## Angle Domain Channel Estimation in Hybrid MmWave Massive MIMO Systems

<table>
<thead>
<tr>
<th>Journal:</th>
<th>IEEE Transactions on Wireless Communications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>Paper-TW-Mar-18-0349.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Transactions Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>08-Jul-2018</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Fan, Dian; Beijing Jiaotong University, Gao, Feifei; Tsinghua University, Department of Automation; Jacobs University Bremen, School of Engineering and Science Liu, Yuanwei; Queen Mary University of London, School of Electric Engineering and Computer Science Deng, Yansha; King's College London, Department of Informatics; Wang, Gongpu; School of Computer and Information Technology, Beijing Jiaotong University; Zhong, Zhangdui; Beijing Jiaotong University, Electronics and Information Engineering Nallanathan, Arumugam; Queen Mary University of London, EECS;</td>
</tr>
<tr>
<td>Keyword:</td>
<td></td>
</tr>
</tbody>
</table>
Responses for the Reviewers’ Comments

The authors would like to first thank the editor and the reviewers for the time and effort they have put into our paper. We have carefully read all of your comments and have made necessary modifications to our revised manuscript, in which the modified portions are highlighted in blue. We believe the presentation of the paper has been improved, thanks to your comments and contributions. The reviewers comments and our respective responses are listed below.

EDITOR

DETAILED COMMENTS

Comment

Specifically, although the theoretical analysis of the paper appears to be sound, I have some concerns on the novelty and contributions of this paper with respect to existing literature. For example, the proposed 2D-DFT based DOA estimation seems to be a direct extension of the beamspace channel representation proposed in [14,23] (and many other references) with 2D UPA antenna array. Moreover, the proposed DOA refinement algorithm (Algorithm 1) is not very clear. The authors claim that the proposed method can obtain the true physical angle information instead of an approximation of the quantized angel with limited resolution, which is rather questionable.

Response

Before starting our point-to-point response to editor, we would like to express our appreciation for his/her judicious reading. These remarks will help improving the quality of our paper and will also guide our future research. Many thanks.

The principle of the beamspace channel estimation method is that using DFT to estimate channel and it is a very rough on-grid approach, which always suffer from a performance loss due to the leakage of energy over some DFT bins. Therefore, many people begin to use the off-grid channel estimation approach. For example, [1] introduced an off-grid model for downlink channel sparse representation with arbitrary 2D-array antenna geometry, and propose an efficient off-grid approach for the sparse channel recovery. [2] proposed an angle domain off-grid channel estimation algorithm for the uplink mmWave massive MIMO systems. [3] proposed a phase shifter network-based precoding structure to solve the power leakage problem. We are also in this routine and propose a method to overcome the on-grid power leakage by angle rotation. Our
The proposed method does not use the channel covariance matrices and do not relay on the grid-of-beam. Moreover, using the 2D-DFT and angle rotation, we can obtain the true physical angle information. Even with limited rotation grids, we can get a more accurate angle information.

We apologize for not making the part of algorithm 1 very clear. The corresponding part has been revised in the new manuscript for better clarity.

**REVIEWER 1**

**DETAILED COMMENTS**

**Comment**

This paper investigates the uplink channel estimation for mmWave MIMO systems with large planar array. Specifically, the authors first estimate the initial path DOAs by exploiting the sparsity of mmWave channel in the angular domain. Then, they refine the estimated path DOAs by using the angle rotation technique. After obtaining the path DOAs, the path gains are estimated by the classical LS algorithm with low pilot overhead. The authors also analyze the MSE performance and the CRLB of the proposed scheme. Generally speaking, the performance analysis in this paper seems solid. However, the presentation is not satisfied and some statements are not correct. The reviewer has the following comments in details.

**Response**

We would like to first thank this reviewer for his/her judicious reading and recognition of our contributions.

**Comment**

1) Some statements in this paper are not accurate. For example: i) page 2, line 22, RF chain includes not only ADC/DAC but also other analog modules such as mixer; ii) page 4, line 14, the hybrid precoding system is not your contribution. You cannot use the word proposed; iii) page 4, line 20, the path of mmWave channel can be described by the angle and gain in nature. It is not your contribution; iv) page 5, line 36, the antenna spacing in the horizon and vertical should be different parameters although they may have the same value.

**Response**
Thank you for your comment. In the revised manuscript, we have modify the corresponding parts. i) The RF chains consist of A/D converters, D/A converters mixers and power amplifiers; ii) We have delated the word proposed; iii) The parameters of each paths in the channel matrix are decomposed into the corresponding channel gain and the DOA information for better estimation; iv) We denote the distance between the neighboring antenna elements in both horizon and vertical as $d$.

Comment

2) Some parameters in this paper are not clearly defined. For example: i) page 5, line 31, it is not clearly which one represents the number of antennas in the horizon; ii) page 5, line 50, $\tau$ is not defined; iii) page 10, line 17, the ranges of $q$ and $p$ are not defined.

Response

Thank you for your comment. In the revised manuscript, we have modify the corresponding parts for better clarity. i) 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; ii) Here, we have rewritten the equation (1) and the corresponding element has been defined; iii) The ranges of $p$ and $q$ have been defined in the modified manuscript.

Comment

3) It is not clear what kind of hybrid precoding architecture do you consider in this paper. According to the statement in page 2 (line 29), it seems that the sub-connected architecture is adopted. However, Fig. 1 shows a fully-connected architecture.

Response

We apologize for not making this part clear. What we are considering is a fully-connected hybrid precoding architecture. We have removed the confusing part.

Comment

4) In Section II, why do you assume that the digital precoder is a vector instead of a matrix? Furthermore, the digital precoder is independent with the proposed channel estimation scheme, why do you include it in the system model?
Response

Thank you for your comment. The digital precoder should be a matrix and we modify it in the new manuscript. Through the digital precoder is independent with the proposed channel estimation scheme, we include it in the system model for maintaining the integrity of the signal transmission process.

Comment

5) How can you get Eq. 11 according to Eq. 10? If $MN$ is divisible by $N_{RF}$, we will have $D = MN$. Then from Eq. 10, we have $\text{vec}\{\hat{H}_k\} = [\text{vec}\{\hat{H}_k\}]_{0:MN}$. Then, what is Eq. 11?

Response

We apologize for not making this part clear. We use another way to explain this process. 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 based on the length of the RF chains and the number of antennas at the BS. Thus we perform the channel vector $\text{vec}\{\hat{H}_k\}$ via stacking the $M_{RF} \times 1$ subvectors as

$$\text{vec}\{\hat{H}_k\} = [\text{vec}^H\{\hat{H}_k\}^{(0)}, \text{vec}^H\{\hat{H}_k\}^{(1)}, \ldots, \text{vec}^H\{\hat{H}_k\}^{(D-1)}]^H,$$

(1)

where

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

(2)

Comment

6) Eq. 13, how can you generate an analog beamformer with both zero-elements and nonzero-elements? Do you add the switches in the analog phase shifter network?

Response

Yes, the switches are included in the analog phase shifter network. So it can generate an analog beamformer with both zero-elements and nonzero-elements.

Comment
7) Eq. 24, the authors state that they can pick $L$ largest peaks of $H_{DFT}$ to estimate the initial DOAs of $L$ paths. This is not correct. In practice, $L$ largest peaks may all belong to one strong path with strong power leakage.

Response

Thank you for your comment. We add some instructions for the meaning of the largest peaks. An example of a two paths channel from $(30^\circ, 140^\circ)$ and $(-50^\circ, 10^\circ)$ with $M = 100$, $N = 100$ as shown in Fig. 1(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 both two paths correspond 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 in each bin also contain considerable power but the power of other points are ignorable. In Fig. 1(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 to indicate the initial DOA estimation.
8) The angle rotation technique has been proposed in [29] of this paper, what is the contribution of this part?

Response

Thank you for your comment. the angle rotation technique in [29] is for decrease the support indexes of the channel. But we use the angle rotation technique for finding the true physical angle information.

Comment

9) It is not clear how to use the initial estimated DOAs in Algorithm 1. It is also not clear the resolution of \( \Delta \phi_i \) and \( \Delta \theta_i \) you used to search in Algorithm 1.

Response

We apologize for not making this part clear. The corresponding part has been revised in the new manuscript for better clarity.

For finding the optimal angle shifter \((\Delta \phi_i, \Delta \theta_i)\) from channel matrix \(\mathbf{H}\), one way is a 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 \((\Delta \phi_i, \Delta \theta_i)\) when the \((p_{i}^{\text{ini}}, q_{i}^{\text{ini}})\)th element of \(\mathbf{F}_M \Phi_M(\Delta \phi_i) \mathbf{H} \Phi_N(\Delta \theta_i) \mathbf{F}_N\) shrink into their highest form, one by one. Mathematically, there is

\[
(\Delta \phi_i, \Delta \theta_i) = \arg \max_{\Delta \phi \in [-\frac{\pi}{M}, \frac{\pi}{M}], \Delta \theta \in [-\frac{\pi}{N}, \frac{\pi}{N}]} \| \mathbf{f}_M^{p_{i}^{\text{ini}}} \Phi_M(\Delta \phi_i) \mathbf{H} \Phi_N(\Delta \theta_i) \mathbf{f}_N^{q_{i}^{\text{ini}}} \|^2,
\]

where \(\mathbf{f}_M^{p_{i}^{\text{ini}}}\) is the \(p_{i}^{\text{ini}}\)th column of \(\mathbf{F}_M\) and \(\mathbf{f}_N^{q_{i}^{\text{ini}}}\) is the \(q_{i}^{\text{ini}}\)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 \((\Delta \phi_i, \Delta \theta_i)\) for each path.

The DOA information of the 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 (3) 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. Note that $K \ll MN$ and $G \ll MN^1$, the complexity of the proposed algorithm is much less than $O(M^2N^2)$.

![Algorithm 1](image)

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

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

1. Find the central point $(p_{ini}^l, q_{ini}^l)$ of each bin in $H_{DFT} = F_M H F_N$, where $(p_{ini}^l, q_{ini}^l) = \arg \max_{(p,q) \in \text{bin}(l)} \|H_{DFT} pq\|^2, l = 1, 2, \cdots, L.$
2. For $\Delta \phi = -\frac{\pi}{M}, \Delta \theta = -\frac{\pi}{N}$
3. For $l = 1 : L$
4. $(\Delta \hat{\phi}_l, \Delta \hat{\theta}_l) = \arg \max_{\Delta \phi \in [-\frac{\pi}{M}, \frac{\pi}{M}], \Delta \theta \in [-\frac{\pi}{N}, \frac{\pi}{N}]} \|f^H_{M,p_{ini}^l} \Phi_M (\Delta \phi_l) H \Phi_N (\Delta \theta_l) f_{N,q_{ini}^l}\|^2$
5. End For
6. End For
7. Obtain $\hat{\phi}_l, \hat{\theta}_l$ from the equation (30).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed 2D-DFT and angle rotation estimate algorithm</td>
<td>$O(MN \log MN + MN + GKMN)$</td>
</tr>
<tr>
<td>Beamspace based estimate algorithm</td>
<td>$O(MN \log MN + MN)$</td>
</tr>
<tr>
<td>CS based estimate algorithm</td>
<td>$O(M^2N^2 + KMN)$</td>
</tr>
</tbody>
</table>

**Comment**

$^1$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.
10) Eq. 48, it is not clear why the system model can be represented without the analog beamformer. This makes the following analysis questionable.

Response

Thank you for your comment. It is worth noting that the MSE of the proposed estimators is irrelevant to analog beamforming. Thus, we omit analog beamforming for simplicity.

Comment

11) In Section V, the references should be cited for the compared channel estimation schemes.

Response

Thank you for your great suggestion. We have indicated the used CS channel estimate approach [4], the beamspace based estimate method [5], and the eigen-decomposition based estimate method [6] in the revised manuscript.

Comment

12) For Fig. 7 and Fig. 8, it is not clear what kind of hybrid precoding algorithm is used.

Response

Thank you for your comment. We have added a small subsection to briefly explain the precoding method which is shown in Page. 15 in the revised manuscript.

Comment

13) There are some typos in this paper. For example: i) page 3, line 27, the sentence has two verbs; ii) Eq. 2, $s(k)$ should be a scalar instead of a vector; iii) page 19, line 46, it should be Jacobian instead of Jocobian.

Response

We apologize for our carelessness. In the revised manuscript, we have modified these typos.
REVIEWER 2

DETAILED COMMENTS

Comment

This paper proposes a DOA-aided channel estimation scheme for hybrid mmWave massive MIMO systems. Specifically, DOA information is firstly estimated through a 2D-DFT and angle rotation technique. Channel gain information is then estimated based on the known DOA information. Theoretical analysis on the MSE bounds and the CRLB bound of the proposed scheme is provided. Generally, this paper is not well-written. Contributions of this paper are very limited. Detailed comments are shown as follows:

Response

Before starting our point-to-point response to this reviewer, we would like to express our appreciation for his/her judicious reading and recognizing all our contributions.

Comment

The proposed 2D-DFT based DOA estimation is a direct extension of the beamspace channel representation [14,23] (and many other references) with 2D UPA antenna array. The proposed DOA refinement algorithm (Algorithm 1) is not clear. The authors claim that their method can obtain the true physical angle information instead of an approximation of the quantized angle with limited resolution. This is not true. Algorithm 1 can only obtain a quantized angle whose resolution is limited to the angle increment of $\Delta \phi$. Moreover, $\Delta \phi = -\frac{\pi}{M} : \frac{\pi}{M}$ is not correct. The angle increment of $\Delta \phi$ should be presented. Besides, the reviewer suggests that the proposed DOA estimation method should be compared with the classical MUSIC and ESPRIT algorithms in simulations.

Response

Thanks for your great suggestion.

Beamspace channel estimation method has been recognized as existing power leakage because it only considers discrete grid points. Meanwhile, there are many papers now considering off-grid channel estimation methods to improve the performance. This is not a minor improvement, but an essential change. For example, [1] introduced an off-grid model for downlink channel
sparse representation with arbitrary 2D-array antenna geometry, and propose an efficient off-grid approach for the sparse channel recovery. [2] proposed an angle domain off-grid channel estimation algorithm for the uplink mmWave massive MIMO systems. [3] proposed a phase shifter network-based precoding structure to solve the power leakage problem. We are also in this routine and propose a method to overcome the on-grid power leakage by angle rotation. Our proposed method does not use the channel covariance matrices and do not relay on the grid-of-beam. Moreover, using the 2D-DFT and angle rotation, we can obtain the true physical angle information. Even with limited rotation grids, we can get a more accurate angle information.

We apologize for not making the part of algorithm 1 very clear. The corresponding part has been revised in the new manuscript for better clarity.

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 (3) is proportional to \( O(MN) \) for the given \( \Delta \phi_t \) and \( \Delta \theta_t \), the complexity of the whole algorithm is about \( O(MN\log MN + MN + GKN) \). The complexity of different algorithms is shown in the Table I. Note that \( K \ll MN \) and \( G \ll MN^2 \), the complexity of the proposed algorithm is much less than \( O(M^2N^2) \).

Since different users can collaborate in the communication process, the angle information of users can be obtained by sending training sequences once. The traditional blind methods such as MUSIC and ESPRIT cannot obtain angle information of users by sending training sequences. Although many people use MUSIC and ESPRIT to do angle estimation in massive MIMO system, it is only suitable for the data transmission stage. In fact, MUSIC and ESPRIT algorithms are originally designed for Radar application where there is no training. Nevertheless, in wireless communications, there do exist training sequences, so we utilize these training sequences to estimate the DOA information.

Comment

\(^2\)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.
The estimation method of channel gain information is not clear. The reviewer notes that a channel gain estimate (36) is given for LOS channels. What about NLOS channels with more than one channel path?

Response

We apologize for not making this part clear. Our method is also applicable to NLOS scenarios. We have added a small subsection to briefly explain the precoding method which is shown in Page 15 in the revised manuscript.

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

$$
Y = F_B^H F_R^H \sum_{k=1}^{K} \text{vec}(H_k)x_k^T + N
$$

$$
= \frac{1}{\sqrt{L}} F_B^H F_R^H \sum_{k=1}^{K} \sum_{l=1}^{L} a_{l,k} \text{vec}(A(\hat{\phi}_{l,k}, \hat{\theta}_{l,k}))x_k^T + N
$$

$$
= \frac{1}{\sqrt{L}} F_B^H F_R^H \sum_{k=1}^{K} A_k a_k x_k^T + N,
$$

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

$$
\hat{\mathbf{a}}_k = \sqrt{L}(F_B^H F_R^H A_k)^\dagger Y x_k = a_k + \sqrt{L}(F_B^H F_R^H A_k)^\dagger N x_k.
$$

Thus, we can obtain the uplink channel estimation for all users as

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

Comment

There are many unclear presentations. For example:

1) The definition of variables in (10) is not clear.

2) How can you obtain (16) from (14)? Is $x(t)$ set as 1 in (14)?

3) Fig. 1 does not show the subarray architecture.

4) What is the definition of searching guide $G$ in Fig. 3?
5) What is the beamspace based method in Fig. 6,7,8
6) Did you only consider one data stream in channel model (1) and (2) What about the inter-user (or inter-stream) interference?
7) \( \phi_l \) in (4) should be \( \phi_{l,k} \).

Response

We apologize for not making these part very clear. The corresponding part has been revised in the new manuscript for better clarity.

1) 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 based on the length of the RF chains and the number of antennas at the BS. Thus we perform the channel vector \( \text{vec}\{\tilde{H}_k\} \) via stacking the \( M_{RF} \times 1 \) subvectors as

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

where

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

2) The baseband signal before digital precoding at the \( p \)th position and \( t \)th time of the BS is written as

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

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

We can estimate \( \text{vec}\{H_k\}^{(p)} \) from (9) via

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

3) What we are considering is a fully-connected architecture. We have removed the confusing part in the introduction and revised Fig. 1.

4) 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 \([-\frac{\pi}{N}, \frac{\pi}{N}]\).

5) The beamspace based method in the simulation is shown in literature [5].

6) We have changed one data stream to multiple data stream in system model.

7) Sorry for our careless. We have fixed it.
REVIEWER 3

DETAILED COMMENTS

Comment

(1) In the paper, You achieved equation 4 by solving equation 5, but there is missing minus sign in equation 4 which may affect your further proof.

Response

Thanks for your carefulness. We should not use the conjugate-transpose, and we have fixed this problem in the revised manuscript.

Comment

(2) In section 2, you considered channel gain as zero mean unit variance, but in the mmWave communication system, channel paths are composed of LoS and NLoS. Furthermore, NLoSs are much weaker than LoS. However, why did you consider the same channel statistics for both?

Response

We apologize for not considering it. What you mentioned is correct. At mmWave frequencies, the amplitudes 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. Nevertheless, since the channel gain statistics has nothing to do with the following analysis, so we have removed the channel gain statistics in the revised manuscript.

Comment

(3) you are estimating DOA and path gain using the channel which is estimated through LS method but LS method is not the most accurate method to estimate the channel then how can you claim that you are calculating the true physical angle of paths and also if you are not able to estimate the true physical angle then you can not estimate channel gain correctly of each path. Please verify it.

Response

Thank you for your comment.
Unlike some other papers, we think channel statistical characteristics under the massive MIMO system are difficult to obtain, so we assume that the channel statistical characteristics are unknown. The LS method is very accurate when the statistical characteristics are unknown. In addition, the truth value we mention is relative to the poorly-accurate on-grid algorithm such as traditional DFT and beamspace methods. Our goal is to calculate the true physical angle information instead of the on-grid angle information, but there will certainly have errors.

Comment

(4) The main strength of this paper is 2D-DFT and angle rotation method. Furthermore, you mentioned that you need to apply angle rotation method to improve the accuracy of DOA but for that, you need a huge number of iteration as a searching guide $G$ that may have some complexity. What will be complexity for this?

Response

Thank you for your comment. We have add the complexity of the proposed algorithm after Algorithm 1 in the revised manuscript. 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 (3) is proportional to $O(MN)$ for the given $\Delta \phi_l$ and $\Delta \theta_l$, the complexity of the whole algorithm is about $O(MN \log MN + MN + GKM N)$. The complexity of different algorithms is shown in the Table I. Note that $K \ll MN$ and $G \ll MN^3$, the complexity of the proposed algorithm is much less than $O(M^2 N^2)$.

To obtain the real DOA, traditional DOA estimation algorithms, such as MUSIC and ESPRIT, also need to a huge searching, so they all are more complicated. Because this is the price of getting true value of DOA.

Comment

$^3$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.
(5) You did not mention the value of searching guide (G) for rest of the plots except figure 3.

Response

We apologize for not making this part very clear. We take angle rotation search grids \( G = G_M G_N = 30 \times 30 = 900 \) unless otherwise mentioned.

Comment

(6) In figure 6 & 7, you mentioned y-axis as SNR/dB. I did not understand this.

Response

Thank you for your comment. It should be SNR(dB) and I have fixed this problem.
REVIEWER 4

DETAILED COMMENTS

Comment

This paper proposes a DOA-aided channel estimation for hybrid millimeter wave (mmWave) massive MIMO systems with the uniform planar array (UPA) at the base station (BS). The initial DOAs can be estimated through 2D-FFT with a resolution inversely proportional to MN, and its resolution can be further enhanced via angle rotation technique. The authors further derive a simple expression for the corresponding theoretical bounds of MSE performance in high SNR region, and the Cramer-Rao lower bounds (CRLBs) to reflect the theoretical lower MSE bound of DOA and channel gain estimation. The following concerns need to be addressed.

Response

We would like to first thank this reviewer for his/her positive comment and recognition of our contributions.

Comment

1. In Eq. (10), why the parameter $D$ is chosen as $(MN/M_{RF})M_{RF}$? What is the physical interpretation of $D$?

Response

We apologize for not making this part clear. We use another way to explain this process. 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 based on the length of the RF chains and the number of antennas at the BS. Thus we perform the channel vector $\text{vec}\{\hat{H}_k\}$ via stacking the $M_{RF} \times 1$ subvectors as

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

where

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

Comment
2. Via angle rotation, Algorithm 1 can enhance the DOA estimation. However, the complexity of one-dimensional search of $\Delta \phi_l$ and $\Delta \theta_l$ is high. How do the authors select the step length?

Response

Thank you for your comment. We have add the complexity of the proposed algorithm after Algorithm 1 in the revised manuscript. Note that the number of search grids $G = G_M G_N$, where $G_M$ means the search grids within $\left[ -\frac{\pi}{M}, \frac{\pi}{M} \right]$ and $G_N$ means the search grids within $\left[ -\frac{\pi}{N}, \frac{\pi}{N} \right]$, 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 (3) is proportional to $O(MN)$ for the given $\Delta \phi_l$ and $\Delta \theta_l$, the complexity of the whole algorithm is about $O(MN \log MN + MN + GKM N)$. The complexity of different algorithms is shown in the Table I. Note that $K \ll MN$ and $G \ll MN^4$, the complexity of the proposed algorithm is much less than $O(M^2 N^2)$.

Comment

3. In Fig. 4, as the number of antennas increases, the MSE of the initial estimation keeps constant for different SNR. As the number of antennas increases, will the performance of initial estimation approach angle rotation scheme? It is interesting to see when the initial estimation can approach the angle rotation scheme.

Response

Thank you for your suggestion. We have modify the Fig. 4 and add what you mentioned.

Fig. 2 plots the MSE performances of DOA estimation as a function of SNR for various URA sizes. We assume that the total transmit power for each BS antenna are constrained constantly. It is clearly seen from Fig. 2 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. 2 that the proposed DOA estimation method outperforms the initial estimation dramatically in the high SNR region. Moreover, the initial

$^4$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.
Fig. 2. The MSE performance 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.

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.

**Comment**

4. The paper requires more elaborative deep work on grammar and typos, e.g., hundreds or thousands antennas $\rightarrow$ hundreds or thousands of antennas, the Cramer-Rao lower bound $\rightarrow$ the Cramer-Rao lower bound.

**Response**

We apologize for our carelessness. In the revised manuscript, we have modified these linguistic errors and we have improved the writing of English.
REVIEWER 5

DETAILED COMMENTS

Comment

please review carefully the text and modify the typos entire the paper. E.g. first paragraph “Meanwhile, For a ...”

Response

We apologize for our carelessness. In the revised manuscript, we have modified these linguistic errors.

Comment

About the complexity, how much less complex than other? I suggest a table with this comparative information.

Response

Thank you for your suggestion. We have add the complexity of the proposed algorithm after Algorithm 1 in the revised manuscript. Note that the number of search grids $G = G_MG_N$, 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 (3) is proportional to $O(MN)$ for the given $\Delta \phi_l$ and $\Delta \theta_l$, the complexity of the whole algorithm is about $O(MN\log MN + MN + GMKMN)$. The complexity of different algorithms is shown in the Table II. Note that $K \ll MN$ and $G \ll MN^5$, the complexity of the proposed algorithm is much less than $O(M^2N^2)$.

Comment

In the simulation results, please define which CS approach was used and explain better the Figures 7 and 8.

$^5$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.
TABLE II  
COMPLEXITY OF DIFFERENT ESTIMATE ALGORITHMS

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed 2D-DFT and angle rotation estimate algorithm</td>
<td>$O(MN \log MN + MN + GKMN)$</td>
</tr>
<tr>
<td>Beamspace based estimate algorithm</td>
<td>$O(MN \log MN + MN)$</td>
</tr>
<tr>
<td>CS based estimate algorithm</td>
<td>$O(M^2N^2 + KMN)$</td>
</tr>
</tbody>
</table>

**Response**

Thank you for your comment. We have already indicated the used CS channel estimate approach [4], the beamspace based estimate method [5], and the eigen-decomposition based estimate method [6] in the revised manuscript.

Fig. 7 plots the achievable sum rate for the downlink data transmission with different channel estimation method using the proposed hybrid precoding method. 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 based 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 based method suffer from severe channel power leakage, but most of the channel power concentrated 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 based 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 achievable 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 based method become smaller. Our results clearly demonstrate the effectiveness of the proposed method.

Finally, we would like to thank you again for serving the second time as our reviewer.
and thank you for your professional comments.

REFERENCES


Angle Domain Channel Estimation in Hybrid MmWave Massive MIMO Systems

Dian Fan, Feifei Gao, Yuanwei Liu, Yansha Deng, Gongpu Wang, Zhangdui Zhong, Arumugam Nallanathan

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) performance and the Cramér-Rao Lower bounds (CRLBs) of the joint DOA and channel gain estimation. Simulation results show that the proposed DOA estimation and channel gain estimation 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.

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 zhfdzhong@bjtu.edu.cn). Feifei Gao is with Department of Automation, Tsinghua University, State Key Lab of Intelligent Technologies and Systems, Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing 100084, P. R. China (e-mail: 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).
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 in the range of 30 to 300 GHz has been proposed for outdoor cellular systems [1]–[4]. The mmWave communication requires massive antennas to overcome high propagation path loss via beamsteering 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 to the number of data streams. 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 analogdigital 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]–[31].

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 in practice. To solve this problem, [17]–[25] proposed the compressive sensing (CS) based channel estimation schemes via efficiently utilize the sparsity of mmWave channel in the angle domain and incorporating the hybrid architecture. However, the complexity of which is still high due to the non-linear optimization, and its effectiveness highly depends on the restricted isometry property (RIP).

The authors in [26], [27] proposed an open-loop channel estimation strategy, which is independent of the hardware constraints. In [28], a grid-of-beams (GoB) based approach was proposed to obtain the angle-of-departure (AOD) and the direction-of-arrival (DOA), which requires large amount of training with high overhead due to the best combinations of analog transmit and receive beams were achieved via exhaustive sequential search. In [29], 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 [30], [31], 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, therefore, there is an urgent requirement for a fast and accurate estimation approach that could efficiently estimate the angle information specifically for mmWave massive MIMO communications systems 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 raised enormous interests inside the array processing community for decades due to their high resolution angle estimation [32]–[34]. Their applications in massive MIMO systems and full-dimension MIMO systems for two-dimensional angles estimation have been extensively studied in [35]–[39]. However, the conventional MUSIC and ESPRIT are not suitable for the mmWave frequency band due to the following main reasons: 1) They are of very high computational complexity during their 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 has single antenna. The parameters of each paths in the channel matrix are decomposed into the corresponding channel gain and the DOA information for better estimation. 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 easy for practical implementation. Moreover, the obtained DOA estimation results are used for the following channel gain estimation. Most importantly, we further derive a simple expression for the corresponding theoretical bounds of mean squared errors (MSEs) performance in high signal-to-noise ratio (SNR) region, as well as the Cramér-Rao lower bounds (CRLBs) to reflect the theoretical lower MSE bounds of DOA and channel gain estimation. 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 presented 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; $tr(A)$ donates the trace of $A$; $[A]_{ij}$ is the $(i, j)$th entry of $A$; Diag{$a$} denotes a diagonal matrix with the diagonal element constructed from $a$, while Diag{$A$} denotes a vector whose elements are extracted from the diagonal components of $A$; vec$(A)$ denotes the vectorization of $A$; $\Re\{A\}$ denotes the real part of $A$; $\Im\{A\}$ 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 $[a]_j$; $E\{\cdot\}$ denotes the statistical expectation, and $\|h\|$ is the Euclidean norm of $h$. 
II. SYSTEM MODEL

We 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 $\| [\mathbf{F}_{RF}]_{:,j} [\mathbf{F}_{RF}]_{:,j}^\ast \|_{F} = \frac{1}{MN}$, $(j = 1, 2, \cdots M_{RF}, i = 1, 2, \cdots MN)$, 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 $k$th mobile user, with block-fading channel. In the uplink transmission stage, the received signal at the BS can be
expressed as
\[
y(t) = F^H_{BB} F^H_{RF} \sum_{k=1}^{K} \text{vec}\{H_k\} s_k(t) + N, t = 1, 2, \ldots, T, \tag{1}
\]
where \(s_k(t)\) is the transmitted signal at time \(t\), \(N \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})\) is the complex Gaussian noise matrix, and \(\sigma_n^2\) is the unit 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
\[
y_k = \text{vec}^H \{H_k\} F_{RF} F_{BB} s + n_k, \tag{2}
\]
where \(s = [s_1, s_2, \ldots, s_K]^T\) is the transmitted signal vector for all \(K\) mobile users. Thus, we can express the received SNR at the \(k\)th mobile user as
\[
\Gamma_k = |\text{vec}^H \{H_k\} F_{RF} F_{BB}|^2 \frac{\sigma_s^2}{\sigma_n^2}, \tag{3}
\]
where \(E\{|s_k|^2\} = \sigma_s^2\) denotes the power of \(s_k\).

**B. Channel Model**

Due to the limited scattering characteristics in the mmWave environment [40]–[43], we assume the channel representation based on the extended Saleh-Valenzuela (SV) model in [26]–[28]. 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 \((l = 1, 2, \ldots, L)\) path of the \(k\)th user. The corresponding steering matrix can be expressed as
\[
A(\phi_{l,k}, \theta_{l,k}) = \frac{1}{\sqrt{MN}} \begin{bmatrix}
1 & \ldots & e^{j2\pi d/\lambda(N-1) \sin \phi_{l,k} \cos \theta_{l,k}} \\
\vdots & \ddots & \vdots \\
e^{j2\pi d/\lambda(M-1) \cos \phi_{l,k}} & \ldots & e^{j2\pi d/\lambda(M-1) \cos \phi_{l,k} + (N-1) \sin \phi_{l,k} \cos \theta_{l,k}}
\end{bmatrix}_{M \times N}, \tag{4}
\]
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
\[
A(\phi_{l,k}, \theta_{l,k}) = A(w_{1,l,k}, w_{2,l,k}) = a(w_{1,l,k}) a^T(w_{2,l,k}), \tag{5}
\]
where \(a(w_{1,l,k}) = \frac{1}{\sqrt{M}} [1, \ldots, e^{j(M-1)w_{1,l,k}}]^T\) and \(a(w_{2,l,k}) = \frac{1}{\sqrt{N}} [1, \ldots, e^{j(N-1)w_{2,l,k}}]^T\).
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

$$
H_k = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} a_{l,k} A(\phi_{l,k}, \theta_{l,k}) = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} a_{l,k} a(w_{1,l,k}) a^T(w_{2,l,k}) = \frac{1}{\sqrt{L}} A_{w_1,k} H_{a,k} A_{w_2,k}^T,
$$

(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 ($l = 0$ 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 $H_k$ 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 $\{|a_{l,k}|\}_{l=2}^{L}$ of the NLOS component [41].

The channel matrix $H_k$ in mmWave massive MIMO system can be very large with up to hundreds of columns and rows depending on the number of BS antennas.

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 $H_k$ has only $L$ complex channel gains and $2L$ real phases $(\phi_{l,k}, \theta_{l,k})$, where the number of paths is usually much smaller 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 [44]. It is noted that the angle information calculated by the beamspace based method and CS based method which use the channel model (6) is not the real physical angle, but only an approximation of the quantized angle space with limited resolution. However, our method of this paper is to calculate the true physical angle information.
III. ARRAY PROCESSING BASED DOA ESTIMATION

In this section, we propose a new DOA estimation algorithm for the hybrid antenna array. To facilitate the understanding of the proposed DOA estimation algorithm, 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 of the BS can be written as

$$Y_{BB} = F_{RF}^H \sum_{k=1}^{K} \text{vec}\{H_k\} x_k^T + N,$$

where $x_k = [x_{k,1}, x_{k,2}, \cdots, x_{k,\tau}]^T$ is the training sequence of the $k$th user, $\tau \geq K$ is the length of training sequences, 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 based on the length of the RF chains and the number of antennas at the BS. Thus we perform the channel vector $\text{vec}\{\tilde{H}_k\}$ via stacking the $M_{RF} \times 1$ subvectors as

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

where

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

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

$$F_{RF}(p) = \begin{bmatrix} 0_{(pM_{RF}) \times M_{RF}} & 0_{(MN-(p+1)M_{RF}) \times M_{RF}} \\ \text{U}_{RF} & 0_{(MN-(p+1)M_{RF}) \times M_{RF}} \end{bmatrix}_{MN \times M_{RF}}.$$
where \( U_{RF} \) is an \( M_{RF} \times M_{RF} \) Hadamard matrix, and its elements are of unitary magnitude, such that \( U_{RF} \) has full rank. For the \( k \)th user, based on (9), the baseband signal before digital precoding at the \( p \)th position and \( t \)th time of the BS is written as

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

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

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

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

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

\[
\tilde{F}_{RF}^H = \begin{bmatrix}
F_{RF}^{1H} & 0 & \cdots & 0 \\
0 & F_{RF}^{2H} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & F_{RF}^{DH}
\end{bmatrix}_{MN \times MN},
\]

where \( F_{RF}^q \) 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

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

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

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

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

\[
\text{vec}\{H_k\} = (\tilde{F}_{RF}^H)^{-1} \tilde{Y}_{BB,k} \frac{\tilde{x}_k^T}{\|\tilde{x}_k^T\|^2} + (\tilde{F}_{RF}^H)^{-1} \tilde{N}_k \frac{\tilde{x}_k^T}{\|\tilde{x}_k^T\|^2}.
\]
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 $X = [x_1, x_2, \ldots, x_K]^T$ 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

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

where

$$H = [\text{vec}\{H_1\}, \text{vec}\{H_2\}, \ldots, \text{vec}\{H_K\}],$$
$$\tilde{X} = [X, X, \ldots, X], \quad \tilde{N} = \sum_{k=1}^{K} [\tilde{N}_k, \tilde{N}_k, \ldots, \tilde{N}_k]. \quad (20)$$

Thus, the channel for all users can be estimated from

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

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 [30], while the remaining channel gain component could be re-estimated via much simplified approach.

After obtaining the initial channel estimation for all users in the preamble stage, 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, where 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 $[F_M]_{pp'} = \frac{1}{\sqrt{M}}e^{-j\frac{2\pi}{M}pp'}$, $p, p' = 0, 1, \ldots, M - 1$ and $[F_N]_{qq'} = \frac{1}{\sqrt{N}}e^{-j\frac{2\pi}{N}qq'}$, $q, q' = 0, 1, \ldots, N - 1$, respectively. Meanwhile, we define the normalized 2D-DFT of the channel matrix $H$ as $H_{DFT} = F_MHF_N$, whose $(p,q)$th element $(p = 0, 1, \ldots, M - 1; q = 0, 1, \ldots, N - 1)$ is computed as

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

$$= \frac{1}{\sqrt{LMN}} \sum_{l=1}^{L} a_l e^{-j\frac{M-1}{2}(\frac{2\pi}{M}p-w_{1,l})} e^{-j\frac{N-1}{2}(\frac{2\pi}{N}q-w_{2,l})}$$

$$\times \frac{\sin \left(\pi p - \frac{Mw_{1,l}}{2}\right)}{\sin \left((\pi p - \frac{Mw_{1,l}}{2})/M\right)} \cdot \frac{\sin \left(\pi q - \frac{Nw_{2,l}}{2}\right)}{\sin \left((\pi q - \frac{Nw_{2,l}}{2})/N\right)}. \quad (22)$$

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 $p_l = \frac{Mw_{1,l}}{2\pi}$ and $q_l = \frac{Nw_{2,l}}{2\pi}$ such that $[H_{DFT}]_{pq_l} = \frac{a_l}{\sqrt{LMN}}$, while the other elements of $H_{DFT}$ are all zero, In other words, all power is concentrated on the $(p_l, q_l)$th 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

$$\phi_l = \cos^{-1} \left(\frac{\lambda p_l}{Md}\right), \quad \theta_l = \cos^{-1} \left(\frac{\lambda q_l}{Nd} \sqrt{1 - \left(\frac{\lambda p_l}{Md}\right)^2}\right). \quad (23)$$

Unfortunately, in the practice, the array aperture cannot be infinitely large, even if $MN$ could be as greater as hundreds or thousands in hybrid mmWave massive MIMO communication 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, it can also makes all power of channel concentrate on some separated single point. We call these on-grid angle. Note that the on-grid angle are very limited and it is positively proportional to $M$ and $N$. In more general case, $Mw_{1,l}/(2\pi)$ and $Nw_{2,l}/(2\pi)$ will not be integers, and the channel power of $H_{DFT}$ will leak from the $([Mw_{1,l}/(2\pi)], [Nw_{2,l}/(2\pi)])$th element to its nearby elements. In fact, the leakage of channel power is positively proportional to the deviation $(Mw_{1,l}/(2\pi) - [Mw_{1,l}/(2\pi)])$ and $(Nw_{2,l}/(2\pi) - [Nw_{2,l}/(2\pi)])$, but is inversely proportional to $M$ and $N$ as shown in (22). However, $H_{DFT}$ can still be approximated as a sparse
Fig. 2. An example of a two paths channel sparse characteristics after 2D-DFT and optimal angle rotation, where BS array has $100 \times 100$ antennas.

matrix with most of power concentrated around the $(\lfloor M w_1/(2\pi) \rfloor, \lfloor N w_2/(2\pi) \rfloor)$th element. Hence, the peak power position of $H_{DFT}$ is still useful for extracting initial DOA information.

An example of a two paths 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 noise-free scenario. It can be seen that both two paths correspond 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 in each bin 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 to indicate the initial DOA estimation.

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

$$
[\hat{H}_{DFT}]_{pq} = [H_{DFT}]_{pq} + [N_{DFT}]_{pq},
$$

(24)

where $N_{DFT} \sim \mathcal{CN}(0, \frac{\sigma_n^2}{\sqrt{MN}}\mathbf{I})$. Denote the $L$ largest peaks in $L$ bins of $\hat{H}_{DFT}$ as $(p_{l_{ini}}, q_{l_{ini}})$. 

July 8, 2018 DRAFT
We can express the initial DOA estimates as
\[ \hat{\phi}_{i}^{\text{ini}} = \cos^{-1} \left( \frac{\lambda p_{i}^{\text{ini}}}{Md} \right), \quad \hat{\theta}_{i}^{\text{ini}} = \cos^{-1} \left( \frac{\lambda q_{i}^{\text{ini}}}{Nd} \sqrt{1 - \left( \frac{\lambda p_{i}^{\text{ini}}}{Md} \right)^2} \right). \]  
(25)

2) Fine DOA Estimation: The resolution of \((\hat{\phi}_{i}^{\text{ini}}, \hat{\theta}_{i}^{\text{ini}})\) 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 \((\hat{\phi}_{i}^{\text{ini}}, \hat{\theta}_{i}^{\text{ini}})\) 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
\[ H^{\text{ro}} = \Phi_{M} (\Delta \phi_{l}) H \Phi_{N} (\Delta \theta_{l}), \]  
(26)
where the diagonal matrices \(\Phi_{M} (\Delta \phi_{l})\) and \(\Phi_{N} (\Delta \theta_{l})\) are given by
\[ \Phi_{M} (\Delta \phi_{l}) = \text{diag} \{1, e^{j \Delta \phi_{1}}, \ldots, e^{j (M-1) \Delta \phi_{l}}\}, \quad \Phi_{N} (\Delta \theta_{l}) = \text{diag} \{1, e^{j \Delta \theta_{1}}, \ldots, e^{j (N-1) \Delta \theta_{l}}\}. \]  
(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 \(H^{\text{ro}}_{\text{DFT}}\) can be calculated as
\[ [H^{\text{ro}}_{\text{DFT}}]_{pq} = \frac{1}{\sqrt{LMN}} \sum_{l=1}^{L} a_{l} e^{-j \frac{M-1}{2} (\frac{\pi p}{M} - p_{w,1,l} - \Delta \phi_{l})} e^{-j \frac{N-1}{2} (\frac{\pi q}{N} - q_{w,2,l} - \Delta \theta_{l})} \times \frac{\sin \left( \pi p - \frac{M w_{1,l}}{2} - \frac{M \Delta \phi_{l}}{2} \right)}{\sin \left( (\pi p - \frac{M w_{1,l}}{2} - \frac{M \Delta \phi_{l}}{2})/M \right)} \times \frac{\sin \left( \pi q - \frac{N w_{2,l}}{2} - \frac{N \Delta \theta_{l}}{2} \right)}{\sin \left( (\pi q - \frac{N w_{2,l}}{2} - \frac{N \Delta \theta_{l}}{2})/N \right)}. \]  
(28)

It can be readily found that the entries of \(H^{\text{ro}}_{\text{DFT}}\) has only \(L\) non-zero elements when the angle shifter satisfying
\[ \Delta \phi_{l} = 2 \pi p_{l} / M - w_{1,l}, \quad \Delta \theta_{l} = 2 \pi q_{l} / N - w_{2,l}, \]  
(29)
where \((\Delta \phi_{l}, \Delta \theta_{l})\) in (29) are the optimal angle shifter.

Based on the derived optimal angle shifter, the elevation angle and the azimuth angle of the \(l\)th path \((\phi_{l}, \theta_{l})\) can be estimated as
\[ \hat{\phi}_{l} = \cos^{-1} \left( \frac{\lambda p_{l}}{Md} - \frac{\lambda \Delta \phi_{l}}{2 \pi d} \right), \]  
\[ \hat{\theta}_{l} = \cos^{-1} \left( \frac{\lambda q_{l}}{Nd} - \frac{\lambda \Delta \theta_{l}}{2 \pi d} \right) / \sqrt{1 - \left( \frac{\lambda p_{l}}{Md} - \frac{\lambda \Delta \phi_{l}}{2 \pi d} \right)^2}. \]  
(30)
For finding the optimal angle shifter \((\Delta \phi_l, \Delta \theta_l)\) from channel matrix \(H\), one way is a 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 \((p_{ini}^{l}, q_{ini}^{l})\)th element of \(F_M \Phi_M(\Delta \phi_l)H\Phi_N(\Delta \theta_l)F_N\) shrink into their highest form, one by one. Mathematically, there is

\[
(\Delta \phi_l, \Delta \theta_l) = \arg \max_{\Delta \phi \in [-\frac{\pi}{M}, \frac{\pi}{M}], \Delta \theta \in [-\frac{\pi}{N}, \frac{\pi}{N}]} \| f^H_{M,p_{ini}^{l}} \Phi_M(\Delta \phi_l)H\Phi_N(\Delta \theta_l)f_{N,q_{ini}^{l}} \|^2 ;
\]

where \(f_{M,p_{ini}^{l}}\) is the \(p_{ini}^{l}\)th column of \(F_M\) and \(f_{N,q_{ini}^{l}}\) is the \(q_{ini}^{l}\)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 the 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 \(O(MN \log MN + MN + G K M N)\). The complexity of different algorithms is shown in the Table I. Note that \(K \ll MN\) and \(G \ll M N\), the complexity of the proposed algorithm is much less than \(O(M^2 N^2)\).

**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.

\(^1\)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.
For \( \hat{\phi}_l, \hat{\theta}_l \) from the equation (30).

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

Input: \( H \). 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 \( H_{DFT} = F_H H F_N \), where \((p_l^{ini}, q_l^{ini}) = \arg \max_{(p,q) \in \text{bin}(l)} \| [H_{DFT}]_{pq} \|^2, l = 1, 2, \cdots, L \).
2. For \( \Delta \phi = -\frac{\pi}{M} : \frac{\pi}{M}, \Delta \theta = -\frac{\pi}{N} : \frac{\pi}{N} \) \( l = 1 : L \)
3. \( (\Delta \hat{\phi}_l, \Delta \hat{\theta}_l) = \arg \max_{\Delta \phi \in [-\pi,M\pi], \Delta \theta \in [-\pi,N\pi]} \| f_{L,l,k}^{H} \Phi M^{l}(\Delta \phi_l)H\Phi N(\Delta \theta_l)f_{N,q,k}^{H} \|^2 \), \( \text{End For} \)
5. \( \text{End For} \)

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

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

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

where \( A_k = [\text{vec}(A(\hat{\phi}_{1,k}, \hat{\theta}_{1,k})), \text{vec}(A(\hat{\phi}_{2,k}, \hat{\theta}_{2,k})), \cdots, \text{vec}(A(\hat{\phi}_{L,k}, \hat{\theta}_{L,k}))] \), and \( a_k = [a_{1,k}, a_{2,k}, \cdots, a_{L,k}]^T \). 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

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

\[
Y = F_{BB}^{H} F_{RF}^{H} \sum_{k=1}^{K} \text{vec}(H_k) x_k^T + N
\]
LS estimation as
\[
\hat{a}_k = \sqrt{L}(F_{BB}^H F_{RF}^H A_k)^\dagger Y x_k = a_k + \sqrt{L}(F_{BB}^H F_{RF}^H A_k)^\dagger N x_k. \tag{33}
\]

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

From (2), the downlink received signals can be expressed as
\[
y^d = H^d F_{RF} F_{BB} s + n, \tag{35}\]
where \(H^d = [\text{vec}\{H_1\}, \text{vec}\{H_2\}, \ldots, \text{vec}\{H_K\}]^H\), and \(n \sim \mathcal{CN}(0, \sigma_n^2 I_K)\) is additive white Gaussian noise vector. We assume \(M_{RF} = KL\). From the previous discussion, the analog precoding matrix can be immediately obtained from
\[
F_{RF} = [\text{vec}\{\Phi_M(\Delta \phi_{1,1})f_{M,p_{1,1}} f_{N,q_{1,1}}^H \Phi_N(\Delta \theta_{1,1})\}, \ldots, \text{vec}\{\Phi_M(\Delta \phi_{L,K})f_{M,p_{L,K}} f_{N,q_{L,K}}^H \Phi_N(\Delta \theta_{L,K})\}], \tag{36}\]
where \((p_{l,k}, q_{l,k})\) denotes the position that contains the largest power after 2D-DFT of the \(l\)th path of the \(k\)th user, and each column of \(F_{RF}\) 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, which is a key difference from beamspace transmission.

Similar to the conventional digital precoding approach, \(F_{BB}\) can be obtained via the zero-forcing (ZF) beamforming algorithm, which can be written as
\[
F_{BB} = \frac{1}{\sqrt{P}}(H^d F_{RF})^H ((H^d F_{RF})(H^d F_{RF})^H)^{-1}, \tag{37}\]
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 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 performance bounds of the joint DOA and channel gain estimation for hybrid mmWave massive MIMO system. Generally, a closed-form MSE performance 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 [45]. We first show that the MSE performance 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 Performance Bounds of The Proposed Estimator

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

$$y = F^H_{RF} \text{vec}(H)s + n = F^H_{RF} \text{vec}(A(w_1, w_2))\alpha s + n,$$

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

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

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

$$[\hat{w}_1, \hat{w}_2] = \arg\max_{w_1, w_2} \| \text{vec}^H(A(w_1, w_2))(F^H_{RF})^\dagger y \|^2$$

$$= \arg\max_{w_1, w_2} y^H F^H_{RF} \text{vec}(A(w_1, w_2)) \text{vec}^H(A(w_1, w_2))(F^H_{RF})^\dagger y,$$

where $\text{vec}(A(w_1, w_2)) = \text{vec}\{ \Phi_M(\Delta \phi)f_M f_N^H \Phi_N(\Delta \theta) \}$. 

July 8, 2018

DRAFT
For given $w_1$, $w_2$ and $\alpha$, the probability density function (PDF) of $y$ can be expressed as

$$f(y|w_1, w_2, \alpha) = \frac{1}{(\pi\sigma_n^2)^{MRF}} \exp \left\{-\frac{\|y - \mathbf{F}_{RF}^H \mathbf{v}^e\|^2}{\sigma_n^2}\right\}.$$ \hspace{1cm} (41)

The joint ML estimates of $w_1$, $w_2$ and $\alpha$ can be obtained via

$$[\hat{w}_1, \hat{w}_2, \hat{\alpha}] = \arg\max_{w_1, w_2, \alpha} f(y|w_1, w_2, \alpha),$$ \hspace{1cm} (42)

or equivalently

$$[\hat{w}_1, \hat{w}_2, \hat{\alpha}] = \arg\min_{w_1, w_2, \alpha} \|y - \mathbf{F}_{RF}^H \mathbf{v}^e\|.$$ \hspace{1cm} (43)

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

For given $w_1$ and $w_2$, the ML estimates of $\alpha$ is obtained from (43) as

$$\hat{\alpha} = \mathbf{v}^H (\mathbf{A}(w_1, w_2)) \mathbf{F}_{RF}^H \mathbf{A}^\dagger \mathbf{s} \mathbf{y}.$$ \hspace{1cm} (44)

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

$$[\hat{w}_1, \hat{w}_2] = \arg\min_{w_1, w_2} \|y - \mathbf{F}_{RF}^H \mathbf{v}^e\|,$$

$$\left\{\mathbf{A}(w_1, w_2) \mathbf{v}^e \mathbf{A}(w_1, w_2) \mathbf{F}_{RF}^H \mathbf{A}^\dagger \mathbf{s} \mathbf{y}\right\},$$

$$= \arg\min_{w_1, w_2} \|y - \mathbf{F}_{RF}^H \mathbf{v}^e\|,$$

$$\left\{\mathbf{A}(w_1, w_2) \mathbf{v}^e \mathbf{A}(w_1, w_2) \mathbf{F}_{RF}^H \mathbf{A}^\dagger \mathbf{s} \mathbf{y}\right\},$$

$$= \arg\max_{w_1, w_2} y^H \mathbf{F}_{RF}^H \mathbf{P}_a \mathbf{F}_{RF}^H \mathbf{A}^\dagger \mathbf{s} \mathbf{y} = \arg\max_{w_1, w_2} g(w_1, w_2),$$ \hspace{1cm} (45)

where $g(w_1, w_2)$ denotes the cost function of $w_1, w_2$, $s^s = \sigma_s^2$, and $\mathbf{P}_a = \mathbf{v}^e \mathbf{A}(w_1, w_2)$ represents the projection matrix onto the subspace spanned by $\mathbf{v}^e \mathbf{A}(w_1, w_2)$.

Interestingly, the MSE performance (40) of the proposed estimator coincides with the ML estimator (45). Till now, we have (44) and (45) as the ML estimates of $\alpha, w_1, w_2$.

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

$$\mathbb{E}\{\Delta w_i^2\} = \frac{\sigma_n^2}{\sigma_s^2 |\alpha|^2} \mathbf{v}^e \mathbf{A}(w_1, w_2) \mathbf{P}_a^\dagger \mathbf{W}_i \mathbf{P}_a \mathbf{W}_i \mathbf{v}^e \mathbf{A}(w_1, w_2),$$ \hspace{1cm} (46)

where $i = 1, 2$, $\mathbf{P}_a^\dagger = \mathbf{I} - \mathbf{P}_a$ is the projection matrix onto the orthogonal space spanned by $\mathbf{A}(w_1, w_2)$, and $\mathbf{W}_1$, $\mathbf{W}_2$ are the diagonal matrices as

$$\mathbf{W}_1 = \text{Diag}\{0, \ldots, 0, 1, \ldots, 1, \ldots, (M - 1), \ldots, (M - 1)\},$$ \hspace{1cm} (47)

$$\mathbf{W}_2 = \text{Diag}\{0, \ldots, 0, 1, \ldots, 1, \ldots, (N - 1), \ldots, (N - 1)\}.$$ \hspace{1cm} (48)
Proof: See Appendix.

In the channel estimation process of hybrid mmWave massive MIMO system, we need to further examine the MSE performance of the azimuth angle \( \phi \) and the elevation angle \( \theta \). 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

\[
E\{\Delta \phi\} = E\{\Delta \theta\} = 0,
\]

\[
E\{\Delta \phi^2\} = E\{(\hat{\phi} - \phi)(\hat{\phi} - \phi)^H\} = \frac{\partial \phi}{\partial w_1} E\{w_1^2\} \left( \frac{\partial \phi}{\partial w_1} \right)^H
= \frac{(\frac{\lambda}{2\pi d})^2}{1 - (\frac{\lambda w_1}{2\pi d})^2} \times 2\sigma_n^2 [\text{vec}(A^H(w_1, w_2))W_1P_a^\perp W_1 \text{vec}(A(w_1, w_2))],
\]

\[
E\{\Delta \theta^2\} = E\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^H\} = \frac{\partial \theta}{\partial w_2} E\{w_2^2\} \left( \frac{\partial \theta}{\partial w_2} \right)^H
= \frac{(\frac{\lambda}{2\pi d \sin \phi})^2}{1 - (\frac{\lambda w_2}{2\pi d \sin \phi})^2} \times 2\sigma_n^2 [\text{vec}(A^H(w_1, w_2))W_2P_a^\perp W_2 \text{vec}(A(w_1, w_2))]. \tag{50}
\]

We further derive the expectation and the variance of the channel gain estimate \( \hat{\alpha} \). Based on (44), we write \( \hat{\alpha} \) as

\[
\hat{\alpha} = \text{vec}^H(\hat{A}(w_1, w_2))(F_{RF}^H)^\dagger s^* (F_{RF}^H \text{vec}(A(w_1, w_2))\alpha s^* + n)
= \sigma_n^2 \text{vec}^H(\hat{A}(w_1, w_2))\text{vec}(A(w_1, w_2))\alpha + \text{vec}^H(\hat{A}(w_1, w_2))(F_{RF}^H)^\dagger s^* n, \tag{51}
\]

where \( \text{vec}^H(\hat{A}(w_1, w_2)) \) is constructed from the estimate \( (\hat{w}_1, \hat{w}_2) \). With the help of Taylor’s expansion, \( \text{vec}^H(\hat{A}(w_1, w_2)) \) can be approximated by

\[
\text{vec}^H(\hat{A}(w_1, w_2)) \approx \text{vec}^H(A(w_1, w_2)) + j\text{vec}^H(A(w_1, w_2))W_i \Delta w_i, \quad i = 1, 2. \tag{52}
\]

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

\[
\hat{\alpha} = \sigma_n^2 (\text{vec}^H(A(w_1, w_2)) + j\text{vec}^H(A(w_1, w_2))W_1 \Delta w_1) \text{vec}(A(w_1, w_2))\alpha
+ \text{vec}^H(\hat{A}(w_1, w_2))(F_{RF}^H)^\dagger s^* n
= \alpha + j\text{vec}^H(A(w_1, w_2))W_1 \text{vec}(A(w_1, w_2))\Delta w_1 \alpha + \text{vec}^H(\hat{A}(w_1, w_2))(F_{RF}^H)^\dagger s^* n. \tag{53}
\]
With the help of (46), we can derive the mean and the MSE of the channel gain estimation as

\[
\begin{align*}
E\{\Delta \alpha\} &= E\{j\text{vec}^H(A(w_1, w_2))W_1\text{vec}(A(w_1, w_2))\Delta w_1\alpha + \text{vec}^H(\hat{A}(w_1, w_2))(F_{RF}^H)^\dagger s^*n\} = 0, \\
E\{\Delta \alpha^2\} &= E\{(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^H\} = \alpha E\{(\Delta w_1)^2\} \alpha^H|\text{vec}^H(A(w_1, w_2))W_1\text{vec}(A(w_1, w_2))|^2 \\
&+ \sigma_n^2|\text{vec}^H(\hat{A}(w_1, w_2))(F_{RF}^H)^\dagger E\{nn^H\}((F_{RF}^H)^\dagger)^H\text{vec}(\hat{A}(w_1, w_2))| \\
&= \frac{\sigma_n^2|\text{vec}^H(A(w_1, w_2))W_1\text{vec}(A(w_1, w_2))|^2}{2\sigma_s^2|\text{vec}^H(A(w_1, w_2))W_1^T F_{RF} W_1\text{vec}(A(w_1, w_2)) + \sigma_n^2}. \quad (54)
\end{align*}
\]

In (54), the first term is caused by the estimation error in \((\phi, \theta)\), and 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 the estimate of \(\alpha\) is then given by

\[
\text{MSE}(\alpha) = \frac{\sigma_n^2|\text{vec}^H(A(w_1, w_2))W_1\text{vec}(A(w_1, w_2))|^2}{2\sigma_s^2|\text{vec}^H(A(w_1, w_2))W_1^T F_{RF} W_1\text{vec}(A(w_1, w_2)) + \sigma_n^2} + \sigma_n^2. \quad (55)
\]

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. A detailed analysis of joint CRLBs for \((\phi, \theta)\) and channel gain \(\alpha\) estimation 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

\[
Y = HS + N = \alpha A(\phi, \theta)s + N = \alpha a(w_1)a^T(w_2)s + N. \quad (56)
\]

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

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

where the real part of the received signal is

\[
y_{m, n}^R = \Re\{y_{m, n}\} = \Re\{\alpha e^{j((m-1)w_1+(n-1)w_2)}s\} + \Re\{n_{m, n}\} = x_{m, n} + n'_{m, n}, \quad (58)
\]
and \(x_{m,n} = \Re\{\alpha e^{j(m-1)w_1 + (n-1)w_2}\}\) and \(n'_{m,n} = \Re\{n_{m,n}\}\). For given \(\alpha, \phi, \) and \(\theta\), the probability density function of \(Y\) can be expressed as

\[
f(Y|\alpha, \phi, \theta) = \frac{1}{(2\pi\sigma^2)^{MN}} \exp\left\{-\frac{\|Y - \alpha A(\phi, \theta)s\|^2}{2\sigma^2}\right\}
\]

\[
= \frac{1}{(2\pi\sigma^2)^{MN}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{m=1}^{M} \sum_{n=1}^{N} (y_{m,n} - x_{m,n})^2\right\}. \tag{59}
\]

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

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

where

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

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

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

**Proof:** See Appendix.

Accordingly, the CRLB for the parameters of channel gain and DOA are \(CRLB = I^{-1}(\Theta)\), thus we have

\[
\text{var}(\hat{\alpha}) \geq A,
\]

\[
\text{var}(\hat{w}_1) \geq \frac{2\sigma^2 M \sum_{n=0}^{N-1} n^2}{(2\pi\alpha)^2 [N \sum_{m=0}^{M-1} m^2] [M \sum_{n=0}^{N-1} n^2] - (2\pi\alpha)^2 \sum_{m=0}^{M-1} m \sum_{n=0}^{N-1} n [\sum_{m=0}^{M-1} m \sum_{n=0}^{N-1} n]}
\]

\[
= AB \frac{6(2N - 1)}{M - 1},
\]

\[
\text{var}(\hat{w}_2) \geq \frac{2\sigma^2 N \sum_{m=0}^{M-1} m^2}{(2\pi\alpha)^2 [N \sum_{m=0}^{M-1} m^2] [M \sum_{n=0}^{N-1} n^2] - (2\pi\alpha)^2 \sum_{m=0}^{M-1} m \sum_{n=0}^{N-1} n [\sum_{m=0}^{M-1} m \sum_{n=0}^{N-1} n]}
\]

\[
= AB \frac{6(2M - 1)}{N - 1}, \tag{63}
\]
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

\[
\text{var}(\hat{\theta}) \geq AB \frac{6(2N-1)\lambda^2}{(M-1)d^2 \sin^2 \theta}, \\
\text{var}(\hat{\phi}) \geq AB \frac{6\lambda^2(2M-1)(M-1) + 6(N-1)Cd \cos \theta \cos \phi}{(M-1)(N-1)d^2 \sin^2 \theta \sin^2 \phi},
\]

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 $\theta$ and elevation angle $\phi$ are required. Therefore, we can use the following transformation to estimate the real angles of azimuth and elevation:

\[
g(\Theta) = \begin{bmatrix} \alpha \\ \theta \\ \phi \end{bmatrix} = \begin{bmatrix} \alpha \\ \arccos(\frac{\lambda w_1}{d}) \\ \arcsin(\frac{\lambda w_2}{d \sin \theta}) \end{bmatrix}. \tag{65}
\]

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

\[
\text{var}(\hat{\theta}) \geq \left[ \frac{\partial g(\Theta)}{\partial \Theta} \right]^{-1} \left[ \frac{\partial g(\Theta)}{\partial \Theta} \right]_{2,2}, \\
\text{var}(\hat{\phi}) \geq \left[ \frac{\partial g(\Theta)}{\partial \Theta} \right]^{-1} \left[ \frac{\partial g(\Theta)}{\partial \Theta} \right]_{3,3}. \tag{66}
\]

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

\[
\text{var}(\hat{\theta}) \geq AB \frac{6(2N-1)\lambda^2}{(M-1)d^2 \sin^2 \theta}, \\
\text{var}(\hat{\phi}) \geq AB \frac{6\lambda^2(2M-1)(M-1) + 6(N-1)Cd \cos \theta \cos \phi}{(M-1)(N-1)d^2 \sin^2 \theta \sin^2 \phi}. \tag{67}
\]

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

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
Fig. 3. Comparison of MSE performances 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.

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, where $L = 10$ paths for each user, and the default value of training sequences $\tau$ is assumed 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 MSE performances 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 performance of proposed DOA estimation method improves with increasing the searching grid, which is due to the improved angle resolution. Theoretically, if the searching grid goes to infinity, the proposed
Fig. 4. The MSE performance 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.

DOA estimation method can achieve the same MSE performance 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 MSE performance of the theoretical bound is very close to the corresponding CRLB.

Fig. 4 plots the MSE performances of DOA estimation as a function of SNR for various URA sizes. We assume that the total transmit power for each BS antenna are 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 performance of the proposed channel gain estimation method with the corresponding CRLB as a function of SNR. It can be seen that the MSE performance 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 proposed channel gain estimation method is very close to the CRLB, especially in the large BS antennas scenario.

Fig. 6 compares the MSE performances of the proposed channel estimation method, the eigen-decomposition based method [15], the beamspace based method [24], and the CS-based method [21]. It can be seen that the MSE performance 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 $M \times N$ dimensional channel covariance matrix for JSDM in practice. It is important to know that the proposed method, the beamspace based method, and the CS-based method could directly handle the instantaneous channel estimation, but the beamspace based method and CS-based method has an error floor due to the power leakage problem in its sparse channel representation.
Fig. 6. The channel estimation MSE performances of the proposed method, eigen-decomposition method, beamspace method, and the CS-based method.

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

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 based 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 based method suffers from severe channel power leakage, but most of the channel power concentrated 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 based 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 achievable 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 based method becomes smaller. Our results clearly demonstrate the effectiveness of the proposed method.

VI. CONCLUSION

In this paper, we proposed a novel DOA estimation and channel gain 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 proposed 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 performance and CRLB performance 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

$$y_d = F_H \text{vec}(A(w_1, w_2)) a_s, \quad (68)$$

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

$$0 = \frac{\partial g(w_1, w_2)}{\partial w_i} \big|_{w_i = \hat{w}_i} \approx \frac{\partial g(w_1, w_2)}{\partial w_i} \big|_{w_i = \hat{w}_i} + \frac{\partial^2 g(w_1, w_2)}{\partial^2 w_i} \big|_{w_i = \hat{w}_i} \Delta w_i, \quad (69)$$

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

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

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

$$\dot{g}(w_1, w_2|w_i) = j y^H F_{RF}^H W_i P_a^\dagger (F_{RF}^H)^\dagger y - j y^H F_{RF}^H P_a^\dagger W_i (F_{RF}^H)^\dagger y. \quad (71)$$
Since $P_a^\perp (F_{RF}^H)\dagger y_d = 0$, we can rewrite (71) as
\[
\dot{g}(w_1, w_2|w_i) = j(y_d + n)^H F_{RF}^H W_i P_a^\perp (F_{RF}^H)\dagger (y_d + n) - j(y_d + n)^H F_{RF}^H P_a^\perp W_i (F_{RF}^H)\dagger (y_d + n) \\
= -2\Re\{y_d^H F_{RF}^H W_i P_a^\perp (F_{RF}^H)\dagger n\} - 2\Re\{n^H F_{RF}^H W_i P_a^\perp (F_{RF}^H)\dagger n\}.
\] (72)

Due to the independence between $y_d$ and $n$, we can obtain
\[
E\{\dot{g}(w_1, w_2|w_i)\} = -2\sigma_n^2 \Re\{tr\{F_{RF}^H W_i P_a^\perp (F_{RF}^H)\dagger\}\} = 0.
\] (73)

The second-order derivative can be calculated as
\[
\ddot{g}(w_1, w_2|w_i) = -y^H F_{RF}^H W_i P_a^\perp (F_{RF}^H)\dagger n + y^H F_{RF}^H W_i P_a^\perp W_i (F_{RF}^H)\dagger y \\
+ y^H F_{RF}^H W_i P_a^\perp W_i (F_{RF}^H)\dagger y - n^H F_{RF}^H P_a^\perp W_i^2 (F_{RF}^H)\dagger y.
\] (74)

Based on (74), we have
\[
E\{\ddot{g}(w_1, w_2|w_i)\} = 2y^H F_{RF}^H W_i P_a^\perp W_i (F_{RF}^H)\dagger y_d + E\{tr(-W_i^2 P_a^\perp + 2W_i P_a^\perp W_i - P_a^\perp W_i^2)\} \\
= 2\sigma_n^2 |\alpha|^2 \text{vec}(A^H(w_1, w_2)) W_i P_a^\perp W_i \text{vec}(A(w_1, w_2)).
\] (75)

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

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

Similarly, $\dot{g}(w_1, w_2|w_i)$ can be expressed as
\[
\dot{g}(w_1, w_2|w_i) = O_1(n) + O_1(n^2),
\] (77)

where $O_1(n)$ and $O_1(n^2)$ represent the linear and quadrature functions of $n$ existing in $\dot{g}(w_1, w_2|w_i)$.

Substituting (76) and (77) into (70) and assuming high SNR, i.e., $||n||^2 \ll ||y_d||^2$, we have
\[
\Delta w_i = - \frac{O_1(n) + O_1(n^2)}{E\{\ddot{g}(w_1, w_2|w_i)\}} + O_2(n) + O_2(n^2) \\
= - \frac{O_1(n) + O_1(n^2)}{E\{\ddot{g}(w_1, w_2|w_i)\}} \times \left(1 - \frac{O_2(n) + O_2(n^2)}{E\{\dot{g}(w_1, w_2|w_i)\}} + \left(\frac{O_2(n) + O_2(n^2)}{E\{\dot{g}(w_1, w_2|w_i)\}}\right)^2 - \cdots \right).
\] (78)

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

$$
E\{\Delta w_i\} = E\left\{ \frac{\partial}{\partial w_i} \ln f(Y|\alpha, \phi, \theta) \right\} = -\frac{1}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \cos(4\pi mw_1 + 4\pi nw_2),
$$

$$
E\{\Delta w_i^2\} = E\left\{ \left( \frac{\partial}{\partial w_i} \ln f(Y|\alpha, \phi, \theta) \right)^2 \right\} = \frac{1}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [m^2 - m_1^2 \cos(4\pi mw_1 + 4\pi nw_2)],
$$

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

$$
E\{\hat{g}(w_1, w_2|w_i)\} = 2E\{y_d^H F_{RF}^H W_i P_a^\perp (F_{RF}^H)^\dagger n n^H F_{RF}^H P_a^\perp W_i (F_{RF}^H)^\dagger y_d \}
$$

$$
+ E\{(n^H F_{RF}^H (W_i P_a^\perp - P_a^\perp W_i) (F_{RF}^H)^\dagger n)^2\},
$$

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

$$
E\{\hat{g}(w_1, w_2|w_i)\} = 2\sigma_n^2 y_d^H F_{RF}^H W_i P_a^\perp W_i (F_{RF}^H)^\dagger y_d
$$

$$
= 2\sigma_n^2 \sigma_o^2 |\alpha|^2 \text{vec}(A^H(w_1, w_2)) W_i P_a^\perp W_i \text{vec}(A(w_1, w_2)).
$$

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

$$
E\{\Delta w_i\} = 0,
$$

$$
E\{\Delta w_i^2\} = \frac{\sigma_n^2}{2\sigma_o^2 |\alpha|^2 \text{vec}(A^H(w_1, w_2)) W_i P_a^\perp W_i \text{vec}(A(w_1, w_2))}. \tag{83}
$$

The proof is completed.

**B. Proof of Lemma 2**

Performing mathematical calculation on (60), we derive

$$
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial^2 \Theta_1} = -\frac{MN}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \cos(4\pi mw_1 + 4\pi nw_2), \tag{84}
$$

$$
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial^2 \Theta_2} = -\frac{(2\pi)^2}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [m^2 - m_1^2 \cos(4\pi mw_1 + 4\pi nw_2)], \tag{85}
$$

$$
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial^2 \Theta_3} = -\frac{(2\pi)^2}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [n^2 - n_1^2 \cos(4\pi mw_1 + 4\pi nw_2)], \tag{86}
$$

\[=\]
\[
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_1 \partial \Theta_2} = \frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_2 \partial \Theta_1} = \frac{2\pi \alpha}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} m \sin(4\pi mw_1 + 4\pi nw_2), \tag{88}
\]
\[
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_1 \partial \Theta_3} = \frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_3 \partial \Theta_1} = \frac{2\pi \alpha}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} n \sin(4\pi mw_1 + 4\pi nw_2), \tag{89}
\]
\[
\frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_2 \partial \Theta_3} = \frac{\partial^2 \ln f(Y|\alpha, \phi, \theta)}{\partial \Theta_3 \partial \Theta_2} = \frac{(2\pi \alpha)^2}{2\sigma^2} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [mn(\cos(4\pi mw_1 + 4\pi nw_2) - 1)]. \tag{90}
\]

**Lemma 4:** For \(x \in [0, 2\pi)\), we have
\[
\lim_{K \to \infty} \sum_{k=1}^{K} k^i \sin(4\pi kx) = 0, \quad i = 0, 1, 2 \tag{91}
\]
\[
\lim_{K \to \infty} \sum_{k=1}^{K} k^i \cos(4\pi kx) = 0, \quad i = 0, 1, 2 \tag{92}
\]

**Proof:** According to [46, Eq. (AD361)], we have
\[
\lim_{K \to \infty} \sum_{k=1}^{K} \sin(4\pi kx) = \frac{\sin((K + 1)2\pi x) \sin(2\pi Kx)}{\sin(2\pi x)} = 0, \tag{93}
\]
\[
\lim_{K \to \infty} \sum_{k=1}^{K} \cos(4\pi kx) = \frac{\cos((K + 1)2\pi x) \sin(2\pi Kx)}{\sin(2\pi x)} = 0, \tag{94}
\]
\[
\lim_{K \to \infty} \sum_{k=1}^{K} k \sin(4\pi kx) = \frac{\sin 4\pi Kx}{4\sin^2(2\pi x)} - \frac{K \cos((2K - 1)2\pi x)}{2\sin(2\pi x)} = 0, \tag{95}
\]
\[
\lim_{K \to \infty} \sum_{k=1}^{K} k \cos(4\pi kx) = \frac{K \sin((2K - 1)2\pi x)}{2\sin(2\pi x)} - \frac{1 - \cos 4\pi Kx}{4\sin^2(2\pi x)} = 0, \tag{96}
\]
for \(i = 2\), where the limits follow from relations similar to (96).

As \(M\) and \(N\) go to infinity in massive MIMO systems, we have
\[
\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \cos(4\pi mw_1 + 4\pi nw_2) = 0, \quad \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} m \sin(4\pi mw_1 + 4\pi nw_2) = 0, \tag{97}
\]
\[
\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} m^2 \cos(4\pi mw_1 + 4\pi nw_2) = 0, \quad \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} mn \cos(4\pi mw_1 + 4\pi nw_2) = 0.
\]

Substituting (84)–(90) into (60), the FIM for the joint channel gain and DOA estimation of a hybrid mmWave massive MIMO system can be expressed as
\[
I(\Theta) = \frac{1}{2\sigma^2} \begin{bmatrix}
MN & 0 & 0 \\
0 & (2\pi \alpha)^2 N \sum_{m=0}^{M-1} m^2 & (2\pi \alpha)^2 \sum_{m=0}^{M-1} m \sum_{n=0}^{N-1} n \\
0 & (2\pi \alpha)^2 \sum_{m=0}^{M-1} m \sum_{n=0}^{N-1} n & (2\pi \alpha)^2 M \sum_{n=0}^{N-1} n^2
\end{bmatrix}. \tag{98}
\]
The proof is completed.

REFERENCES


