



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/142274/>

Version: Accepted Version

Proceedings Paper:

You, Y, Zhang, L and Liu, M (2019) IP Aided OMP Based Channel Estimation for Millimeter Wave Massive MIMO Communication. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2019). IEEE Wireless Communications and Networking Conference (WCNC 2019), 15-18 Apr 2019, Marrakesh, Morocco. IEEE. ISBN: 978-1-5386-7646-2. ISSN: 1558-2612. EISSN: 1558-2612.

<https://doi.org/10.1109/WCNC.2019.8885881>

© 019 IEEE. This is an author produced version of a paper published in Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2019). Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

IP Aided OMP Based Channel Estimation for Millimeter Wave Massive MIMO Communication

You You

School of Electronic and
Electrical Engineering
University of Leeds
Email: el14yy@leeds.ac.uk

Li Zhang

School of Electronic and
Electrical Engineering
University of Leeds
Email: l.x.zhang@leeds.ac.uk

MiaoMiao Liu

School of Electronic and
Electrical Engineering
University of Leeds
Email: ml13m2l@leeds.ac.uk

Abstract—Adoption of hybrid precoding is the key element of reducing the hardware cost of radio-frequency compared with the conventional full-digital precoding approach in millimeter wave (mmWave) MIMO systems. For hybrid precoding, channel state information (CSI) is needed. However, the use of analog precoding in hybrid architecture and the large antenna array make channel estimation difficult for mmWave system. It has been shown that mmWave channels exhibit sparsity, thus compressive sensing (CS) techniques can be leveraged to conduct channel estimation. Conventional CS based channel estimation methods for mmWave MIMO are based on quantized angle grid. However, the performance would be severely affected by off grid angles which can not be improved by increasing the resolution because this will increase the coherence between the grid points. In this paper, we propose an Interior Point (IP) aided orthogonal matching pursuit (OMP) algorithm. It significantly improves the channel estimation accuracy by reducing the estimation error of angle-of-departure (AoD) and angle-of-arrival (AoA). The simulation results demonstrate the advantage of the proposed IP-OMP over the existing methods such as least squares and the conventional OMP.

I. INTRODUCTION

Millimeter wave (mmWave) communication is considered a potential approach for the future wireless systems because of the huge amount of available spectrum [1]. Thanks to the short wave length, large antenna arrays can be packed into small areas so that mmWave systems can achieve desirable beamforming gains to overcome the huge propagation loss compared with existing micro wave systems. However, cost and power consumption of device operating at mmWave make it unfeasible to apply full-digital MIMO architecture as in micro wave communication [2]. To overcome the radio frequency (RF) power consumption, a hybrid MIMO architecture consisting of an analog beamformer in radio frequency domain cascaded with a digital MIMO processor in baseband has been proposed for mmWave communication [3].

As in conventional microwave systems, channel state information (CSI) is needed to design precoding and combining procedures at transmitters and receivers in mmWave systems. Channel estimation for mmWave hybrid MIMO systems is challenging, because the number of antennas in mmWave at both transmitter and receiver can be much larger than that in conventional microwave massive MIMO due to the much smaller wavelength of mmWave. The different “virtual array”

channel models widely used for mmWave massive MIMO [5] and the new constraints on the hardware of hybrid architecture also make channel estimation different in millimeter wave frequencies with that at lower frequencies [6]. Hence new channel estimation methods are required. Due to the sparse nature of mmWave channel [7], compressive sensing (CS) theory [8] can be leveraged to effectively estimate mmWave channels [10]. Instead of estimating all the entries in the channel matrix, only the angle-of-departure (AoD), angle-of-arrival (AoA) of dominant paths and the corresponding path gains are estimated [9]. Recently, a closed-loop beam training-based method were proposed in [10]. This method is a multistage process that can avoid exhaustive search. In [10], they first estimate the AoDs/AoAs by closed-loop beam training and then estimate the path gain associated with each pair of AoD and AoA. At each stage the transmitter emits the pilot beams, and the receiver selects the best beam and feeds back its decision. This process starts with wide beams that cover all of the angles of interest and improves the angle resolution only around the angles where AoDs/AoAs are present. The performance of close-loop method tends to be limited by the beamforming dictionary (codebook) designed for beam training. For example, an improved codebook employing continuous basis pursuit (CBP) method instead of conventional grid-based approach was proposed in [11]. Compared with [10], [11] significantly improves the estimation accuracy. However, the application of close-loop methods would be difficult for outdoor environment channel. Outdoor environment needs larger beamforming gain which prevents from using wide beam for first step beam training due to limited transmitted power.

An alternative approach is to use open-loop estimator exploiting sparse nature of mmWave channels. An open-loop channel estimator was developed using fixed width training beam without feedback from receiver in [12]. To reduce computation, orthogonal matching pursuit (OMP) algorithm was used to solve the sparse signal recovery problem [13]. Conventional OMP is a grid based algorithm. Despite the continuous nature of AoDs/AoAs, only G values are considered in estimation referred to as grid points (circles in Fig. 1). All AoDs/AoAs are approximated as the nearest grid points. The black dots indicate off grid angles. The approximation results in off grid errors. The estimation accuracy of grid based

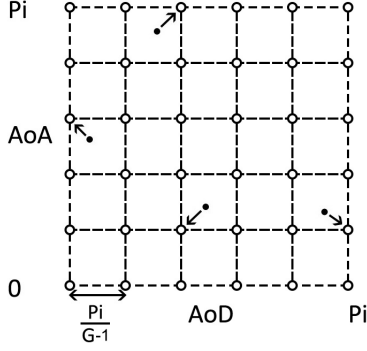


Fig. 1. An illustration of angle grid and the off grid angles.

CS algorithm is deteriorated by off grid angles severely. The impact of off grid angle errors is shown in the simulation results Fig. 3. A multi-grid OMP (MG-OMP) method was proposed to achieve better precision [12]. The MG-OMP starts with a coarse grid and refines the grid only around the regions where AoDs and AoAs are present. A finely quantized angle grid is proposed for OMP method in [14] to reduce the coherence of the redundant dictionary and improve estimation performance. While even with finely quantized angle grid [14] or multi-grid [12], off grid errors still exist and adversely affect the estimation performance.

In this paper, we propose an enhanced approach employing OMP algorithm for mmWave MIMO channel estimation, and IP algorithm to improve angle estimation accuracy. The proposed approach can be also employed with MG-OMP and other grid-based CS algorithms. It is shown that the IP-OMP can significantly improve the channel estimation accuracy by reducing the estimation error of AoDs/AoAs.

The organization of the paper is as follows. Section II presents the system model. The Least Square (LS) based channel estimation and CS based channel estimation problem are formulated and solved in Section III. The improved algorithm for the hybrid MIMO channel estimation are designed in Section IV. Simulation results illustrating the performance of the proposed algorithm are given in Section V. Finally, the conclusion is presented in Section VI.

II. SYSTEM MODEL

We consider a single user hybrid MIMO system shown in Fig. 2 [12], where the transmitter equipped with N_T antennas and N_{RF} RF chains communicating with a receiver with N_R antennas and N_{RF} RF chains ($N_{RF} \leq (\min N_T, N_R)$).

In the channel estimation stage, transmitter uses N_T^{Beam} ($N_T^{Beam} \leq N_T$) pilot beam training patterns denoted as $\{\mathbf{f}_m \in \mathbb{C}^{N_T \times 1} : m = 1, \dots, N_T^{Beam}\}$ and receiver uses N_R^{Beam} ($N_R^{Beam} \leq N_R$) beam patterns denoted as $\{\mathbf{w}_n \in \mathbb{C}^{N_R \times 1} : n = 1, \dots, N_R^{Beam}\}$. At training period, transmitter sends training beams \mathbf{f}_m to receiver successively. We consider the transmitter beam \mathbf{f}_m one by one and each \mathbf{f}_m is received

through receiver beam patterns \mathbf{w}_n . Because the receiver has N_{RF} RF chains. The receiver can generate N_{RF} receiver beams simultaneously and receive signal $\mathbf{y}_q \in \mathbb{C}^{N_{RF} \times 1}$ for $q \in \{1, \dots, N_R^{Block}\}$ in one time slot. Here q denotes the received block index and $N_R^{Block} = \frac{N_R^{Beam}}{N_{RF}}$ is the number of received blocks. We assume N_R^{Beam} and N_T^{Beam} are multiples of N_{RF} . Collecting all q received block signals can represent the received signal $\mathbf{y}_m \in \mathbb{C}^{N_R^{Beam} \times 1}$ for one transmitter beam \mathbf{f}_m in q time slots. The received vector for the q -th block and the m -th transmit beam is given by

$$\mathbf{y}_{q,m} = \mathbf{W}_q^H \mathbf{H} \mathbf{f}_m x_p + \mathbf{W}_q^H \mathbf{n}_{q,m}, \quad (1)$$

where $\mathbf{W}_q = [\mathbf{w}_{(q-1)N_{RF}+1}, \dots, \mathbf{w}_{qN_{RF}}] \in \mathbb{C}^{N_R \times N_{RF}}$ is the beam pattern matrix for RF receive beam patterns at one time slot for \mathbf{f}_m . x_p is the transmitted pilot symbol. Each \mathbf{w} is a beam pattern generated by one RF chain at receiver. $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ represents the channel matrix, and $\mathbf{n} \in \mathbb{C}^{N_R \times 1}$ is the noise vector. Collecting $\mathbf{y}_{q,m}$ for $q \in \{1, \dots, N_R^{Block}\}$, we get $\mathbf{y}_m \in \mathbb{C}^{N_R^{Beam} \times 1}$ given by

$$\mathbf{y}_m = \mathbf{W}^H \mathbf{H} \mathbf{f}_m x_p + \text{diag}(\mathbf{W}_1^H, \dots, \mathbf{W}_{N_R^{Block}}^H) \times [\mathbf{n}_{1,m}^T, \dots, \mathbf{n}_{N_R^{Block},m}^T]^T, \quad (2)$$

where $\mathbf{W} = [\mathbf{W}_1, \dots, \mathbf{W}_{N_R^{Block}}] \in \mathbb{C}^{N_R \times N_R^{Beam}}$. \mathbf{y}_m is the received signal for \mathbf{f}_m in q time slots. To represent the signals for all N_T^{Beam} transmit beams, we collect \mathbf{y}_m for $m \in \{1, \dots, N_T^{Beam}\}$ to get

$$\mathbf{Y} = \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{X} + \mathbf{N} = \sqrt{P} \mathbf{W}^H \mathbf{H} \mathbf{F} + \mathbf{N} \quad (3)$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_{N_T^{Beam}}] \in \mathbb{C}^{N_R^{Beam} \times N_T^{Beam}}$, $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_{N_T^{Beam}}] \in \mathbb{C}^{N_T \times N_T^{Beam}}$ and $\mathbf{N} \in \mathbb{C}^{N_R^{Beam} \times N_T^{Beam}}$ is the noise matrix given by

$$\mathbf{N} = \text{diag}(\mathbf{W}_1^H, \dots, \mathbf{W}_{N_R^{Block}}^H) [\mathbf{n}_{1,1}^T, \dots, \mathbf{n}_{N_R^{Block},1}^T]^T, \quad (4)$$

$$\dots, [\mathbf{n}_{1,N_T^{Beam}}^T, \dots, \mathbf{n}_{N_R^{Block},N_T^{Beam}}^T]^T.$$

The matrix $\mathbf{X} \in \mathbb{C}^{N_T^{Beam} \times N_T^{Beam}}$ is a diagonal matrix with x_p on its diagonal. Throughout the paper, we assume identical pilot symbols so that $\mathbf{X} = \sqrt{P} \mathbf{I}_{N_T^{Beam}}$ where P is the pilot power.

In the mmWave communication, hybrid MIMO architecture is employed. The transmit and receive training matrices are regarded as hybrid beamforming matrix and they can be decomposed as $\mathbf{F} = \mathbf{F}_{RF} \mathbf{F}_{BB}$ and $\mathbf{W} = \mathbf{W}_{RF} \mathbf{W}_{BB}$, where $\mathbf{F}_{RF} \in \mathbb{C}^{N_T \times N_T}$ and $\mathbf{W}_{RF} \in \mathbb{C}^{N_R \times N_R}$ represent the RF beamforming matrices, $\mathbf{F}_{BB} \in \mathbb{C}^{N_T \times N_T^{Beam}}$ and $\mathbf{W}_{BB} \in \mathbb{C}^{N_R \times N_R^{Beam}}$ represent the baseband processing matrices. In this case, (3) can be formulated as

$$\mathbf{Y} = \sqrt{P} (\mathbf{W}_{RF} \mathbf{W}_{BB})^H \mathbf{H} (\mathbf{F}_{RF} \mathbf{F}_{BB}) + \mathbf{N}. \quad (5)$$

\mathbf{F}_{RF} , \mathbf{W}_{RF} , \mathbf{W}_{BB} and \mathbf{F}_{BB} will be designed in Section V.

The mmWave narrowband channel can be approximated by a geometric channel mode with L scatters due to its limited

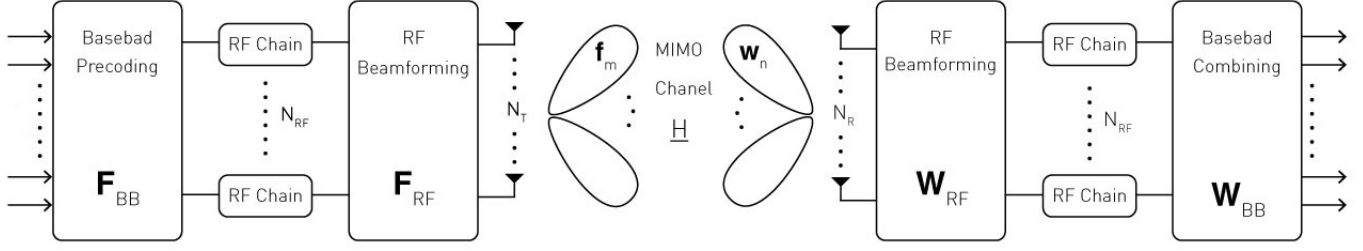


Fig. 2. Hybrid Massive MIMO system for mmWave communication

scattering feature [3]. Each scatterer contributes only one path of propagation between transmitter and receiver. The channel matrix can be written as

$$\mathbf{H} = \sqrt{\frac{N_T N_R}{L}} \sum_{\ell=1}^L \alpha_{\ell} \mathbf{a}_r(\theta_{\ell}^r) \mathbf{a}_t^H(\theta_{\ell}^t), \quad (6)$$

where L is the number of scatterers, α_{ℓ} is the complex gain, θ_{ℓ}^r and θ_{ℓ}^t are the AoA and AoD of the ℓ -th path, respectively. We assume the uniform linear arrays (ULA) whose array response vectors are denoted as $\mathbf{a}_r(\theta_{\ell}^r) \in \mathbb{C}^{N_R \times 1}$ for the receiver and $\mathbf{a}_t(\theta_{\ell}^t) \in \mathbb{C}^{N_T \times 1}$ for the transmitter. For an N -element ULA, the steering vector can be given by

$$\mathbf{a}(\theta) = [1, e^{-j2\pi\vartheta}, e^{-j4\pi\vartheta}, \dots, e^{-j2\pi\vartheta(N-1)}]^T, \quad (7)$$

where the normalized spatial angle ϑ is related to the physical angle (of arrival or departure) $\theta \in [0, \pi)$ as $\vartheta = \frac{d}{\lambda} \cos \theta = \beta \cos \theta$, d denotes the antenna spacing, λ denotes the wavelength of operation and β is the normalized antenna spacing. We assume that $N = N_T$ when $\mathbf{a}_t(\theta_t)$ represents the array weights needed to transmit a beam focused in direction θ_t , and $N = N_R$ when $\mathbf{a}_r(\theta_r)$ represents the signal response at the receiver array due to a point source in direction θ_r . In this paper, we consider $d = \frac{\lambda}{2}$. The channel gains $\{\alpha_{\ell}\}_{\ell=1}^L$ are modeled by i.i.d. random variables with distribution $\mathcal{CN}(0, \sigma_{\alpha}^2)$. The AoAs and AoDs are modeled by the Laplacian distribution whose mean is uniformly distributed over $[0, \pi)$, and angular standard deviation is σ_{AS} . The channel model in (6) can be rewritten in matrix form as

$$\mathbf{H} = \mathbf{A}_R \mathbf{H}_a \mathbf{A}_T^H, \quad (8)$$

where $\mathbf{H}_a = \sqrt{\frac{N_T N_R}{L}} \text{diag}(\alpha_1, \dots, \alpha_{\ell}, \dots, \alpha_L)$, $\mathbf{A}_R = [\mathbf{a}_r(\theta_1^r), \dots, \mathbf{a}_r(\theta_{\ell}^r), \dots, \mathbf{a}_r(\theta_L^r)] \in \mathbb{C}^{N_R \times L}$, and $\mathbf{A}_T = [\mathbf{a}_t(\theta_1^t), \dots, \mathbf{a}_t(\theta_{\ell}^t), \dots, \mathbf{a}_t(\theta_L^t)] \in \mathbb{C}^{N_T \times L}$.

To exploit the sparsity of mmWave channel, typically, the AoAs/AoDs (θ_t, θ_r) are estimated as one point in an uniform grid of size G as $\varphi_{\ell}^t, \varphi_{\ell}^r$ ($\varphi_{\ell}^t, \varphi_{\ell}^r \in \{0, \frac{\pi}{G-1}, \dots, \frac{\pi(G-1)}{G-1}\}$), with $G \gg L$ to achieve desired resolution in [3],[10],[12]. $\bar{\mathbf{A}}_T = [\mathbf{a}_t(\varphi_1^t), \dots, \mathbf{a}_t(\varphi_{\ell}^t), \dots, \mathbf{a}_t(\varphi_G^t)] \in \mathbb{C}^{N_T \times G}$ and $\bar{\mathbf{A}}_R = [\mathbf{a}_r(\varphi_1^r), \dots, \mathbf{a}_r(\varphi_{\ell}^r), \dots, \mathbf{a}_r(\varphi_G^r)] \in \mathbb{C}^{N_R \times G}$ are defined as array response matrices. Using these matrices, \mathbf{H} can be approximated in terms of a L -sparse matrix $\mathbf{H}_b \in \mathbb{C}^{G \times G}$,

with L non zero elements in the positions corresponding to the AoAs and AoDs.

$$\mathbf{H} = \bar{\mathbf{A}}_R \mathbf{H}_b \bar{\mathbf{A}}_T^H + \mathbf{E} \quad (9)$$

There is a grid error \mathbf{E} in (9), because the true continuous AoDs/AoAs do not fall onto the uniform grid points precisely as illustrated in Fig. 1. Intuitively, the grid errors can be mitigated by increasing the grid size. However, for CS channel estimation in mmWave communication, using larger G is not desirable due to the increasing coherence between steering vectors for each grid angle. In this case, the sensing matrix employed in channel estimation does not satisfy Restricted Isometry Property (RIP) and leads to even worse estimation performance [16]. Also a larger G leads to exponentially increasing complexity of OMP algorithm. Most works on channel estimation for mmWave MIMO communication leave grid errors as unexplored area. Therefore, to improve the achievable channel estimation performance with a reasonable complexity, in this paper, we propose to employ Interior Point (IP) method to minimize the off grid angle error and refine the grid accordingly in every iteration of OMP algorithm.

III. FORMULATION OF MMWAVE CHANNEL ESTIMATION PROBLEM

In this section, two different formulations of mmWave channel estimation problem are presented. Least Square formulation is first presented for the purpose of comparison.

A. Least Square Channel Estimation

To formulate the channel estimation problem, it is necessary to vectorize the received signal matrix \mathbf{Y} in (5). Denoting $\text{vec}(\mathbf{Y})$ by \mathbf{y}_v and therefore (5) is rewritten as

$$\begin{aligned} \mathbf{y}_v &= \sqrt{P} (\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H) \cdot \text{vec}(\mathbf{H}) \\ &\quad + \text{vec}(\mathbf{N}) \\ &= \mathbf{Q} \cdot \text{vec}(\mathbf{H}) + \mathbf{n}_Q, \end{aligned} \quad (10)$$

Using the property of the Khatri-Rao product

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \cdot \text{vec}(\mathbf{B}). \quad (11)$$

The matrices $\mathbf{W}_{BB}^H \mathbf{W}_{RF}^H$, \mathbf{H} and $\mathbf{F}_{RF} \mathbf{F}_{BB}$ in (5) are regarded as \mathbf{A} , \mathbf{B} and \mathbf{C} in (11) respectively. \mathbf{n}_Q is the vectorized noise. Let $\mathbf{Q} = \sqrt{P} (\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H) \in \mathbb{C}^{N_T^{beam} N_R^{beam} \times N_T N_R}$, a natural approach to estimating

$\text{vec}(\mathbf{H})$ is the LS approach, which results in a closed-form solution given by $(\mathbf{Q}^H \mathbf{Q})^{-1} \mathbf{Q}^H \mathbf{y}_v$. The use of LS solution for mmWave communication is difficult. Because the LS solution requires $N_T^{Beam} N_R^{Beam} \geq N_T N_R$ so that $\mathbf{Q}^H \mathbf{Q}$ has full rank. However, (N_T, N_R) are usually large integers for mmWave MIMO system and $N_T^{Beam} N_R^{Beam} \leq N_T N_R$. This difficulty can be overcome in the CS approach because the number of entries to be estimated in the CS formulation is proportional to the sparsity level which is much less than $(N_T N_R)$.

B. Compressive Sensing Channel Estimation

Considering the system model in (5) and channel model in (9) neglecting grid error \mathbf{E} , the mmWave channel estimation can be formulated as a sparse problem by vectorizing \mathbf{Y} in (5). Using property of Khatri-Rao product (11) for (5) and \mathbf{H}_b in (9), equation (10) can be rewritten as

$$\begin{aligned} \mathbf{y}_v &= \sqrt{P} (\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H) \cdot \text{vec}(\mathbf{H}_b) + \mathbf{n}_Q \\ &= \sqrt{P} (\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H) \mathbf{A}_D \mathbf{h}_b + \mathbf{n}_Q \quad (12) \\ &= \bar{\mathbf{Q}} \cdot (\mathbf{h}_b) + \mathbf{n}_Q, \end{aligned}$$

where $\mathbf{A}_D = \bar{\mathbf{A}}_T^* \otimes \bar{\mathbf{A}}_R$ is an $N_T N_R \times G^2$ dictionary matrix that consists of the G^2 column vectors of the form $\mathbf{a}_t^H(\theta_u) \otimes \mathbf{a}_r(\theta_v)$, with θ_u and θ_v , the u th and v th points, respectively, of the angle uniform grid. For example, $\theta_u = \pi u / (G - 1)$ ($u = 0, 1, \dots, G - 1$) and $\theta_v = \pi v / (G - 1)$ ($v = 0, 1, \dots, G - 1$). $\mathbf{h}_b = \text{vec}(\mathbf{H}_b)$ is an $G^2 \times 1$ vector which represents the path gains of the corresponding quantized directions. In (12), $\bar{\mathbf{Q}} = \sqrt{P} (\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H) \mathbf{A}_D \in \mathbb{C}^{N_T^{Beam} N_R^{Beam} \times G^2}$ is the sensing matrix. The formulation of the vectorized received signal in (12) represents a sparse formulation of the channel estimation problem as \mathbf{h}_b has only L non-zero elements and $L \ll G^2$. This implies that the number of required measurements to detect the non-zero elements of \mathbf{h}_b is much less than G^2 . Given the formulation in (12), CS algorithms such as OMP can be adapted to solve this channel estimation problem.

IV. PROPOSED ESTIMATION ALGORITHM FOR MMWAVE MIMO CHANNELS

Considering the previous estimation problem using CS method in (12), given that the true continuous-domain AoDs and AoAs may lie off the grid, the grid representation in this case will result in the degradation of estimation performance. This can be mitigated to a certain extent by finer discretization of the grid, but that may lead to longer computation time and higher mutual coherence of the sensing matrix, thus becoming less effective for sparse signal recovering. To effectively estimate the position of non-zero values, and consequently the corresponding AoDs/AoAs and path gains, OMP method is used in conjunction with the IP method. In this paper, it is named as IP-OMP. The proposed IP-OMP algorithm solving (12) is summarized in Algorithm 1.

Algorithm 1 operates as follows. In the initial stage, when $t = 1$, this algorithm chooses the column j of $\bar{\mathbf{Q}}$ that is the most strongly correlated with the residual \mathbf{r}_{t-1} in step

3. Each column index obtained in step 3 corresponds to an AoD/AoA pair and is called AoD/AoA pair index. In step 4, the column number j is added to set Ω_t . The most strongly correlated column in $\bar{\mathbf{Q}}$ is determined by the column of the dictionary matrix \mathbf{A}_D when hybrid precoding and combining matrix are given. Because $\mathbf{A}_D = \bar{\mathbf{A}}_T^* \otimes \bar{\mathbf{A}}_R$ is an $N_T N_R \times G^2$ dictionary matrix that consists of the G^2 column vectors of the form $\mathbf{a}_t^H(\theta_u) \otimes \mathbf{a}_r(\theta_v)$, with θ_u and θ_v , the u th and v th discrete points, respectively, of the uniform angle grid. We can find the estimated AoD/AoA value through column index j in the t th iteration as $AoD_t = 0 + \text{ceil}(\frac{j}{G}) \frac{\pi}{G-1}$ and $AoA_t = 0 + (\text{mod}(j - 1, G) + 1) \frac{\pi}{G-1}$ where $u = \text{ceil}(\frac{j}{G})$, $v = \text{mod}(j - 1, G) + 1$ as described in step 5. However, the main problem for conventional OMP method is that the off grid angles deteriorate the accuracy in step 3. Because the true AoD/AoA are continuous values instead of the discrete values in step 5. It means that, in step 3, $|\bar{\mathbf{Q}}(i)^H \mathbf{r}_{t-1}|$ can be even larger than the value corresponding to the j th column if we can choose a more accurate AoD/AoA pair. In this case, it is obvious that we can obtain improved AoD/AoA pair through maximizing $|\bar{\mathbf{Q}}(i)^H \mathbf{r}_{t-1}|$. Considering the order of complexity, we choose to employ IP method to minimize the off grid error and estimate more accurate AoD/AoA pair index based on the result from step 3. We set $x_t = (AoD_t, AoA_t)$ as original point corresponding to the j th column in $\bar{\mathbf{Q}}$. We define the correlation between the sensing column and the residual as $f(AoD'_t, AoA'_t) = |((\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H)(\mathbf{a}^*(AoD'_t) \otimes \mathbf{a}(AoA'_t)))^H \mathbf{r}_{t-1}|$ and set $-f(AoD'_t, AoA'_t)$ as objective function. Through minimizing objective function between the adjacent grid points, we can obtain new angle pair $x'_t = (AoD'_t, AoA'_t)$ which is most correlated with residual \mathbf{r}_{t-1} . This optimization problem in step 6 can be formulated as

$$\begin{aligned} & \min_{AoD'_t, AoA'_t} -f(AoD'_t, AoA'_t) \\ \text{s.t.} & \begin{cases} |AoD'_t - AoD_t| < \frac{\pi}{2(G-1)}, \\ |AoA'_t - AoA_t| < \frac{\pi}{2(G-1)}. \end{cases} \end{aligned}$$

When we obtain x'_t using IP method, the new most strongly correlated column is calculated as $\mathbf{p} = (\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H)(\mathbf{a}^*(AoD'_t) \otimes \mathbf{a}(AoA'_t))$ in step 7. Use \mathbf{p} to replace the column j in sensing matrix $\bar{\mathbf{Q}}$ as step 8. The updated matrix $\bar{\mathbf{Q}}$ is the new sensing matrix with the corrected grid. In this way, we adjust the grid point and sensing matrix in every iterative step to find a more accurate angle and corresponding path gain. The channel gains associated with the new grid points are obtained by evaluating the LS solution of $\mathbf{y}_v = \bar{\mathbf{Q}}_{\Omega_t} \mathbf{h}$ in step 9, where $\bar{\mathbf{Q}}_{\Omega_t} \in \mathbb{C}^{N_T^{Beam} N_R^{Beam} \times t}$ is the sub-matrix of $\bar{\mathbf{Q}}$ that only contains the columns whose indices are included in Ω_t and $\mathbf{h} \in \mathbb{C}^{t \times 1}$ is a vector with varying size. In step 10, the contributions of the chosen column vectors to \mathbf{y}_v are subtracted to update the residual. This procedure is repeated until $t = K$. In step 13, the algorithm constructs the sparse channel vector $\mathbf{h}_b \in \mathbb{C}^{G^2 \times 1}$ by putting K estimated channel gains into the corresponding position according to

elements in Ω_t . So that $\mathbf{h}_b(i) = \mathbf{h}_{t-1}$ for $i \in \Omega_{t-1}$ and $\mathbf{h}_b(i) = 0$, otherwise. \mathbf{h}_b is the channel matrix as in (12).

Algorithm 1 IP-OMP method for mmWave channel estimation

Require: sensing matrix $\bar{\mathbf{Q}}$, measurement vector \mathbf{y}_v , sparsity K and grid G

- 1: $\Omega_{t-1} = \text{empty set}$, residual $\mathbf{r}_0 = \mathbf{y}_v$, set the iteration counter $t = 1$
- 2: **while** $t \leq K$ **do**
- 3: $j = \arg \max_{i=1, \dots, G^2} |\bar{\mathbf{Q}}(i)^H \mathbf{r}_{t-1}|$
- 4: $\Omega_t = \Omega_{t-1} \cup \{j\}$
- 5: $AoD_t = 0 + \text{ceil}(\frac{j}{G}) \frac{\pi}{G}$
 $AoA_t = 0 + (\text{mod}(j-1, G) + 1) \frac{\pi}{G}$
 $x_t = (AoD_t, AoA_t)$
- 6: $\min_{AoD'_t, AoA'_t} f(AoD'_t, AoA'_t)$, $x'_t = (AoD'_t, AoA'_t)$
- 7: $\mathbf{p} = (\mathbf{F}_{BB}^T \mathbf{F}_{RF}^T \otimes \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H) (\mathbf{a}^*(AoD'_t) \otimes \mathbf{a}(AoA'_t))$
- 8: $\mathbf{Q}_j = \mathbf{p}$
- 9: $\mathbf{h}_t = \arg \min_{\mathbf{h}} \|\mathbf{y}_v - \bar{\mathbf{Q}}_{\Omega_t} \mathbf{h}\|_2$
- 10: $\mathbf{r}_t = \mathbf{y}_v - \bar{\mathbf{Q}}_{\Omega_t} \mathbf{h}_t$
- 11: $t = t + 1$
- 12: **end while**
- 13: $\mathbf{h}_b(i) = \mathbf{h}_{t-1}$ for $i \in \Omega_{t-1}$ and $\mathbf{h}_b(i) = 0$ otherwise
- 14: **return** \mathbf{h}_b

V. SIMULATION RESULTS

The performance of the proposed method is examined through computer simulation with the following parameters. ULAs are assumed at both transmitter and receiver with $N_T = N_R = 32$. They have DFT training beams with $N_T^{Beam} = N_R^{Beam} = 32$. All simulation results are averaged over 500 channel realizations with a carrier frequency of 60GHz. At each channel realization, the number of scatterers L is determined by $L = \max\{P_{10}, 1\}$ where P_{10} is the outcome of the Poisson random variable with mean 10. The design of analog/digital hybrid precoding and combining matrices have been extensively investigated [3], [4]. We use phase shifts to generate DFT beams for analog beamforming. So \mathbf{F}_{RF} and \mathbf{W}_{RF} can be designed as DFT matrices. The transmit and receive weight vectors are given by the columns of $N_T^{Beam} \times N_T^{Beam}$ and $N_R^{Beam} \times N_R^{Beam}$ DFT matrices respectively. We use the approach in [15] and [12] to generate precoding matrix for baseband through minimizing the coherence of sensing matrix $\bar{\mathbf{Q}}$. \mathbf{F}_{BB} and \mathbf{W}_{BB} are block diagonal matrices given by $\mathbf{F}_{BB} = \text{diag}(\mathbf{F}_{BB,1}, \dots, \mathbf{F}_{BB,i}, \dots, \mathbf{F}_{BB, N_T^{block}})$ and $\mathbf{W}_{BB} = \text{diag}(\mathbf{W}_{BB,1}, \dots, \mathbf{W}_{BB,i}, \dots, \mathbf{W}_{BB, N_R^{block}})$ whose diagonal entries, $\mathbf{F}_{BB,i}$ and $\mathbf{W}_{BB,i}$, consist of $N_{RF} \times N_{RF}$ complex valued matrices. $N_R^{block} = \frac{N_R^{Beam}}{N_{RF}}$ and $N_T^{block} = \frac{N_T^{Beam}}{N_{RF}}$ are the number of receive blocks and transmit block respectively. It is shown in [15] that the optimal solution of

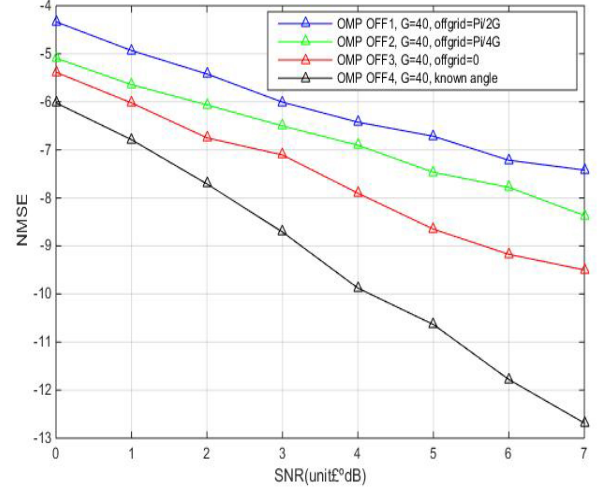


Fig. 3. NMSEs at different SNR levels (dB)

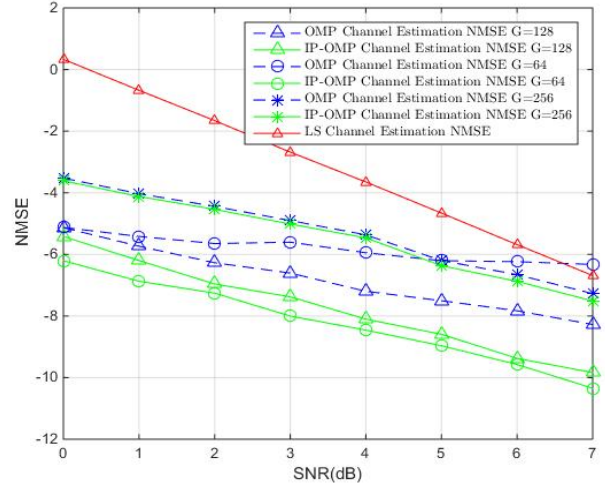


Fig. 4. NMSEs at different SNR levels (dB)

\mathbf{W}_{BB} and \mathbf{F}_{BB} to minimize coherence of sensing matrix are given by (13) and (14).

$$\mathbf{W}_{BB,i} = \mathbf{U}_1 (\boldsymbol{\Lambda}_1^{-1/2})^H, 1 \leq i \leq N_R^{Block}, \quad (13)$$

where \mathbf{U}_1 and $\boldsymbol{\Lambda}_1$ are the matrices of the eigenvectors and eigenvalues, respectively, satisfying $\mathbf{W}_{RF,i}^H \bar{\mathbf{A}}_R \bar{\mathbf{A}}_R^H \mathbf{W}_{RF,i} = \mathbf{U}_1 \boldsymbol{\Lambda}_1 \mathbf{U}_1^H$.

$$\mathbf{F}_{BB,i} = \mathbf{U}_2^* (\boldsymbol{\Lambda}_2^{-1/2})^T, 1 \leq i \leq N_T^{Block}, \quad (14)$$

where \mathbf{U}_2 and $\boldsymbol{\Lambda}_2$ are the matrices of the eigenvectors and eigenvalues, respectively, satisfying $\mathbf{F}_{RF,i}^T \bar{\mathbf{A}}_T^* (\mathbf{F}_{RF,i}^T \bar{\mathbf{A}}_T^*)^H = \mathbf{U}_2 \boldsymbol{\Lambda}_2 \mathbf{U}_2^H$. \mathbf{F}_{BB} and \mathbf{W}_{BB} are calculated as (13) and (14).

Fig. 3 compares the normalized mean square error (NMSE) defined as $10 \log_{10} (\mathbb{E}(\|\mathbf{H} - \mathbf{H}^{LS/CS}\|_F^2 / \|\mathbf{H}\|_F^2))$. Noted, normalization is used because the sparsity of channel makes MSE always extremely small. We consider different AoDs/AoAs

with the same grid size $G = 40$ using conventional OMP algorithm, referred to as OFF1, OFF2, OFF3 for off grid angle $\Delta\theta = (0, \frac{\pi}{4G}, \frac{\pi}{2G})$. We also consider the estimation using known AoDs/AoAs for the purpose of comparison as OFF4. The grid points used in OMP algorithms are uniformly distributed in $[0, \pi)$. We set the true continuous AoD/AoA as $(\theta_{t,\ell}, \theta_{r,\ell})$ with $\ell \in \{0, 1, \dots, L\}$. For OFF1, OFF2, OFF3, $(\theta_{t,\ell}, \theta_{r,\ell})$ take values from set $\{\theta_{r,\ell}, \theta_{t,\ell}\} \in \{0, \frac{\pi}{G-1}, \dots, \frac{\pi(G-1)}{G-1}\}$, $\{\theta_{r,\ell}, \theta_{t,\ell}\} \in \{0 + \frac{\pi}{4G}, \frac{\pi}{G-1} + \frac{\pi}{4G}, \dots, \frac{\pi(G-1)}{G-1} + \frac{\pi}{4G}\}$ and $\{\theta_{r,\ell}, \theta_{t,\ell}\} \in \{0 + \frac{\pi}{2G}, \frac{\pi}{G-1} + \frac{\pi}{2G}, \dots, \frac{\pi(G-1)}{G-1} + \frac{\pi}{2G}\}$ respectively. As shown in Fig. 3, OFF3 is not close to OFF4. Because AoDs/AoAs can not be perfectly estimated by OMP algorithm even without off grid error. OFF1, OFF2 and OFF3 shows that the off grid angles deteriorate channel estimation performance severely. It demonstrates that the channel estimation performance can be improved by enhancing the angle estimation.

In Fig. 4, we consider OMP algorithms and IP-OMP algorithms with different grid size G . For $G = 64, 128, 256$, OMP algorithms are named as OMP1, OMP2 and OMP3 respectively. And IP-OMP algorithms with $G = 64, 128, 256$ are named as the IP-OMP1, IP-OMP2 and IP-OMP3 respectively. G should be large enough to guarantee the sparsity of channel representation. The grid points used in OMP algorithms are uniformly distributed over $[0, \pi)$. We also consider the conventional LS algorithm for comparison. Fig. 4 compares the NMSE of the above algorithms. As shown in Fig. 4, LS method has the worst performance with complexity $O((N_T N_R)^2 N_T^{Beam} N_R^{Beam})$. And all of the OMP based methods with complexity $O(LN_T^{Beam} N_R^{Beam} G^2)$ can achieve better performance compared to the LS method. Among three conventional OMP methods, as expected, the performance is better when G increases from 64 to 128. However, when G grows from 128 to 256, the estimation becomes worse. Because the large grid size induces a higher mutual coherence of sensing matrix which does not satisfy RIP. In CS theory, sensing matrix should satisfy RIP to guarantee recovery performance. So we can not improve estimation performance by further increasing G . In order to achieve a desirable estimation performance, IP-OMP algorithms are employed. Comparing with OMP, IP-OMP algorithm performs better when $G = 64, 128, 256$. Especially, for $G = 64, 128$, the impact of grid error is significantly mitigated and the performances are much better than the corresponding OMP algorithm with the same G . IP-OMP3 improves little because of the great number of G results in limited space to further improve the angle estimation. Because IP-OMP and OMP have the same order of complexity. We use MATLAB to calculate the computational complexity of IP-OMP and OMP for $G = 64, 128, 256$ respectively. If we consider the complexity of OMP $G = 64$ as 1. Then the complexity is 1, 4 and 16 for OMP $G = 64, 128, 256$. The results show that the complexity of IP-OMP is 6, 12 and 24. That is to say IP-OMP with $G = 64$ can achieve much better performance than that of OMP with $G = 128, 256$, at the cost of slightly increased complexity compared with

OMP $G = 128$ and significantly reduced complexity compared with OMP $G = 256$. In summary, IP-OMP algorithm can use a small G value to achieve significant improved estimation performance without causing unaffordable computational load.

VI. CONCLUSION

In this paper, we presented a novel approach for channel estimation in mmWave MIMO communication. To solve the problem in the conventional grid-based OMP, IP method was applied to improve the angle estimation, and thereby improve the channel estimation. The simulation results demonstrated that the IP-OMP can outperform OMP, while requiring an affordable computation, and that the achievable best performance of estimation is much better than that of the OMP with increased grid number. Interesting extensions to this work will be to improve the angle optimization algorithms or to design the CS algorithm without grid.

REFERENCES

- [1] T.S Rappaport et al., "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!", IEEE Access, vol. 1, pp. 335-349, 2013.
- [2] Z. Pi, F. Khan, "An introduction to millimeter-wave mobile broadband systems", IEEE Commun. Mag., vol. 49, no. 6, pp. 101-107, June 2011.
- [3] O. El Ayach et al., "Spatially Sparse Precoding in Millimeter Wave MIMO Systems", IEEE Trans. Wireless Commun., vol. 13, no. 3, pp. 1499-1513, Mar. 2014.
- [4] X. Gao, L. Dai, S. Han, C. L. I, R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays", IEEE Journal on Sel. Areas in Communications, vol. 34, no. 4, pp. 998-1009, April 2016.
- [5] M. Rossi, A. Haimovich, Y. Eldar, "Spatial compressive sensing for MIMO radar", IEEE Trans. Sig. Proc., vol. 62, no. 2, pp. 419-430, Jan. 2014
- [6] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems", IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 436-453, Apr. 2016.
- [7] P. Smulders and L. Correia, Characterisation of propagation in 60 GHz radio channels, Electron. Commun. Eng. J., vol. 9, no. 2, pp. 7380, 1997.
- [8] E. J. Candes and T. Tao, Decoding by linear programming, IEEE Transactions on Information Theory, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
- [9] O. E. Ayach et al., "Low Complexity Precoding for Large Millimeter Wave MIMO Systems", IEEE International Conference on Communications (ICC), pp. 3724-3729, Jun. 2012.
- [10] A. Alkhateeb, O. El Ayach, G. Leus, R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems", IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 831-846, Oct. 2014.
- [11] Shu Sun, Theodore S. Rappaport, "Millimeter Wave MIMO Channel Estimation Based on Adaptive Compressed Sensing", IEEE International Conference on Communications Workshop (ICCW), May 2017.
- [12] J. Lee, G.-T. Gil, Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications", Proc. IEEE Global Commun. Conf. (GLOBECOM), pp. 3326-3331, Dec. 2014.
- [13] J. Tropp, A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", Information Theory IEEE Transactions on, vol. 53, no. 12, pp. 4655-4666, Dec 2007.
- [14] J. Lee, G. T. Gil, Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications", IEEE Trans. Commun., vol. 64, no. 6, pp. 2370-2386, Apr. 2016.
- [15] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," IEEE Trans. Signal Proc., vol. 59, no. 9, pp. 4300-4312, 2011.
- [16] E. J. Candes and M. B. Waken, "An Introduction to Compressive Sampling", IEEE Sig. Proc. Magazine, Mar 2008.