by the estimation error in (ϕ, θ), while the second part is caused by the noise only. If (ϕ, θ) are perfectly estimated, $\text{E}\{\Delta\alpha^2\}$ only depends on the second term in (54), which makes it equivalent to the covariance of the traditional channel estimation methods.

Theorem 1. *The MSE of $\hat{\alpha}$ is then given by*

$$\mathit{MSE}(\hat{\alpha}) = \frac{\sigma_n^2|vec^H(\mathbf{A})\mathbf{W}_1 vec(\mathbf{A})|^2}{2\sigma_s^2 vec^H(\mathbf{A})\mathbf{W}_1\mathbf{P}_a^\perp\mathbf{W}_1 vec(\mathbf{A})} + \sigma_n^2. \tag{55}$$

From (50) and (54), we know that the joint ML estimator is unbiased for both (ϕ, θ) and α. Thus the analysis on their CRLBs are necessary to show the effectiveness of these estimators, which will be provided in the next subsection.

B. CRLB Analysis

In this subsection, we compute the CRLBs for the channel gain and the DOA estimation under UPA antenna configurations. It is worth noting that the MSE of the proposed estimators is irrelevant to analog beamforming. Thus, we omit analog beamforming for simplicity. With single LOS path ($L = 1$), the received signal $\mathbf{Y}$ can be expressed as

$$\begin{aligned}\mathbf{Y} =& \mathbf{H}s + \mathbf{N} = \alpha\mathbf{A}(\phi,\theta)s + \mathbf{N} \\ =& \alpha\mathbf{a}(w_1)\mathbf{a}^T(w_2)s + \mathbf{N}.\end{aligned} \tag{56}$$

The (m, n)th received signal is given by

$$y_{m,n} = \alpha e^{j((m-1)w_1+(n-1)w_2)}s + n_{m,n}, \tag{57}$$

where the real part of the received signal is

$$\begin{aligned}y_{m,n}^R =& \Re\{y_{m,n}\} = \Re\{\alpha e^{j((m-1)w_1+(n-1)w_2)}s\} + \Re\{n_{m,n}\} \\ =& x_{m,n} + n'_{m,n},\end{aligned} \tag{58}$$

$$\mathbf{I}(\mathbf{\Theta}) = \frac{1}{2\sigma^2}\left[\begin{array}{ccc} MN & 0 & 0 \\ 0 & (2\pi\alpha)^2 N\sum_{m=0}^{M-1} m^2 & (2\pi\alpha)^2\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n \\ 0 & (2\pi\alpha)^2\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n & (2\pi\alpha)^2 M\sum_{n=0}^{N-1} n^2 \end{array}\right]. \tag{62}$$

$$\begin{aligned}
\text{var}(\hat{\alpha}) \ &\geq A, \\
\text{var}(\hat{w_1}) \geq &\frac{2\sigma^2 M\sum_{n=0}^{N-1} n^2}{(2\pi\alpha)^2[N\sum_{m=0}^{M-1} m^2][M\sum_{n=0}^{N-1} n^2] - (2\pi\alpha)^2[\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n][\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n]} = AB\frac{6(2N-1)}{M-1}, \\
\text{var}(\hat{w_2}) \geq &\frac{2\sigma^2 N\sum_{m=0}^{M-1} m^2}{(2\pi\alpha)^2[N\sum_{m=0}^{M-1} m^2][M\sum_{n=0}^{N-1} n^2] - (2\pi\alpha)^2[\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n][\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n]} = AB\frac{6(2M-1)}{N-1},
\end{aligned} \tag{63}$$

$x_{m,n} = \Re\{\alpha e^{j((m-1)w_1+(n-1)w_2)}s\}$, and $n'_{m,n} = \Re\{n_{m,n}\}$. For given α, ϕ, and θ, the probability density function of $\mathbf{Y}$ can be expressed as

$$\begin{aligned} f(\mathbf{Y}|\alpha,\phi,\theta) &= \frac{1}{(2\pi\sigma^2)^{\frac{MN}{2}}}\exp\{-\frac{\|\mathbf{Y}-\alpha\mathbf{A}(\phi,\theta)s\|}{2\sigma^2}\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{MN}{2}}}\exp\{-\frac{1}{2\sigma^2}\sum_{m=1}^{M}\sum_{n=1}^{N}(y^R_{m,n}-x_{m,n})^2\}. \end{aligned} \tag{59}$$

Let us define $\mathbf{\Theta} = [\alpha, w_1, w_2]^T$ as the unknown parameter vector. The Fisher information matrix (FIM) is defined as

$$[\mathbf{I}(\mathbf{\Theta})]_{i,j} = -E\left[\frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial\mathbf{\Theta}_i\partial\mathbf{\Theta}_j}\right], \tag{60}$$

where

$$\ln f(\mathbf{Y}|\alpha,\phi,\theta) = -\frac{MN}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{m=1}^{M}\sum_{n=1}^{N}{n'}^2_{m,n}, \tag{61}$$

and $\sigma^2 = \sigma_n^2/\sigma_s^2$.

Lemma 2. *The FIM for the joint channel gain and DOA estimation of hybrid mmWave massive MIMO systems can be expressed as* (62) *shown on top of this page.*

Proof. See Appendix. □

Accordingly, the CRLB for the parameters of channel gain and DOA are CRLB $= \mathbf{I^{-1}}(\mathbf{\Theta})$. Thus we have (63) shown on top of this page, where $A = \frac{2\sigma^2}{MN}$, and $B = \frac{1}{(\pi\alpha)^2(7MN+M+N-5)}$.

Lemma 3. *The CRLB of the azimuth and elevation angle can be expressed as*

$$\begin{aligned} var(\hat{\theta}) &\geq AB\frac{6(2N-1)\lambda^2}{(M-1)d^2\sin^2\theta}, \\ var(\hat{\phi}) &\geq AB\frac{6\lambda^2(M-1)(2M-1)+6(N-1)Cd\cos\theta\cos\phi}{(M-1)(N-1)d^2\sin^2\theta\sin^2\phi}, \end{aligned} \tag{64}$$

where $C = 3\lambda(M-1) + (2N-1)d\cos\theta\cos\phi$.

Proof. For DOA estimation of hybrid mmWave massive MIMO system, the performance of azimuth angle θ and elevation angle ϕ are required. Therefore, we can use the following transformation to estimate the real angles of azimuth and elevation:

$$g(\mathbf{\Theta}) = \left[\begin{array}{c}\alpha\\ \theta\\ \phi\end{array}\right] = \left[\begin{array}{c}\alpha\\ \arccos(\frac{\lambda w_1}{d})\\ \arcsin(\frac{\lambda w_2}{d\sin\theta})\end{array}\right]. \tag{65}$$

Then, we can obtain the CRLBs of the azimuth and elevation angle of hybrid mmWave massive MIMO system through the following Jacobian matrix:

$$\begin{aligned} \text{var}(\hat{\theta}) &\geq [\frac{\partial g(\mathbf{\Theta})}{\partial\mathbf{\Theta}}\mathbf{I^{-1}}(\mathbf{\Theta})\frac{\partial g(\mathbf{\Theta}^T)}{\partial\mathbf{\Theta}}]_{2,2}, \\ \text{var}(\hat{\phi}) &\geq [\frac{\partial g(\mathbf{\Theta})}{\partial\mathbf{\Theta}}\mathbf{I^{-1}}(\mathbf{\Theta})\frac{\partial g(\mathbf{\Theta}^T)}{\partial\mathbf{\Theta}}]_{3,3}. \end{aligned} \tag{66}$$

Next, the CRLBs of the azimuth and elevation angle can be expressed as

$$\begin{aligned} \text{var}(\hat{\theta}) &\geq AB\frac{6(2N-1)\lambda^2}{(M-1)d^2\sin^2\theta}, \\ \text{var}(\hat{\phi}) &\geq 6AB\frac{\lambda^2(M-1)(2M-1)+Cd\cos\theta\cos\phi}{(M-1)(N-1)d^2\sin^2\theta\sin^2\phi}, \end{aligned} \tag{67}$$

where $C = 3\lambda(M-1)(N-1)+(N-1)(2N-1)d\cos\theta\cos\phi$. □

It is observed from (67) that the MSEs for both angle and channel gain estimators are inversely proportional to the SNR of the received signal, and the CRLBs decreases with increasing the number of antenna array.

V. Simulation Results

In this section, we show the effectiveness of the proposed estimation method through numerical examples. In our simulation, we consider a TDD mmWave massive MIMO system, where the UPA at the BS has $M = 100, N = 100$ antennas of $d = \lambda/2$, with $M_{RF} = 100$ RF chains. There are $K = 10$ single-antenna users uniformly distributed, and each user has $L = 10$ paths. The default value of preamble τ is set to be $\tau = 10$. We use the ray-tracing way to model the mmWave channels, and the channel matrix of different users are formulated according to (6). We take angle rotation search grids $G = G_M G_N = 30\times 30 = 900$ unless otherwise mentioned. In all examples, the DOA information of all users are estimated from the preamble. With $\tau = 10$, the overall

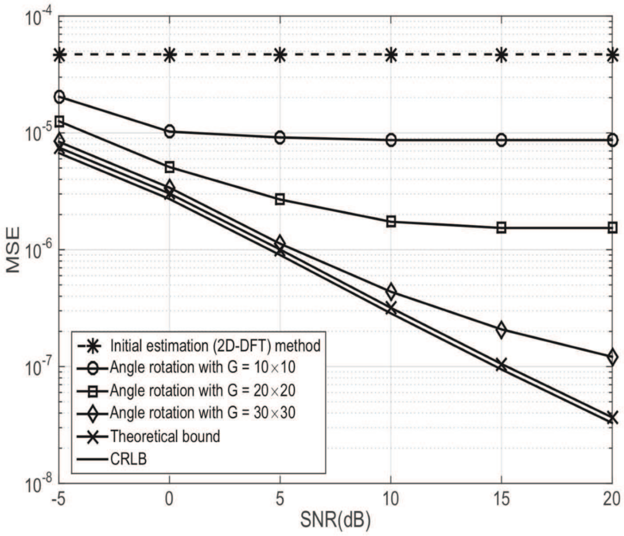


Fig. 3. Comparison of MSEs of the theoretical bound, CRLB, initial estimation method and the proposed DOA estimation schemes with searching guides $G = 10 \times 10, 20 \times 20, 30 \times 30$, respectively.

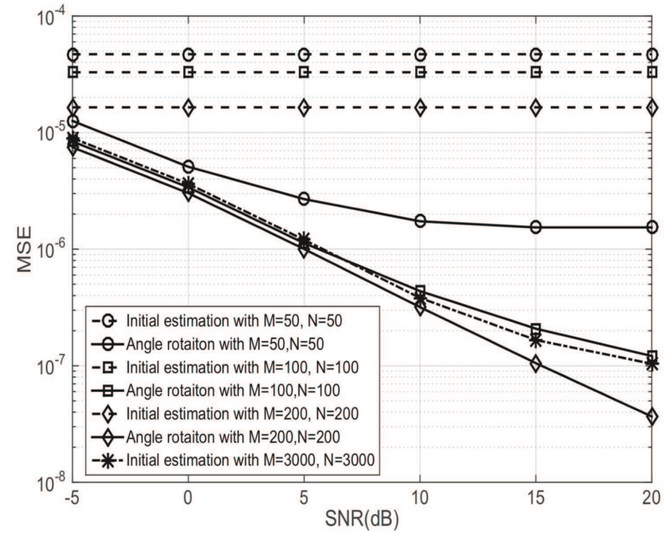


Fig. 4. The MSE comparison of the proposed DOA estimation and the initial estimation, with $M = 50, N = 50, M = 100, N = 100, M = 200, N = 200, M = 3000, N = 3000$, respectively.

users can transmit pilot synchronously such that the orthogonal training can be applied to obtain the DOA information.

Fig. 3 plots the MSEs of DOA estimation as a function of SNR for initial 2D-DFT, our proposed estimation method, theoretical bound, and CRLB. The total transmission power for uplink training is constrained to ρ for all users. It can be seen that our proposed DOA estimation method performs slightly worse than that of theoretical bound, but performs much better than the initial estimation (2D-DFT) method. Interestingly, the MSE of proposed DOA estimation method improves with increasing the searching grid, which is due to the improved angle resolution. If the searching grid goes to infinity, the proposed DOA estimation method can achieve the same MSE as the theoretical bound. It can also be seen from Fig. 3 that the traditional initial estimation method remains constant for any SNRs. The reason is that the Gaussian noise will keep the same level after the 2D-DFT, such that the power of noise will keep constant in all SNRs. In addition, the theoretical MSE is very close to the corresponding CRLB.

Fig. 4 plots the MSEs of DOA estimation as a function of SNR for various URA sizes. We assume that the total transmit power for each BS antenna is constrained constantly. It is clearly seen from Fig. 4 that increasing the number of BS antennas improves the DOA estimation accuracy due to the improved spatial signatures accuracy in both initial estimate and angle rotation estimate. It can also be seen from Fig. 4 that the proposed DOA estimation method outperforms the initial estimation dramatically in the high SNR region. Moreover, the initial estimation algorithm with the number of antennas at the base station reaching 3000×3000 and the proposed estimation algorithm with 100×100 antennas at the base station have the almost same performance. When the number of antennas is large enough, the initial estimation can approach the proposed method. Therefore, the proposed DOA estimation algorithm can greatly reduce the number of BS antennas while ensuring the accuracy of the estimation.

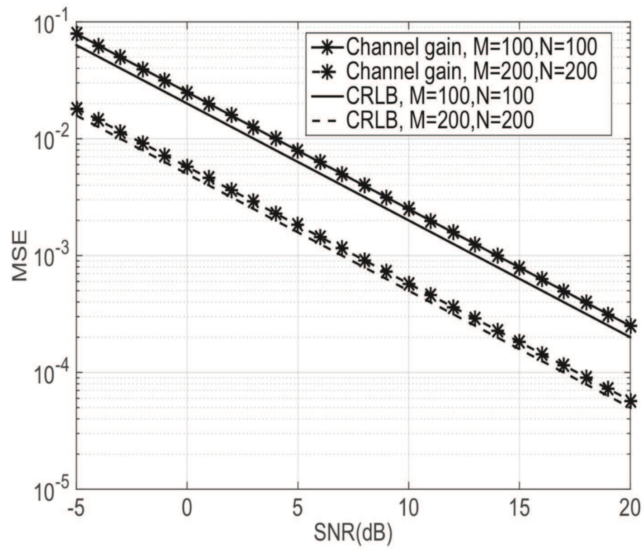


Fig. 5. Comparison of channel gain MSEs and the corresponding CRLB with $M = 100, N = 100, M = 200, N = 200$, respectively.

Fig. 5 plots the MSE of the proposed channel gain estimation method with the corresponding CRLB as a function of SNR. It can be seen that the MSE improves with increasing the number of antennas, due to the fact that the total training power is proportional to MN. It is also seen that the MSE of proposed channel gain estimation method is very close to the CRLB, especially in the large BS antennas scenario.

Fig. 6 compares the MSEs of the proposed channel estimation method, the eigen-decomposition based method [15], the beamspace method [23], and the CS method [21]. It can be seen that the MSE of joint spatial division and multiplexing (JSDM) is slightly better than the proposed one, since the former catches the exact eigen-direction to recover the channel. Nevertheless, it is not an easy and stable task to obtain the

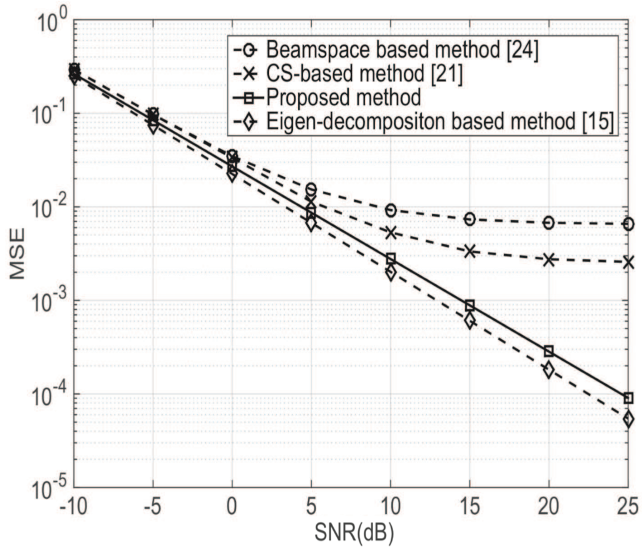


Fig. 6. The channel estimation MSEs of the proposed method, eigen-decomposition method, beamspace method, and the CS method.

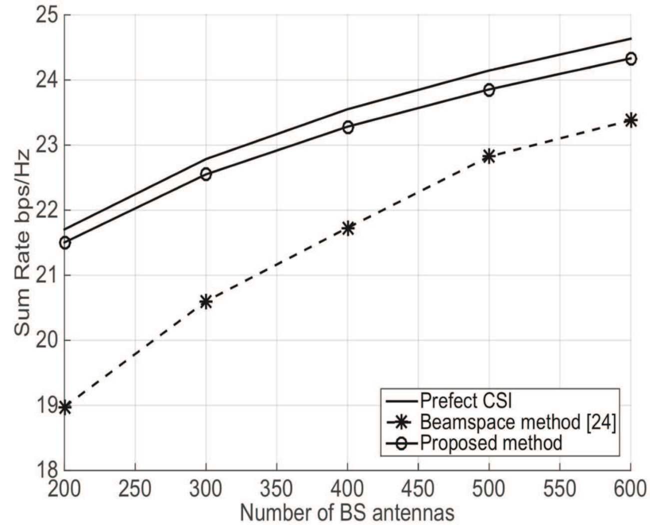


Fig. 8. Sum rate comparison of different methods, where the proposed method, beamspace method and prefect CSI are displayed for comparison as a function of the number of BS antennas.

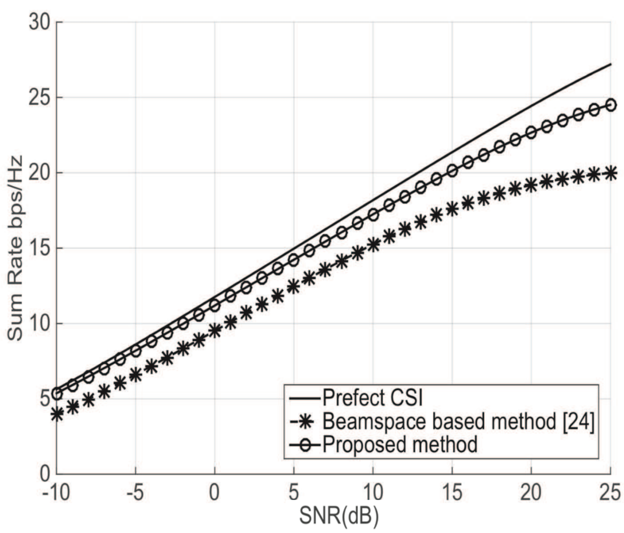


Fig. 7. Sum rate comparison of different methods, where the proposed method, beamspace method and prefect CSI are displayed for comparison as a function of SNR.

$M \times N$ dimensional channel covariance matrix for JSDM in practice. Though the proposed method, the beamspace method, and the CS method could directly handle the instantaneous channel estimation, the beamspace method and CS method have an error floor due to the power leakage problem in its sparse channel representation.

Fig. 7 plots the achievable sum rate for the downlink data transmission with different channel estimation method using the proposed hybrid precoding method (37). To make the comparison fair, the overall data power are set as the same for all methods. It can be seen from Fig. 7 that with the increasing of SNR, the performances of all methods become better. The achievable sum rate of the proposed method is much higher than that of beamspace method in any SNR values, and it is comparable to the performance of the prefect CSI case, especially in low SNR case. Note that beamspace method suffers from severe channel power leakage, while most of the channel power could concentrate on only few points through angle rotation. Therefore, the proposed method have a more desirable sum rate.

Fig. 8 plots the achievable sum rate for the downlink data transmission with the proposed method, beamspace method, and prefect CSI as a function of the number of BS antennas. To keep the comparison fair, the overall data power are set to be the same for each method. It is seen that with the increasing of the number of BS antennas, the performances of all methods become better. The sum rate achieved by the proposed channel estimation method greatly outperforms that of the beamspace method, but is slightly worse than that of the perfect CSI. When the number of BS antennas increases, the channel power leakage of beamspace method will decrease due to the improves of angle resolution. Thus as the number of antennas increases, the gap between the proposed method and the beamspace method becomes smaller.

VI. Conclusion

In this paper, we proposed a novel channel estimation for hybrid digital and analog mmWave massive MIMO system, where the channel is decomposed into DOA information and channel gain information. In our estimation method, a fast DOA estimation algorithm was designed based on 2D-DFT and angle rotation, and the channel gain estimation was performed with very small amount of training resources, which significantly reduces the training overhead and the feedback cost. To evaluate the benefits of our proposed method, we derived the theoretical bounds of MSE and CRLB of the joint DOA and channel gain estimation in high SNR region. It is shown that our proposed estimation method is very close to the CRLB, especially in the large BS antennas case.

VII. Appendix

A. Proof of Lemma 1

Let us define

$$\mathbf{y}_d = \mathbf{F}_{RF}^H \text{vec}(\mathbf{A})\alpha s. \tag{68}$$

The received signal $\mathbf{y} = \mathbf{y}_d + \mathbf{n}$ is actually a perturbed version of $\mathbf{y}_d$. At high SNR, the first derivative of the cost function can be approximated using Taylors expansion as

$$\begin{aligned} 0 =& \frac{\partial g(w_1, w_2)}{\partial w_i}|_{w_i=\hat{w}_i} \\ \approx& \frac{\partial g(w_1, w_2)}{\partial w_i}|_{w_i=\hat{w}_i} + \frac{\partial^2 g(w_1, w_2)}{\partial^2 w_i}|_{w_i=\hat{w}_i}\Delta w_i, \end{aligned} \tag{69}$$

where $\Delta w_i = \hat{w}_i - w_i$ is the perturbation in the w_i. Thus, Δw_i can be represented as

$$\Delta w_i = -\frac{\frac{\partial g(w_1,w_2)}{\partial w_i}|_{w_i=\hat{w}_i}}{\frac{\partial^2 g(w_1,w_2)}{\partial^2 w_i}|_{w_i=\hat{w}_i}} = -\frac{\dot{g}(w_1, w_2|w_i)}{\ddot{g}(w_1, w_2|w_i)}. \tag{70}$$

In (70), the first order derivative can be calculated as

$$\begin{aligned} \dot{g}(w_1, w_2|w_i) =& j\mathbf{y}^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y} \\ &- j\mathbf{y}^H \mathbf{F}_{RF}^H \mathbf{P}_a^{\perp} \mathbf{W}_i (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y}. \end{aligned} \tag{71}$$

Since $\mathbf{P}_a^{\perp}(\mathbf{F}_{RF}^H)^{\dagger}\mathbf{y}_d = \mathbf{0}$, we can rewrite (71) as

$$\begin{aligned} \dot{g}(w_1, w_2|w_i) =& j(\mathbf{y}_d + \mathbf{n})^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} (\mathbf{F}_{RF}^H)^{\dagger} (\mathbf{y}_d + \mathbf{n}) \\ &- j(\mathbf{y}_d + \mathbf{n})^H \mathbf{F}_{RF}^H \mathbf{P}_a^{\perp} \mathbf{W}_i (\mathbf{F}_{RF}^H)^{\dagger} (\mathbf{y}_d + \mathbf{n}) \\ =& -2\Im\{\mathbf{y}_d^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{n}\} \\ &- 2\Im\{\mathbf{n}^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{n}\}. \end{aligned} \tag{72}$$

Due to the independence between $\mathbf{y}_d$ and $\mathbf{n}$, we can obtain

$$\text{E}\{\dot{g}(w_1, w_2|w_i)\} = -2\sigma_n^2 \Im\{tr\{\mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} (\mathbf{F}_{RF}^H)^{\dagger}\}\} = 0. \tag{73}$$

The second-order derivative can be calculated as

$$\begin{aligned} \ddot{g}(w_1, w_2|w_i) =& -\mathbf{y}^H \mathbf{F}_{RF}^H \mathbf{W}_i^2 \mathbf{P}_a^{\perp} (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{n} \\ &+ \mathbf{y}^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y} \\ &+ \mathbf{y}^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y} \\ &- \mathbf{n}^H \mathbf{F}_{RF}^H \mathbf{P}_a^{\perp} \mathbf{W}_i^2 (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y}. \end{aligned} \tag{74}$$

Based on (74), we have

$$\begin{aligned} \text{E}\{\ddot{g}(w_1, w_2|w_i)\} =& 2\mathbf{y}_d^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y}_d \\ &+ \text{E}\{tr(-\mathbf{W}_i^2 \mathbf{P}_a^{\perp} + 2\mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i - \mathbf{P}_a^{\perp} \mathbf{W}_i^2)\} \\ =& 2\sigma_s^2 |\alpha|^2 \text{vec}(\mathbf{A}^H) \mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i \text{vec}(\mathbf{A}). \end{aligned} \tag{75}$$

Therefore, $\ddot{g}(w_1, w_2|w_i)$ can be rewritten as

$$\ddot{g}(w_1, w_2|w_i) = \text{E}\{\ddot{g}(w_1, w_2|w_i)\} + O_2(\mathbf{n}) + O_2(\mathbf{n}^2), \tag{76}$$

where $O_2(\mathbf{n})$ and $O_2(\mathbf{n}^2)$ represent the linear and quadrature functions of $\mathbf{n}$ existing in $\ddot{g}(w_1, w_2|w_i)$. Similarly, $\dot{g}(w_1, w_2|w_i)$ can be expressed as

$$\dot{g}(w_1, w_2|w_i) = O_1(\mathbf{n}) + O_1(\mathbf{n}^2), \tag{77}$$

where $O_1(\mathbf{n})$ and $O_1(\mathbf{n}^2)$ represent the linear and quadrature functions of $\mathbf{n}$ existing in $\dot{g}(w_1, w_2|w_i)$. Substituting (76) and (77) into (70) and assuming high SNR, i.e., $\|\mathbf{n}\|^2 \ll \|\mathbf{y}_d\|^2$, we have

$$\begin{aligned} \Delta w_i =& -\frac{O_1(\mathbf{n}) + O_1(\mathbf{n}^2)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\} + O_2(\mathbf{n}) + O_2(\mathbf{n}^2)} \\ =& -\frac{O_1(\mathbf{n}) + O_1(\mathbf{n}^2)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}} \times \left(1 - \frac{O_2(\mathbf{n}) + O_2(\mathbf{n}^2)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}} \right. \\ &\left. + \left(\frac{O_2(\mathbf{n}) + O_2(\mathbf{n}^2)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}}\right)^2 - \cdots \right). \end{aligned} \tag{78}$$

Knowing that $\frac{O_1(\mathbf{n})+O_1(\mathbf{n}^2)}{\text{E}\{\ddot{g}(w_1,w_2|w_i)\}}$ and $\frac{O_2(\mathbf{n})+O_2(\mathbf{n}^2)}{\text{E}\{\ddot{g}(w_1,w_2|w_i)\}}$ are small terms at high SNR, their product can be ignored, and Δw_i can be approximated as

$$\Delta w_i \approx -\frac{O_1(\mathbf{n}) + O_1(\mathbf{n}^2)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}} = -\frac{\dot{g}(w_1, w_2|w_i)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}}. \tag{79}$$

The above analysis gives the explicit form of perturbation in the estimate of w_i for a specific realization of the training. Thus, we can express the expectation and variance of the DOA estimation as

$$\begin{aligned} \text{E}\{\Delta w_i\} =& \text{E}\left\{-\frac{\dot{g}(w_1, w_2|w_i)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}}\right\} = -\frac{\text{E}\{\dot{g}(w_1, w_2|w_i)\}}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}}, \\ \text{E}\{\Delta w_i^2\} =& \text{E}\left\{\left(-\frac{\dot{g}(w_1, w_2|w_i)}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}}\right)^2\right\} \\ =& \frac{\text{E}\{\dot{g}(w_1, w_2|w_i)^2\}}{\text{E}\{\ddot{g}(w_1, w_2|w_i)\}^2}, \end{aligned} \tag{80}$$

respectively. To derive the variance of DOA information, we express

$$\begin{aligned} \text{E}\{\dot{g}(w_1, w_2|w_i)^2\} =& 2\text{E}\{\mathbf{y}_d^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{n}\mathbf{n}^H \mathbf{F}_{RF}^H \mathbf{P}_a^{\perp} \\ &\mathbf{W}_i (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y}_d\} + \text{E}\{(\mathbf{n}^H \mathbf{F}_{RF}^H (\mathbf{W}_i \mathbf{P}_a^{\perp} - \mathbf{P}_a^{\perp} \mathbf{W}_i)(\mathbf{F}_{RF}^H)^{\dagger} \mathbf{n})^2\}, \end{aligned} \tag{81}$$

where the second term can be ignored for high SNR to obtain

$$\begin{aligned} \text{E}\{\dot{g}(w_1, w_2|w_i)^2\} =& 2\sigma_n^2 \mathbf{y}_d^H \mathbf{F}_{RF}^H \mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i (\mathbf{F}_{RF}^H)^{\dagger} \mathbf{y}_d \\ =& 2\sigma_s^2 \sigma_n^2 |\alpha|^2 \text{vec}(\mathbf{A}^H) \mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i \text{vec}(\mathbf{A}). \end{aligned} \tag{82}$$

Substituting (73), (75), (82) into (80), we obtain the mean and the MSE of estimation $\hat{w}_i$ as

$$\begin{aligned} \text{E}\{\Delta w_i\} =& 0, \\ \text{E}\{\Delta w_i^2\} =& \frac{\sigma_n^2}{2\sigma_s^2 |\alpha|^2 \text{vec}^H(\mathbf{A}) \mathbf{W}_i \mathbf{P}_a^{\perp} \mathbf{W}_i \text{vec}(\mathbf{A})}. \end{aligned} \tag{83}$$

The proof is completed.

B. Proof of Lemma 2

Performing mathematical calculation on (60), we obtain

$$\begin{aligned} &\frac{\partial^2 \ln f(\mathbf{Y}|\alpha, \phi, \theta)}{\partial^2 \mathbf{\Theta}_1} \\ =& -\frac{MN}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \cos(4\pi m w_1 + 4\pi n w_2), \end{aligned} \tag{84}$$

$$\frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial^2 \mathbf{\Theta}_2} = -\frac{(2\pi\alpha)^2}{2\sigma^2}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[m^2 - m^2\cos(4\pi m w_1 + 4\pi n w_2)], \tag{85}$$

$$\frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial^2 \mathbf{\Theta}_3} = -\frac{(2\pi\alpha)^2}{2\sigma^2}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[n^2 - n^2\cos(4\pi m w_1 + 4\pi n w_2)], \tag{86}$$

$$\frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial \mathbf{\Theta}_1 \partial \mathbf{\Theta}_2} = \frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial \mathbf{\Theta}_2 \partial \mathbf{\Theta}_1} = \frac{2\pi\alpha}{2\sigma^2}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} m\sin(4\pi m w_1 + 4\pi n w_2), \tag{87}$$

$$\frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial \mathbf{\Theta}_1 \partial \mathbf{\Theta}_3} = \frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial \mathbf{\Theta}_3 \partial \mathbf{\Theta}_1} = \frac{2\pi\alpha}{2\sigma^2}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} n\sin(4\pi m w_1 + 4\pi n w_2), \tag{88}$$

$$\frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial \mathbf{\Theta}_2 \partial \mathbf{\Theta}_3} = \frac{\partial^2 \ln f(\mathbf{Y}|\alpha,\phi,\theta)}{\partial \mathbf{\Theta}_3 \partial \mathbf{\Theta}_2} = \frac{(2\pi\alpha)^2}{2\sigma^2}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[mn(\cos(4\pi m w_1 + 4\pi n w_2) - 1)]. \tag{89}$$

Lemma 4. *For $x \in [0, 2\pi)$, we have*

$$\lim_{K\to\infty}\sum_{k=1}^{K} k^i \sin(4\pi k x) = 0, \quad i = 0,1,2, \tag{90}$$

$$\lim_{K\to\infty}\sum_{k=1}^{K} k^i \cos(4\pi k x) = 0, \quad i = 0,1,2. \tag{91}$$

Proof. According to [45, Eq. (AD361)], we have

$$\lim_{K\to\infty}\sum_{k=1}^{K}\sin(4\pi k x) = \frac{\sin((K+1)2\pi x)\sin(2\pi K x)}{\sin(2\pi x)} = 0, \tag{92}$$

$$\lim_{K\to\infty}\sum_{k=1}^{K}\cos(4\pi k x) = \frac{\cos((K+1)2\pi x)\sin(2\pi K x)}{\sin(2\pi x)} = 0, \tag{93}$$

$$\lim_{K\to\infty}\sum_{k=1}^{K} k\sin(4\pi k x) = \frac{\sin 4\pi K x}{4\sin^2(2\pi x)} - \frac{K\cos((2K-1)2\pi x)}{2\sin(2\pi x)} = 0, \tag{94}$$

$$\lim_{K\to\infty}\sum_{k=1}^{K} k\cos(4\pi k x) = \frac{K\sin((2K-1)2\pi x)}{2\sin(2\pi x)} - \frac{1-\cos 4\pi K x}{4\sin^2(2\pi x)} = 0, \tag{95}$$

for $i = 2$, where the limits follow from relations similar to (95). □

As M and N go to infinity in massive MIMO systems, there are

$$\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\cos(4\pi m w_1 + 4\pi n w_2) = 0,$$
$$\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} m\sin(4\pi m w_1 + 4\pi n w_2) = 0,$$
$$\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} m^2\cos(4\pi m w_1 + 4\pi n w_2) = 0,$$
$$\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} mn\cos(4\pi m w_1 + 4\pi n w_2) = 0. \tag{96}$$

Substituting (84)–(89) into (60), the FIM for the joint channel gain and DOA estimation of hybrid mmWave massive MIMO system can be expressed as (62). The proof is completed.

References

[1] C. Wang, F. Haider, X. Gao, H. You, and Y. Yang, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–130, Feb. 2014.

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

[3] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, May. 2013.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?", *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] X. Zhang, A. F. Molisch, and S. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[6] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

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

[8] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[9] A. Alkhateeb and R. W. Heath Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May. 2016.

[10] X. Yang, W. Lu, N. Wang, K. Nieman, C. Wen, C. Zhang, S. Jin, H. Zhu, X. Mu, I. Wong, Y. Huang, and X. You, "Design and Implementation of a TDD-Based 128-Antenna Massive MIMO Prototype System," *China Commun.*, vol.14, no. 12, pp. 162–187, Dec. 2017.

[11] L. Jiang, C. Qin, X. Zhang, and H. Tian, "Secure Beamforming Design for SWIPT in Cooperative D2D Communications," *China Commun.*, vol.14, no.1, pp. 20–33, Jan. 2017.

[12] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2014.

[13] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, Nov. 2016.

[14] C. Sun, X. Q.Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

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

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

[17] S. L. H. Nguyen and A. Ghrayeb, “Compressive sensing-based channel estimation for massive multiuser MIMO systems,” in *Proc. IEEE WCNC*, Shanghai, China, Apr. 2013, pp. 2890–2895.

[18] X. Rao and V. K.Lau, “Distributed compressive CSIT estimation and feed-back for FDD multi-user massive MIMO systems,” *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

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

[20] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., “Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2909–2913.

[21] Z. Gao, L. Dai, Z. Wang, and S. Chen, “Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO,” *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[22] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[23] X. Gao, L. Dai, S. Han, C. I, X. Wang, “Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,” *IEEE Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.

[24] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcicy, and R. W. Heath Jr., “Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems,” in *Intern. Conf. Commun. (ICC'15)*. IEEE, Jun. 2015, pp. 1340–1345.

[25] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. Gonzalez-Prelcicy, and R. W. Heath Jr., “Channel estimation and hybrid combining for mmWave: phase shifters or switches?,” in *Proc. Inf. Theory Appl. Workshop*, Feb. 2015, pp. 90–97.

[26] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcicy, A. Alkhateeb, and R. W. Heath Jr., “Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches?,” *IEEE Access*, vol. 4, pp. 247–267, Jan 2016.

[27] J. Singh and S. Ramakrishna, “On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2670–2683, May 2015.

[28] R. Mendez-Rial, N. Gonzalez-Prelcicy, and R. W. Heath Jr., “Adaptive hybrid precoding and combining in mmWave multiuser MIMO systems based on compressed covariance estimation,” in *Proc. Intern. Workshop on Computational Advances in Multi-Sensor Adaptive Process.*, Dec. 2015, pp. 213–216.

[29] H. Xie, F. Gao, S. Zhang, and S. Jin, “A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[30] H. Xie, B. Wang, F. Gao, and S. Jin, “A full-space spectrum-sharing strategy for massive MIMO cognitive radio systems,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2537–2549, Oct. 2016.

[31] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[32] R. Roy and T. Kailath, “Esprit-estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[33] H. Krim and M. Viberg, “Two decades of array signal processing research: the parametric approach,” *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[34] T. Wang, B. Ai, R. He, and Z. Zhong, “Two-dimension direction-of-arrival estimation for massive MIMO systems,” *IEEE Access*, vol. 3, pp. 2122–2128, Nov. 2015.

[35] A. Hu, T. Lv, H. Gao, Z. Zhang, and S. Yang, “An ESPRIT-based approach for 2-D localization of incoherently distributed sources in massive MIMO systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 996–1011, Oct. 2014.

[36] A. Wang, L. Liu, and J. Zhang, “Low complexity direction of arrival (DOA) estiamtion for 2D massive MIMO systems,” in *Proc. Global Telecomm. Conf.*, Dec. 2012, pp. 703–707.

[37] 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.*, vol. 63, no. 22, pp. 5897–5909, Nov. 2015.

[38] R. Shafin, L. Liu, and J. Zhang, “DoA estimation and capacity analysis for 3D massive-MIMO/FD-MIMO OFDM system,” in *Proc. Global Conf. on Signal and Information Process.*, Dec. 2015, pp. 181–184.

[39] T. Rappaport, Y. Qiao, J. Tamir, J. Murdock, and E. Ben-Dor, “Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems,” in *Proc. Radio and Wireless Symp. (RWS)*, Santa Clara, CA, USA, Jan. 2012, pp. 151–154.

[40] J. Murdock, E. Ben-Dor, Y. Qiao, J. Tamir, and T. Rappaport, “A 38 GHz cellular outage study for an urban outdoor campus environment,” in *Proc. Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2012, pp. 3085–3090.

[41] A. Sayeed and V. Raghavan, “Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.

[42] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, “Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications,” *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[43] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, “Angle Domain Signal Processing aided Channel Estimation for Indoor 60GHz TDD/FDD Massive MIMO Systems,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Sep. 2017.

[44] P. Stoica and N. Arye, “MUSIC, maximum likelihood, and Cramer-Rao bound,” *IEEE Trans. Acoustics Speech and Signal Process.*, vol. 37, no. 5, pp. 720–741, 1989.

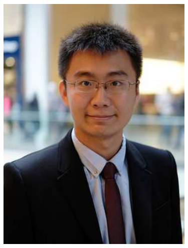
[45] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic, 1980.

Dian Fan received his B.Eng. degree from the School of Science, Beijing Jiaotong University (BJTU), Beijing, China, in 2014. He is currently pursuing the Ph.D. degree at the School of Computer and Information Technology, BJTU. He was a Visiting Ph.D. Student at the Department of Informatics at King's College London from October 2016 to March 2017, and at the School of Electronic Engineering and Computer Science at Queen Mary University of London from October 2017 to September 2018. His research interests include MIMO techniques, massive MIMO systems, millimeter wave systems and array signal processing.

Feifei Gao (M'09, SM'14) received the B.Eng. degree from Xian Jiaotong University, Xi'an, China in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada in 2004, and the Ph.D. degree from National University of Singapore, Singapore in 2007. He was a Research Fellow with the Institute for Infocomm Research (I2R), A*STAR, Singapore in 2008 and an Assistant Professor with the School of Engineering and Science, Jacobs University, Bremen, Germany from 2009 to 2010. In 2011, he joined the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor.

Prof. Gao's research areas include communication theory, signal processing for communications, array signal processing, and convex optimizations, with particular interests in MIMO techniques, multi-carrier communications, cooperative communication, and cognitive radio networks. He has authored/coauthored more than 120 refereed IEEE journal papers and more than 120 IEEE conference proceeding papers.

Prof. Gao has served as an Editor of IEEE Transactions on Wireless Communications, IEEE Signal Processing Letters, IEEE Communications Letters, IEEE Wireless Communications Letters, International Journal on Antennas and Propagations, and China Communications. He has also serves as the symposium co-chair for 2018 IEEE Vehicular Technology Conference Spring (VTC), 2015 IEEE Conference on Communications (ICC), 2014 IEEE Global Communications Conference (GLOBECOM), 2014 IEEE Vehicular Technology Conference Fall (VTC), as well as Technical Committee Members for many other IEEE conferences.



Yuanwei Liu (S'13-M'16) received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications in 2011 and 2014, respectively, and the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2016. He was with the Department of Informatics, Kings College London, from 2016 to 2017, where he was a Post-Doctoral Research Fellow. He has been a Lecturer (Assistant Professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London, since 2017.

His research interests include 5G wireless networks, Internet of Things, machine learning, stochastic geometry, and matching theory. He received the Exemplary Reviewer Certificate of the IEEE wireless communication letters in 2015, the IEEE transactions on communications in 2016 and 2017, the IEEE transactions on wireless communications in 2017. Currently, He is in the editorial board of serving as an Editor of the IEEE communication letters and the IEEE access. He also serves as a guest editor for IEEE JSTSP special issue on "Signal Processing Advances for Non-Orthogonal Multiple Access in Next Generation Wireless Networks ". He has served as the Publicity Co-Chairs forVTC2019-Fall. Hehas served as a TPC Member for many IEEE conferences, such as GLOBECOM and ICC.

Y ansha Deng (S'13M'18) received the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2015. From 2015 to 2017, she was a Post-Doctoral Research Fellow with Kings College London, U.K., where she is currently a Lecturer (Assistant Professor) with the Department of Informatics. Her research interests include molecular communication, Internet of Things, and 5G wireless networks. She was a recipient of the Best Paper Awards from ICC 2016 and Globecom 2017 as the first author. She is currently an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE COMMUNICATION LETTERS. She also received an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2016 and 2017. She has also served as a TPC Member for many IEEE conferences, such as IEEE GLOBECOM and ICC.

Gongpu Wang received the B.Eng. degree in communication engineering from Anhui University, Hefei, China, in 2001, the M.Sc. degree from the Beijing University of Posts and Telecommunications (BUPT), China, in 2004, and the Ph.D. degree from the University of Alberta, Edmonton, AB, Canada, in 2011. From 2004 to 2007, he was an Assistant Professor with the School of Network Edu- cation, BUPT. After graduation, he joined the School of Computer and Information Technology, Beijing Jiaotong University, China, where he is currently an Associate Professor. His research interests include Internet of Things, wireless communication theory, and signal processing technologies.

Zhangdui Zhong received the B.E. and M.S. degrees from Beijing Jiaotong University, Beijing, China, in 1983 and 1988, respectively. He is currently a Professor and an advisor of Ph.D. candidates with Beijing Jiaotong University, Beijing, China, where he is also a Chief Scientist of the State Key Laboratory of Rail Traffic Control and Safety. He is also a Director of the Innovative Research Team of the Ministry of Education, Beijing, and a Chief Scientist of the Ministry of Railways, Beijing. His interests include wireless communications for railways, control theory and techniques for railways, and GSM-R systems. His research has been widely used in railway engineering, such as at the Qinghai-Xizang railway, the Datong-Qinhuangdao Heavy Haul railway, and many high-speed railway lines in China. He is an Executive Council member of the Radio Association of China, Beijing, and a Deputy Director of Radio Association, Beijing. He has authored or co-authored seven books, five invention patents, and over 200 scientific research papers in his research area. He received the MaoYiSheng Scientific Award of China, the ZhanTianYou Railway Honorary Award of China, and the Top 10 Science/Technology Achievements Award of Chinese Universities.

A rumugam Nallanathan (S'97-M'00-SM'05-F'17) is Professor of Wireless Communications and Head of the Communication Systems Research (CSR) group in the School of Electronic Engineering and Computer Science at Queen Mary University of London since September 2017. He was with the Department of Informatics at Kings College London from December 2007 to August 2017, where he was Professor of Wireless Communications from April 2013 to August 2017 and a Visiting Professor from September 2017. He was an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore from August 2000 to December 2007. His research interests include 5G Wireless Networks, Internet of Things (IoT) and Molecular Communications. He published nearly 400 technical papers in scientific journals and international conferences. He is a co-recipient of the Best Paper Awards presented at the IEEE International Conference on Communications 2016 (ICC'2016) , IEEE Global Communications Conference 2017 (GLOBECOM'2017) and IEEE Vehicular Technology Conference 2018 (VTC'2018). He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science Highly Cited Researcher in 2016.

He is an Editor for IEEE Transactions on Communications. He was an Editor for IEEE Transactions on Wireless Communications (2006-2011), IEEE Transactions on Vehicular Technology (2006-2017), IEEE Wireless Communications Letters and IEEE Signal Processing Letters. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society and Technical Program Chair and member of Technical Program Committees in numerous IEEE conferences. He received the IEEE Communications Society SPCE outstanding service award 2012 and IEEE Communications Society RCC outstanding service award 2014.