Angle Domain Channel Estimation in Hybrid MmWave Massive MIMO Systems

Dian Fang, 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 consumption 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 independent of the hardware constraints. In [27], a grid-of-beams (GoB) based approach
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 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; $tr(A)$ denotes the trace of $A$; $|A|_{ij}$ is the $(i, j)$th entry of $A$; $\text{Diag}(a)$ denotes a diagonal matrix with the diagonal element constructed from $a$, while $\text{Diag}(A)$ denotes a vector whose elements are extracted from the diagonal components of $A$; $\text{vec}(A)$ denotes the vectorization of $A$; $\mathcal{R}\{A\}$ denotes the real part of $A$; $\mathcal{S}\{A\}$ denotes the Imaginary part of $A$; $|a|_{i,j}$ denotes the subvector of $a$ that starts with $[a]_i$ and ends at $[a]_j$; $|A|_{i,j}$ denotes the submatrix of $A$ that starts with row $[a]_i$ and ends at row $[a]_j$; $\mathbb{E}\{\cdot\}$ denotes the statistical expectation, and $\|\cdot\|$ is the Euclidean norm of $h$.

**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 $M_{RF} \leq MN$ 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 $F_{BB}(F_{BB} \in \mathbb{C}^{M_{RF} \times K})$, followed by an analog beamformer $F_{RF}(F_{RF} \in \mathbb{C}^{MN \times M_{RF}})$. To simplify the hardware implementation, each element of $F_{RF}$ has unitary magnitude with arbitrary phase. As $F_{RF}$ is implemented using analog phase shifters, its elements are constrained to satisfy $[[F_{RF}]_{i,j}]_{k} = \frac{1}{\sqrt{M_{RF}}} (j = 1, 2, \cdots M_{RF}, i = 1, 2, \cdots MN)$, where all elements of $F_{RF}$ have equal norm. The total power constraint is enforced by normalizing $F_{BB}$ to satisfy $\|F_{RF}F_{BB}\|_F^2 = 1$, $k = 1, 2, \cdots, K$.

**A. Transmitter Model**

Denote $H_k$ as the $M \times N$ channel matrix between the BS and the $k$th mobile user. In the uplink transmission stage, the
received signal at the BS can be expressed as

\[ y(t) = F_{BB}^H F_{RF}^H \sum_{k=1}^{K} \text{vec}(H_k) s_k(t) + N, \quad t = 1, \cdots, T, \] (1)

where \( s_k(t) \) is the transmitted signal at time \( t \), \( N \sim \mathcal{CN}(0, \sigma_n^2 I) \) is the complex Gaussian noise matrix, and \( \sigma_n^2 \) 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 \( k \)th mobile user is given by

\[ y_k = \text{vec}^H(H_k) F_{RF} F_{BB} s + n_k, \] (2)

where \( s = [s_1, s_2, \cdots, s_K]^T \) is the transmitted signal vector for all \( K \) mobile users. Thus, we can express the received SNR at the \( k \)th mobile user as

\[ \Gamma_k = |\text{vec}^H(H_k) F_{RF} F_{BB}|^2 \frac{\sigma_n^2}{\sigma_a^2}, \] (3)

where \( E(|s_k|^2) = \sigma_n^2 \) denotes the power of \( s_k \).

### 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 \( l \)th path of the \( k \)th user. The corresponding steering matrix can be expressed as (4) shown on top of this page, where \( \Lambda \) 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

\[ A(\phi_{l,k}, \theta_{l,k}) = A(w_{1,l,k}, w_{2,l,k}) = a(w_{1,l,k}) a^T(w_{2,l,k}), \] (5)

where

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

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

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

\[ H_k = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} a_{l,k} A(\phi_{l,k}, \theta_{l,k}) = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} a_{l,k} a(w_{1,l,k}) a^T(w_{2,l,k}) = \frac{1}{\sqrt{L}} A_{w_{1,k}} H_{a,k} A^T_{w_{2,k}}, \] (6)

where

\[ A_{w_{1,k}} = [a(w_{1,1,k}), a(w_{1,2,k}), \cdots, a(w_{1,L,k})], \]
\[ A_{w_{2,k}} = [a(w_{2,1,k}), a(w_{2,2,k}), \cdots, a(w_{2,L,k})], \]
\[ H_{a,k} = \text{diag}(a_{1,k}, a_{2,k}, \cdots, a_{L,k}), \] (7)

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

The \((m,n)\)th element of the channel matrix \( H_k \) can be written as

\[ H_{k,m,n} = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} a_{l,k} e^{j(mw_{1,l,k} + nw_{2,l,k})}, \] (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 \( \{ |a_{l,k}| \}_{l=2}^{K} \) 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 \( H_k \) 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 \( H_k \), 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 only will be sent once at the beginning of the transmission. The received signal at the BS can be written as

\[ Y_{BB} = F_{BB}^H F_{RF}^H \sum_{k=1}^{K} \text{vec}(H_k) x_k^T + N, \] (9)

where \( x_k = [x_k, x_{k,2}, \cdots, x_{k,T}]^T \) is the preamble of the \( k \)th user, \( T \geq K \) is the length of preamble, and \( Y_{BB} \) 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 \( k \)th 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 \( \text{vec}\{H_k\} \) as

\[
\text{vec}\{H_k\} = [\text{vec}^H\{H_k\}(0), \cdots, \text{vec}^H\{H_k\}(D-1)]^H,
\]

(10)

where

\[
\text{vec}\{H_k\}(p) = \text{vec}\{H_k\}(pM_{RF}):(p+1)M_{RF}-1).
\]

(11)

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

\[
F_{RF}(p) = \begin{bmatrix}
0_{(pM_{RF}) \times M_{RF}} \\
U_{RF}^T_{(MN-(p+1)M_{RF}) \times M_{RF}}
\end{bmatrix}_{MN \times M_{RF}},
\]

(12)

where \( U_{RF} \) is an \( M_{RF} \times M_{RF} \) Hadamard matrix, and its elements are of unitary magnitude. Based on (9), the baseband signal of the \( k \)th user before digital precoding at the \( p \)th position and the \( t \)th time is written as

\[
y_{BB,k}(p, t) = (F_{RF}(p, t))^H \text{vec}\{H_k\} x_k(t) + n(p, t)
\]

\[
= U_{RF}^H \text{vec}\{H_k\}(p)x_k(t) + n_k(p, t),
\]

(13)

where \( n_k(p, t) = [n_k(t)(pM_{RF}):(p+1)M_{RF} - 1), t = 1, 2, \cdots, T, \text{ and } p = 0, 1, \cdots D - 1.\)

We can then estimate \( \text{vec}\{H_k\}(p) \) from (13) via

\[
\text{vec}\{\hat{H}_k\}(p) = (U_{RF}^H)^{-1}y_{BB,k}(p, t) - \text{vec}\{H_k\}(p) + n_k(p, t) \frac{x_k(t)}{\|x_k(t)\|^2}.
\]

(14)

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

\[
\tilde{F}_{RF}^H = \begin{bmatrix}
F_{RF}^H & 0 & \cdots & 0 \\
0 & F_{RF}^H & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & F_{RF}^D
\end{bmatrix}_{MN \times MN},
\]

(15)

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

\[
\tilde{y}_{BB,k} = \tilde{F}_{RF}^H \text{vec}\{H_k\} \tilde{x}_k + \tilde{N}_k,
\]

(16)

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

\[
\tilde{N}_k = \begin{bmatrix}
n_k(1,1) & n_k(1,2) & \cdots & n_k(1,D) \\
n_k(2,1) & n_k(2,2) & \cdots & n_k(2,D) \\
\vdots & \vdots & \ddots & \vdots \\
n_k(D,1) & n_k(D,2) & \cdots & n_k(D,D)
\end{bmatrix}.
\]

(17)

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

\[
\text{vec}\{\hat{H}_k\} = (\tilde{F}_{RF}^H)^{-1}y_{BB,k} \frac{\tilde{x}_k^T}{\|\tilde{x}_k\|^2} + (\tilde{F}_{RF}^H)^{-1}\tilde{N}_k \frac{\tilde{x}_k^T}{\|\tilde{x}_k\|^2}.
\]

(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 \( \text{vec}\{H_k\} \). However, it is not enough to estimate the complete channel. Define \( X = [x_1, x_2, \cdots, x_K]^T \) as the orthogonal training matrix, and \( \|x_k\|^2 = 1 \). We can send the same orthogonal training matrix for \( D \) times to estimate the whole channel matrix. Similar to (16), we have

\[
\tilde{Y}_{BB} = \tilde{F}_{RF}^H \tilde{H} \tilde{X} + \tilde{N},
\]

(19)

where

\[
\tilde{H} = [\text{vec}\{H_1\}, \text{vec}\{H_2\}, \cdots, \text{vec}\{H_K\}],
\]

\[
\tilde{X} = \left[\begin{array}{c}
x_1 \\
\vdots \\
x_K
\end{array}\right], \quad \tilde{N} = \left[\begin{array}{c}
\tilde{N}_1 \\
\vdots \\
\tilde{N}_K
\end{array}\right].
\]

(20)

Thus, the channel for all users can be estimated from

\[
\tilde{H} = (\tilde{F}_{RF}^H)^{-1} \tilde{Y}_{BB} \tilde{X}^T + (\tilde{F}_{RF}^H)^{-1} \tilde{N} \tilde{X}^T.
\]

(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 \( F_M \) and \( F_N \), whose elements are given by \( F_M_{pq} = \frac{1}{M} e^{-j \frac{2\pi pq}{M}}, p, q = 0, 1, \cdots, M - 1 \) and \( F_N_{qq'} = \frac{1}{N} e^{-j \frac{2\pi q q'}{N}}, q, q' = 0, 1, \cdots, N - 1 \), respectively. Meanwhile, we define the normalized 2D-DFT of the channel matrix \( H \) as \( H_{DFT} = F_M H F_N \), whose \((p, q)\)th element
\( p = 0, 1, \ldots, M - 1; q = 0, 1, \ldots, N - 1 \) is computed as

\[
[H_{DFT}]_{pq} = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [H]_{pn} e^{-j2\pi \frac{m}{M} (\frac{p}{M} + \frac{n}{N})}
\]

\[
= \frac{1}{\sqrt{LMN}} \sum_{l=1}^{L} a_l e^{-j\frac{Mw_1}{2}} \left( e^{-j\frac{\pi p - Mw_1}{2}} e^{-j\frac{\pi q - Nw_2}{2}} - \frac{\sin((\pi p - Mw_1)/M)}{\sin((\pi q - Nw_2)/N)} \right)
\]

\[
\times \frac{\sin(\pi p - Mw_1/2)}{\sin((\pi q - Nw_2)/2)}/(\sin((\pi p - Mw_1)/M) \sin((\pi q - Nw_2)/N)). \tag{22}
\]

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 \( p_l = \frac{Mw_1}{2} \) and \( q_l = \frac{Nw_2}{2} \) such that \( [H_{DFT}]_{pq} = \frac{e^{-j\phi}}{\sqrt{LMN}} \) while the other elements of \( H_{DFT} \) are all zero. In other words, all power is concentrated on the \( (p_l, q_l) \) elements and the elements of \( H_{DFT} \) possess sparse property, such that the elevation and the azimuth DOA of the \( l \)th path \((\phi_l, \theta_l)\) can be immediately estimated from the non-zero positions \((p_l, q_l)\) of \( H_{DFT} \) using

\[
\phi_l = \cos^{-1} \left( \frac{\lambda_{p_l}}{M_d} \right),
\]

\[
\theta_l = \cos^{-1} \left( \frac{\lambda_{q_l}}{N_d} \sqrt{1 - \left( \frac{\lambda_{p_l}}{M_d} \right)^2} \right). \tag{23}
\]

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 \( Mw_1/(2\pi) \) and \( Nw_2/(2\pi) \) are integers, all power of channel will concentrate on some separated singular points. We call these as on-grid angles. In more general case, \( Mw_1/(2\pi) \) and \( Nw_2/(2\pi) \) will not be integers, and the channel power of \( H_{DFT} \) will leak from the \((\lfloor Mw_1/2\pi \rfloor, \lfloor Nw_2/2\pi \rfloor)\)th element to its nearby elements. In fact, the leakage of channel power is positively proportional to the deviation \( (Mw_1/(2\pi) - \lfloor Mw_1/2\pi \rfloor) \) and \( (Nw_2/(2\pi) - \lfloor Nw_2/2\pi \rfloor) \), but is inversely proportional to \( M \) and \( N \) as shown in (22). However, \( H_{DFT} \) can still be approximated as a sparse matrix with most of power concentrated around the \((\lfloor Mw_1/2\pi \rfloor, \lfloor Nw_2/2\pi \rfloor)\)th element. Hence, the peak power position of \( H_{DFT} \) is still useful for extracting initial DOA information.

An example of a two-path channel from \((30^\circ, 140^\circ)\) and \((-50^\circ, 10^\circ)\) 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 noiseless 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 \( H \), with its \((p, q)\)th element being

\[
[H_{DFT}]_{pq} = [H_{DFT}]_{pq} + [N_{DFT}]_{pq}, \tag{24}
\]

where \( N_{DFT} \sim CN \left( 0, \frac{\sigma^2}{MN} I \right) \). Denote the \( L \) largest peaks in \( L \) bins of \( H_{DFT} \) as \((p_1^{ini}, q_1^{ini})\). We can express the initial DOA estimates as

\[
\hat{\phi}_l^{ini} = \cos^{-1} \left( \frac{\lambda_{p_l}^{ini}}{M_d} \right),
\]

\[
\hat{\theta}_l^{ini} = \cos^{-1} \left( \frac{\lambda_{q_l}^{ini}}{N_d} \sqrt{1 - \left( \frac{\lambda_{p_l}^{ini}}{M_d} \right)^2} \right). \tag{25}
\]

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

\[
H^{ro} = \Phi_M(\Delta \phi_l) H \Phi_N(\Delta \theta_l), \tag{26}
\]

where the diagonal matrices \( \Phi_M(\Delta \phi_l) \) and \( \Phi_N(\Delta \theta_l) \) are given by

\[
\Phi_M(\Delta \phi_l) = \text{diag}\{1, e^{j\Delta \phi_l}, \ldots, e^{j(M-1)\Delta \phi_l}\},
\]

\[
\Phi_N(\Delta \theta_l) = \text{diag}\{1, e^{j\Delta \theta_l}, \ldots, e^{j(N-1)\Delta \theta_l}\}. \tag{27}
\]

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

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

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

where \((\Delta \phi_l, \Delta \theta_l)\) in (29) are the optimal angle shifts.
Based on the derived optimal angle shifts, the elevation angle and the azimuth angle of the lth path \((\phi_l, \theta_l)\) can be estimated as

\[
\phi_l = \cos^{-1}\left(\frac{\lambda \pi}{M d} - \frac{\lambda \Delta \phi_l}{2 \pi d}\right) \quad \text{and} \quad \theta_l = \cos^{-1}\left(\frac{\lambda \pi}{N d} - \frac{\lambda \Delta \theta_l}{2 \pi d}\right)/\sqrt{1 - \left(\frac{\lambda \pi}{M d} - \frac{\lambda \Delta \phi_l}{2 \pi d}\right)^2}.
\]

For finding the optimal angle shifts \((\Delta \phi_l, \Delta \theta_l)\) from channel matrix \(H\), one way is the simple two-dimensional searching of \(\Delta \phi\) and \(\Delta \theta\) over the very small region \(\Delta \phi \in [-\pi, \pi]\) and \(\Delta \theta \in [-\pi, \pi]\). Then we can extract the corresponding \((\Delta \phi_l, \Delta \theta_l)\) when the \((p_{ini}^{\text{th}}, q_{ini}^{\text{th}})\)th element of \(F_M \hat{\Phi}_M(\Delta \phi_l) \hat{H}_{PN}(\Delta \theta_l) F_N\) shrink into their highest form. Mathematically, there is

\[
(\Delta \phi_l, \Delta \theta_l) = \arg \max_{\Delta \phi \in [-\pi, \pi], \Delta \theta \in [-\pi, \pi]} \|F_M H_{p_{ini}^{\text{th}}}^{H} \hat{\Phi}_M(\Delta \phi_l) \hat{H}_{PN}(\Delta \theta_l) F_N q_{ini}^{\text{th}}\|^2,
\]

where \(F_M H_{p_{ini}^{\text{th}}}^{H}\) is the \(p_{ini}^{\text{th}}\)th column of \(F_M\) and \(F_N q_{ini}^{\text{th}}\) is the \(q_{ini}^{\text{th}}\)th column of \(F_N\).

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 \([-\frac{\pi}{2}, \frac{\pi}{2}]\) and \(G_N\) means the search grids within \([-\pi, \pi]\), 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 \(O(MN \log MN + MN + GKMN)\). The complexity of different algorithms is shown in the Table I. Since \(K \ll MN\) and \(G \ll MN^2\), the complexity of the proposed algorithm is much less than \(O(M^2 N^2)\).

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

**Input:** \(H\). **Output:** \(\phi_l, \theta_l, l = 1, 2, \ldots, L\).

1. Find the central point \((p_{ini}^{\text{th}}, q_{ini}^{\text{th}})\) of each bin in \(H_{DFT} = F_M H F_N\), where \((p_{ini}^{\text{th}}, q_{ini}^{\text{th}}) = \arg \max_{(p,q) \in \text{bin of } H_{DFT}} \|F_M H_{p_{ini}^{\text{th}}}^{H} \hat{\Phi}_M(\Delta \phi) \hat{H}_{PN}(\Delta \theta) F_N q_{ini}^{\text{th}}\|^2, l = 1, 2, \ldots, L\).

2. For \(l = 1 : L\):

3. For \(l = 1 : L\):

4. \((\Delta \phi_l, \Delta \theta_l) = \arg \max_{\Delta \phi \in [-\pi, \pi], \Delta \theta \in [-\pi, \pi]} \|F_M H_{p_{ini}^{\text{th}}}^{H} \hat{\Phi}_M(\Delta \phi_l) \hat{H}_{PN}(\Delta \theta_l) F_N q_{ini}^{\text{th}}\|^2\).

5. End For

6. End For

7. Obtain \(\hat{\phi}_l, \hat{\theta}_l\) from the equation (30).

Remarks 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 \((\Delta \phi_{l,k}, \Delta \theta_{l,k})\), \(l = 1, 2, \ldots, L\), \(k = 1, 2, \ldots, K\). The received signal of the BS can be written as

\[
Y = F_{BB}^{H} F_{HF}^{H} \sum_{k=1}^{K} \text{vec}(H_k) x_k^T + N
\]

\[
= \frac{1}{\sqrt{L}} F_{BB}^{H} F_{HF}^{H} \sum_{k=1}^{K} a_l, \text{vec}(A(\hat{\phi}_{l,k}, \hat{\theta}_{l,k})) x_k^T + N
\]

\[
= \frac{1}{\sqrt{L}} F_{BB}^{H} F_{HF}^{H} \sum_{k=1}^{K} A_k a_k x_k^T + N,
\]

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

\[
\hat{a}_k = \sqrt{T}(F_{BB}^{H} F_{HF}^{H} A_k)^{\dagger} Y x_k
\]

\[
= a_k + \sqrt{T}(F_{BB}^{H} F_{HF}^{H} A_k)^{\dagger} N x_k.
\]

Thus, with the DOAs information from (30) and gains information from (33), we can obtain the uplink channel estimation.
for all users as
\[
\hat{h}_k = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \hat{a}_{l,k} \text{vec}\{A(\hat{\phi}_{l,k}, \hat{\theta}_{l,k})\}. \tag{34}
\]

Remark where algorithm, as exactly towards its signal direction and is thus named as precoding indicates that each path of each user is transmitting to the frequency division duplex (FDD) system due to the different transmission frequencies of the uplink and downlink for all users as a common propagation space between the BS and the user. The reciprocity of the channel cannot be applied to FDD system and the channel gain component of the DOA information of both uplink and downlink are 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 $\phi, \theta$ 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
\[
y = F_{RF}^H \text{vec}(H)s + n = F_{RF}^H \text{vec}(A)\alpha s + n, \tag{38}
\]
where $A \triangleq A(w_1, w_2)$ is the $M \times N$ steering matrix with its $(p, q)$th entry given by
\[
[A]_{pq} = e^{j((p-1)w_1 + (q-1)w_2)}, \tag{39}
\]
and $n$ is the $M_{RF} \times 1$ vector representing the white Gaussian noise with zero mean and variance $\sigma_n^2$.

The proposed estimator can be rewritten as
\[
[w_1, w_2] = \arg\max_{w_1, w_2} \|\text{vec}(H)(A)(F_{RF}^H)^\dagger y\|^2
= \arg\max_{w_1, w_2} y^H F_{RF}^H \text{vec}(A) \text{vec}(H)(F_{RF}^H)^\dagger y, \tag{40}
\]
where $\text{vec}(A) = \{\Phi_M(\Delta \phi)f_{M,1}f_{N,1}^H \Phi_N(\Delta \theta_{1,1})\}, \ldots, \{\Phi_M(\Delta \phi_{L,K})f_{M,L,K}f_{N,q_{L,K}}^H \Phi_N(\Delta \theta_{L,K})\}$.

For given $w_1, w_2$ and $\alpha$, the probability density function (PDF) of $y$ can be expressed as
\[
f(y|w_1, w_2, \alpha) = \frac{1}{(\pi \sigma_n^2)^M_{RF}} \exp\left\{-\frac{\|y - F_{RF}^H \text{vec}(A)\alpha s\|^2}{\sigma_n^2}\right\}. \tag{41}
\]

The joint ML estimates of $w_1, w_2$ and $\alpha$ can be obtained via
\[
[w_1, w_2, \hat{\alpha}] = \arg\max_{w_1, w_2, \alpha} f(y|w_1, w_2, \alpha), \tag{42}
\]
or equivalently
\[
[w_1, w_2, \hat{\alpha}] = \arg\min_{w_1, w_2, \alpha} \|y - F_{RF}^H \text{vec}(A)\alpha s\|^2. \tag{43}
\]

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

For given $w_1$ and $w_2$, the ML estimate of $\alpha$ is obtained from (43) as
\[
\hat{\alpha} = \text{vec}(A)(F_{RF}^H)^\dagger s^*y. \tag{44}
\]

2This is a key difference from beamspace transmission.
Substituting (44) into (43), the ML estimates of \( w_1, w_2 \) can be written as

\[
\begin{aligned}
[\hat{w}_1, \hat{w}_2] &= \arg\min_{w_1, w_2} \| y - \mathbf{F}_{RF}^H \text{vec}(A) \text{vec}(A)(\mathbf{F}_{RF}^H)^\dagger s^* s \|^2 \\
&= \arg\min_{w_1, w_2} \| y - \sigma_s^2 \mathbf{F}_{RF}^H \text{vec}(A) \text{vec}(A)(\mathbf{F}_{RF}^H)^\dagger y \|^2 \\
&= \arg\max_{w_1, w_2} y^H \mathbf{F}_{RF}^H \mathbf{P}_A (\mathbf{F}_{RF}^H)^\dagger y \\
&= \arg\max_{w_1, w_2} g(w_1, w_2),
\end{aligned}
\]

(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 \( \mathbf{P}_A = \text{vec}(A)(\mathbf{F}_{RF}^H) \) represents the projection matrix onto the subspace spanned by \( \text{vec}(A) \). 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 \( \alpha, w_1, \) and \( w_2 \).

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

\[
E\{\Delta w_i^2\} = \frac{\sigma_n^2}{2\|\alpha\|^2 \text{vec}(A)|\mathbf{W}_1 \mathbf{P}_a \mathbf{W}_1 \text{vec}(A)}
\]

(46)

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

\[
\mathbf{W}_1 = \text{Diag}\{0, \ldots, 0, \ldots, (M-1), \ldots, (M-1)\},
\]

(47)

\[
\mathbf{W}_2 = \text{Diag}\{0, \ldots, 0, \ldots, (N-1), \ldots, (N-1)\}.
\]

(48)

**Proof.** See Appendix.

In the channel estimation process, we need to further examine the MSE of the azimuth angle \( \phi \) and the elevation angle \( \theta \). Based on the fact that \( w_1 = \frac{\lambda \omega_1}{2\pi d} \cos \phi, w_2 = \frac{\lambda \omega_2}{2\pi d} \sin \phi \cos \theta \), we have

\[
\phi = \cos^{-1}\left(\frac{\lambda \omega_1}{2\pi d}\right), \quad \text{and} \quad \theta = \cos^{-1}\left(\frac{\lambda \omega_2}{2\pi d \sin \phi}\right).
\]

(49)

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

\[
\begin{aligned}
E\{\Delta \phi\} &= E\{\Delta \theta\} = 0, \\
E\{\Delta \phi^2\} &= E\{[(\hat{\phi} - \phi)\hat{\phi} - \phi]^H\} = \frac{\partial^2 \hat{\phi}}{\partial w_1^2} E\{w_1^2\} \left(\frac{\partial^2 \hat{\phi}}{\partial w_1^2}\right)^H \\
&= \frac{\lambda^2}{(2\pi d)^2} \frac{\sigma_n^2}{\sigma_s^2} \left(1 - \frac{\lambda \omega_1}{2\pi d}\right)^2 \times \frac{1}{2\|\alpha\|^2 \text{vec}(A)|\mathbf{W}_1 \mathbf{P}_a \mathbf{W}_1 \text{vec}(A)}, \\
E\{\Delta \theta^2\} &= E\{[(\hat{\theta} - \theta)\hat{\theta} - \theta]^H\} = \frac{\partial^2 \hat{\theta}}{\partial w_2^2} E\{w_2^2\} \left(\frac{\partial^2 \hat{\theta}}{\partial w_2^2}\right)^H \\
&= \frac{\lambda^2}{(2\pi d \sin \phi)^2} \frac{\sigma_n^2}{\sigma_s^2} \left(1 - \frac{\lambda \omega_2}{2\pi d \sin \phi}\right)^2 \times \frac{1}{2\|\alpha\|^2 \text{vec}(A)|\mathbf{W}_2 \mathbf{P}_a \mathbf{W}_2 \text{vec}(A)}.
\end{aligned}
\]

(50)

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

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

(51)

where \( \text{vec}(\hat{A}) \) is constructed from the estimate \( (\hat{w}_1, \hat{w}_2) \).

With the help of Taylor’s expansion, \( \text{vec}(\hat{A}) \) can be approximated by

\[
\text{vec}(\hat{A}) \approx \text{vec}(A) + j\text{vec}(\hat{A})\mathbf{W}_1 \Delta w_1, \quad i = 1, 2.
\]

(52)

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

\[
\hat{\alpha} = \sigma_s^2 (\text{vec}(A) + j\text{vec}(\hat{A})\mathbf{W}_1 \Delta w_1) \text{vec}(A) \alpha + \text{vec}(\hat{A})(\mathbf{F}_{RF}^H)^\dagger s^* n.
\]

(53)

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

\[
\begin{aligned}
E\{\Delta \alpha\} &= E\{j\text{vec}(\hat{A})\mathbf{W}_1 \text{vec}(A) \Delta w_1 \alpha + j\text{vec}(\hat{A})(\mathbf{F}_{RF}^H)^\dagger s^* n\} = 0, \\
E\{\Delta \alpha^2\} &= E\{((\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha))^H\} \\
&= \alpha E\{\{\Delta w_1 \}^2 \alpha^H \text{vec}(A) W_1 \text{vec}(A)\} \\
&= \sigma_s^2 \text{vec}(\hat{A})(\mathbf{F}_{RF}^H)^\dagger E\{m \text{vec}(A)\} (\mathbf{F}_{RF}^H)^\dagger \text{vec}(A) \\
&= \frac{\sigma_n^2 \text{vec}(\hat{A}) \text{vec}(A) |\mathbf{W}_1 \mathbf{P}_a \mathbf{W}_1 \text{vec}(A)} + \sigma_n^2.
\end{aligned}
\]

(54)

In (54), the first term is caused by the estimation error in \( (\phi, \theta) \), while the second part is caused by the noise only. If \( (\phi, \theta) \) are perfectly estimated, \( E\{\Delta \alpha^2\} \) 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

\[
\text{MSE}(\hat{\alpha}) = \frac{\sigma_n^2 \text{vec}(\hat{A}) \text{vec}(A) |\mathbf{W}_1 \mathbf{P}_a \mathbf{W}_1 \text{vec}(A)} + \sigma_n^2.
\]

(55)

From (50) and (54), we know that the joint ML estimator is unbiased for both \( (\phi, \theta) \) and \( \alpha \). 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 \( Y \) can be expressed as

\[
Y = \mathbf{H}s + \mathbf{N} = \alpha \mathbf{A}(\phi, \theta) s + \mathbf{N} = \alpha \mathbf{A}(w_1) \alpha^T (w_2) s + \mathbf{N}.
\]

(56)

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

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

(57)

where the real part of the received signal is

\[
y_{m,n}^R = \Re\{y_{m,n}\} = \Re\{\alpha e^{j(m-1)w_1 + (n-1)w_2} s\} + \Re\{n_{m,n}\} = y_{m,n} + n_{m,n}^R.
\]

(58)
\[ I(\Theta) = \frac{1}{2\sigma^2} \begin{bmatrix}
MN & 0 & 0 \\
0 & (2\pi^2)N\sum_{m=0}^{M-1} m^2 & (2\pi^2)\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n \\
0 & (2\pi^2)\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n & (2\pi^2)(2M)^2\sum_{n=0}^{N-1} n^2
\end{bmatrix}. \] (62)

\[ \text{var}(\hat{\alpha}) \geq A, \]
\[ \text{var}(\hat{u}_1) \geq (2\pi^2)(N\sum_{m=0}^{M-1} m^2)[M\sum_{n=0}^{N-1} n^2] - (2\pi^2)[M\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n][M\sum_{n=0}^{N-1} n] = AB \frac{6(2N-1)}{M-1}, \]
\[ \text{var}(\hat{u}_2) \geq (2\pi^2)(N\sum_{m=0}^{M-1} m^2)[M\sum_{n=0}^{N-1} n^2] - (2\pi^2)[M\sum_{m=0}^{M-1} m\sum_{n=0}^{N-1} n][M\sum_{n=0}^{N-1} n] = AB \frac{6(2M-1)}{N-1}. \] (63)

For given \( \alpha, \phi, \) and \( \theta \), the probability density function of \( Y \) can be expressed as
\[ f(Y|\alpha, \phi, \theta) = \frac{1}{(2\pi\sigma^2)^{2M+2N}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{m=1}^{M} \sum_{n=1}^{N} (y_{m,n} - x_{m,n})^2 \right\}. \] (59)

Let us define \( \Theta = [\alpha, w_1, w_2]^T \) as the unknown parameter vector. The Fisher information matrix (FIM) is defined as
\[ [I(\Theta)]_{i,j} = -E\left[ \frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_i \partial \Theta_j} \right], \] (60)
where
\[ \ln f(Y|\alpha, \phi, \theta) = \frac{MN}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{m=1}^{M} \sum_{n=1}^{N} n^2 m_{m,n}, \] (61)
and \( \sigma^2 = \sigma_n^2/\sigma_\alpha^2 \).

**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. \( \square \)

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

**Lemma 3.** The CRLB of the azimuth and elevation angle can be expressed as
\[ \text{var}(\hat{\theta}) \geq AB \frac{6(2N-1)^2}{(M-1)d^2 \sin^2 \theta}, \] (64)
\[ \text{var}(\hat{\phi}) \geq AB \frac{6\lambda^2(M-1)(2M-1) + 6(N-1)Cd \cos \theta \cos \phi}{(M-1)(N-1)d^2 \sin^2 \theta \sin^2 \phi}, \] (65)
where \( C = 3\lambda(M-1)(N-1) + (N-1)(2N-1)d \cos \theta \cos \phi \).

**Proof.** For DOA estimation of hybrid mmWave massive MIMO system, the performance of azimuth angle \( \theta \) and elevation angle \( \phi \) are required. Therefore, we can use the following transformation to estimate the real angles of azimuth and elevation:
\[ g(\Theta) = \begin{bmatrix}
\alpha \\
\theta \\
\phi
\end{bmatrix} = \begin{bmatrix}
\arccos\left(\frac{\lambda_{in}}{d}\right) \\
\arcsin\left(\frac{\lambda_{in}}{d\sin \gamma}\right)
\end{bmatrix}. \] (66)

Then, we can obtain the CRLBs of the azimuth and elevation angle of hybrid mmWave massive MIMO system through the following Jacobian matrix:
\[ \text{var}(\hat{\theta}) \geq \left| \frac{\partial g(\Theta)}{\partial \Theta} I^{-1}(\Theta) \frac{\partial g(\Theta^T)}{\partial \Theta} |_{2,2}, \right. \]
\[ \text{var}(\hat{\phi}) \geq \left| \frac{\partial g(\Theta)}{\partial \Theta} I^{-1}(\Theta) \frac{\partial g(\Theta^T)}{\partial \Theta} |_{3,3}. \] (66)

Next, the CRLBs of the azimuth and elevation angle can be expressed as
\[ \text{var}(\hat{\theta}) \geq AB \frac{6(2N-1)^2}{(M-1)d^2 \sin^2 \theta}, \]
\[ \text{var}(\hat{\phi}) \geq AB \frac{6\lambda^2(M-1)(2M-1) + 6(N-1)Cd \cos \theta \cos \phi}{(M-1)(N-1)d^2 \sin^2 \theta \sin^2 \phi}, \] (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 \( \tau \) 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...
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 $\rho$ 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 $3000 \times 3000$ and the proposed estimation algorithm with $100 \times 100$ 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 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 $M N$. 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. 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.

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.

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 prefect 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.
A. Proof of Lemma 1

Let us define
\[ y_d = F^H_{RF} \text{vec}(A) \alpha s. \]  
(68)

The received signal \( y = y_d + n \) is actually a perturbed version of \( y_d \). At high SNR, the first derivative of the cost function can be approximated using Taylor's expansion as
\[ 0 = \frac{\partial g(w_1, w_2)}{\partial w_i} |_{w_i=\hat{w}_i} \approx \frac{\partial g(w_1, w_2)}{\partial w_i} |_{w_i=\hat{w}_i} + \frac{\partial^2 g(w_1, w_2)}{\partial^2 w_i} |_{w_i=\hat{w}_i} \Delta w_i, \]  
(69)

where \( \Delta w_i = \hat{w}_i - w_i \) is the perturbation in the \( w_i \). Thus, \( \Delta w_i \) can be represented as
\[ \Delta w_i = -\frac{\partial g(w_1, w_2)}{\partial w_i} |_{w_i=\hat{w}_i} \frac{\partial^2 g(w_1, w_2)}{\partial^2 w_i} |_{w_i=\hat{w}_i} \hat{w}_i. \]  
(70)

In (70), the first order derivative can be calculated as
\[ \hat{g}(w_1, w_2 | w_i) = jy^H F^H_{RF} W_i P^\perp_a (F^H_{RF})^i y - jy^H F^H_{RF} P^\perp_a W_i (F^H_{RF})^i y. \]  
(71)

Since \( P^\perp_a (F^H_{RF})^i y = 0 \), we can rewrite (71) as
\[ \hat{g}(w_1, w_2 | w_i) = jy^H F^H_{RF} W_i P^\perp_a (F^H_{RF})^i (y_d + n) - jy^H F^H_{RF} P^\perp_a W_i (F^H_{RF})^i (y_d + n) = -23 \{ y^H F^H_{RF} W_i P^\perp_a (F^H_{RF})^i n \} \]  
(72)

Due to the independence between \( y_d \) and \( n \), we can obtain
\[ \text{E}\{\hat{g}(w_1, w_2 | w_i)\} = -2\sigma_n^2 \text{E}\{\text{vec}(A)^H W_i P^\perp_a (F^H_{RF})^i \} = 0. \]  
(73)

The second-order derivative can be calculated as
\[ \hat{g}(w_1, w_2 | w_i) = -y^H F^H_{RF} W_i^2 P^\perp_a (F^H_{RF})^i n + y^H F^H_{RF} W_i P^\perp_a W_i (F^H_{RF})^i y + y^H F^H_{RF} P^\perp_a W_i (F^H_{RF})^i y - n^H F^H_{RF} P^\perp_a W_i (F^H_{RF})^i y. \]  
(74)

Based on (74), we have
\[ \text{E}\{\hat{g}(w_1, w_2 | w_i)\} = 2y^H F^H_{RF} W_i P^\perp_a W_i (F^H_{RF})^i y_d + \text{E}\{\text{vec}(A)^H \text{vec}(A)\} = 2\sigma_n^2 \text{vec}(A)^H W_i P^\perp_a W_i \text{vec}(A). \]  
(75)

Therefore, \( \hat{g}(w_1, w_2 | w_i) \) can be rewritten as
\[ \hat{g}(w_1, w_2 | w_i) = \text{E}\{\hat{g}(w_1, w_2 | w_i)\} + O_2(n) + O_2(n^2), \]  
(76)

where \( O_2(n) \) and \( O_2(n^2) \) represent the linear and quadrature functions of \( n \) existing in \( \hat{g}(w_1, w_2 | w_i) \). Similarly, \( \hat{g}(w_1, w_2 | w_i) \) can be expressed as
\[ \hat{g}(w_1, w_2 | w_i) = O_1(n) + O_1(n^2), \]  
(77)

where \( O_1(n) \) and \( O_1(n^2) \) represent the linear and quadrature functions of \( n \) existing in \( \hat{g}(w_1, w_2 | w_i) \). Substituting (76) and (77) into (70) and assuming high SNR, i.e., \( \|n\|^2 \ll \|y_d\|^2 \), we have
\[ \Delta w_i = -\frac{O_1(n) + O_1(n^2)}{\text{E}\{\hat{g}(w_1, w_2 | w_i)\} + O_2(n) + O_2(n^2)} \]  
(78)

Knowing that \( O_1(n) + O_1(n^2) \) and \( O_2(n) + O_2(n^2) \) are small terms at high SNR, their product can be ignored, and \( \Delta w_i \) can be approximated as
\[ \Delta w_i \approx -\frac{O_1(n) + O_1(n^2)}{\text{E}\{\hat{g}(w_1, w_2 | w_i)\}} = -\frac{\hat{g}(w_1, w_2 | w_i)}{\text{E}\{\hat{g}(w_1, w_2 | w_i)\}}. \]  
(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
\[ \text{E}\{\Delta w_i\} = \text{E}\left\{ -\frac{\hat{g}(w_1, w_2 | w_i)}{\text{E}\{\hat{g}(w_1, w_2 | w_i)\}} \right\} = \text{E}\{\hat{g}(w_1, w_2 | w_i)\} \]  
\[ \text{E}\{\Delta w_i^2\} = \text{E}\left\{ -\frac{\hat{g}(w_1, w_2 | w_i)}{\text{E}\{\hat{g}(w_1, w_2 | w_i)\}} \right\}^2 = \frac{\text{E}\{\hat{g}(w_1, w_2 | w_i)\}^2}{\text{E}\{\hat{g}(w_1, w_2 | w_i)\}}, \]  
(80)

respectively. To derive the variance of DOA information, we express
\[ \text{E}\{\hat{g}(w_1, w_2 | w_i)\}^2 = 2\text{E}\{y_d^H F^H_{RF} W_i P^\perp_a (F^H_{RF})^i n n^H F^H_{RF} P^\perp_a W_i (F^H_{RF})^i y_d \} + \text{E}\{\text{vec}(A)^H \text{vec}(A)\} \]  
(81)

where the second term can be ignored for high SNR to obtain
\[ \text{E}\{\hat{g}(w_1, w_2 | w_i)\}^2 = 2\sigma_n^2 \text{vec}(A)^H W_i P^\perp_a W_i (F^H_{RF})^i y_d = 2\sigma_n^2 \text{vec}(A)^H W_i P^\perp_a W_i \text{vec}(A). \]  
(82)

Substituting (73), (75), (82) into (80), we obtain the mean and MSE of estimation \( \hat{w}_i \) as
\[ \text{E}\{\Delta w_i\} = 0, \]  
\[ \text{E}\{\Delta w_i^2\} = \frac{\sigma_n^2}{2\sigma_n^2 |\alpha|^2 \text{vec}(A)^H W_i P^\perp_a W_i \text{vec}(A)}. \]  
(83)

The proof is completed.

B. Proof of Lemma 2

Performing mathematical calculation on (60), we obtain
\[ \frac{\partial^2 \ln f(Y | \alpha, \phi, \theta)}{\partial^2 \Theta_i} = -\frac{M N}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \cos(4\pi m w_1 + 4\pi n w_2), \]  
(84)
\[
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_1 \partial \Theta_2} = -\frac{(2\pi \alpha)^2}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[ m^2 \sin(4\pi mw_1 + 4\pi nw_2) \right], \\
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_2 \partial \Theta_1} = -\frac{(2\pi \alpha)^2}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[ n^2 \cos(4\pi mw_1 + 4\pi nw_2) \right], \\
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_3 \partial \Theta_1} = -\frac{2\pi \alpha}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} n \sin(4\pi mw_1 + 4\pi nw_2), \\
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_3 \partial \Theta_2} = -\frac{2\pi \alpha}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} m \sin(4\pi mw_1 + 4\pi nw_2), \\
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_3 \partial \Theta_3} = -\frac{(2\pi \alpha)^2}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[ mn \cos(4\pi mw_1 + 4\pi nw_2) - 1 \right]. 
\]

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

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

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

\[
\lim_{K \to \infty} \sum_{k=1}^{K} \sin(4\pi k x) = \frac{\sin((K + 1)2\pi x) \sin(2\pi K x)}{\sin(2\pi x)} = 0, \\
\lim_{K \to \infty} \sum_{k=1}^{K} \cos(4\pi k x) = \frac{\cos((K + 1)2\pi x) \sin(2\pi K x)}{\sin(2\pi x)} = 0, \\
\lim_{K \to \infty} \sum_{k=1}^{K} k \sin(4\pi k x) = \frac{\sin 4\pi K x}{4 \sin^2(2\pi x)} - \frac{K \cos((2K - 1)2\pi x)}{2 \sin(2\pi x)} = 0, \\
\lim_{K \to \infty} \sum_{k=1}^{K} k \cos(4\pi k x) = \frac{K \sin((2K - 1)2\pi x)}{2 \sin(2\pi x)} - \frac{1 - \cos 4\pi K x}{4 \sin^2(2\pi x)} = 0. 
\]
Yanhua Deng (S’13-M’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 TRANSACTIONS ON WIRELESS COMMUNICATIONS 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.

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.

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 Education, 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.