Enhanced Channel Estimation and Codebook Design for Millimeter-Wave Communication

Zhenyu Xiao, *Senior Member, IEEE*, Hang Dong, Lin Bai, *Senior Member, IEEE*, Pengfei Xia, *Senior Member, IEEE*, and Xiang-Gen Xia, *Fellow, IEEE*

***Abstract*—The existing channel estimation methods for millimeter-wave communications, e.g., hierarchical search and compressed sensing, either acquire only one single multipath component (MPC) or require considerably high training overhead. To realize fast yet accurate channel estimation, we propose a multipath decomposition and recovery approach in this paper. The proposed approach has two stages. In the first stage, instead of directly searching the real MPCs, we decompose each real MPC into several *virtual* MPCs and acquire the *virtual* MPCs by using the hierarchical search based on a normal-resolution codebook. Then, in the second stage, the real MPCs are recovered from the acquired virtual MPCs in the first stage, which turns out to be a sparse reconstruction problem, where the size of the dictionary matrix is greatly reduced by exploiting the results of the virtual multipath acquisition. Moreover, to make the proposed approach applicable for both analog and hybrid beamforming/combining devices with strict constant-modulus constraint, we particularly design a codebook for the hierarchical search by using an enhanced subarray technique, and the codebook is also applicable in other hierarchical search methods. Performance comparisons show that the proposed approach achieves a superior tradeoff between estimation performance and training penalty over the state-of-the-art alternatives.**

***Index Terms*—Channel estimation, millimeter-wave (mmWave), compressed sensing, hierarchical search, beamforming.**

Manuscript received December 5, 2017; revised June 2, 2018; accepted June 26, 2018. Date of publication July 9, 2018; date of current version October 15, 2018. This work was supported in part by the National Key Research and Development Program under Grants 2016YFB1200100 and 2017YFB0503002, in part by the National Natural Science Foundation of China under Grants 61571025, 91538204, 71731001, and 61671327, in part by the Open Research Fund of Key Laboratory of Space Utilization, Chinese Academy of Sciences (LSU-DZXX-2017-02), and in part by the Huawei under Grant YBN2017030039. The review of this paper was coordinated by Dr. L. Zhao. *(Corresponding author: Lin Bai.)*

Z. Xiao, H. Dong, and L. Bai are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China, with the Key Laboratory of Advanced technology of Near Space Information System (Beihang University), Ministry of Industry and Information Technology of the People's Republic of China, Beijing 100804, China, and also with the National Engineering Laboratory for Big Data Application Technologies for Comprehensive Traffic, Beijing 100191, China (e-mail: xiaozy@buaa.edu.cn; donghang_0607@buaa.edu.cn; l.bai@buaa.edu.cn).

P. Xia is with the School of Electronics and Information Engineering and the Key Laboratory of Embedded System and Service Computing, Tongji University, Shanghai 200092, China (e-mail: pengfei.xia@gmail.com).

X.-G. Xia is with the College of Information Engineering, Shenzhen University, Shenzhen 518060, China, and also with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: xxia@ee.udel.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2854369

I. Introduction

MILLIMETER-WAVE (mmWave) communication has drawn global attention in the past a few years due to its abundant frequency resource [1]–[8]. To bridge the link budget gap due to the extremely high path loss in the mmWave band, and meanwhile to fulfil the low hardware cost requirement, analog beamforming/combining is usually preferred, where all the antennas share a single radio-frequency (RF) chain and have constant-modulus (CM) constraint on their weights [9]. Meanwhile, a hybrid analog/digital precoding/combining structure has also been proposed to realize multi-stream/multi-user transmission [1]–[3], where a small number of RF chains are tied to a large antenna array. Subject to the beamforming/combining structures, the conventional channel estimation methods developed for classic multiple-input multiple-output (MIMO) communications is rather inefficient for mmWave communications due to high pilot overhead as well as high computational cost [10]. Instead, several approaches have been proposed to enable estimation of the mmWave channel and/or the beamforming vectors. Among them, hierarchical search [9], [11], [12] and compressed sensing (CS) [1], [13], [14] were two typical kinds of methods, which have been extensively exploited to acquire channel state information (CSI) with affordable overhead.

To enable hierarchical search [9], [11], [12], a multi-resolution/hierarchical codebook needs to be defined first, where the beamwidth of a lower-resolution codeword is the summation of those of several higher-resolution codewords. Then a multi-stage search may be conducted based on the codebook to fast acquire a multipath component (MPC) by refining the beamwidth stage by stage. Although the hierarchical search method is time efficient and can achieve a high detection rate to acquire a MPC [11], [12], it is usually limited to acquire only *one single* MPC. Extending it to multiple MPCs is nontrivial due to the limited angle resolution of a normal codebook, since the angle estimation error would result in significant effect on the search of the remaining MPCs. Increasing the angle resolution of the codebook may help to search multiple MPCs, though it not only increases the training overhead, but also requires high-resolution phase shifters, which are expensive and may not be feasible with current mmWave RF technology.

Since mmWave channel is generally sparse in the angle/spatial domain, the CS approach is also an attractive candidate, and many CS based channel estimation methods have been proposed [1], [13]–[15]. Different from the hierarchical search method, the CS based approaches are open-looped, which means that the pilot overhead does not increase in the multi-user case [13]. However, the performance of the CS based schemes is highly dependent on the number of measurements. To achieve a satisfactory estimation performance, the training overhead is in fact considerably high. In [10], an adaptive CS (ACS) method was proposed, where a hierarchical codebook was also designed to reduce the required number of measurements. Although the ACS method can reduce the overhead by iterative training [13], it requires a large number of RF chains, which may make the method less attractive for devices with only a few RF chains. Except these CS methods which make use of the spatial sparsity of mmWave channel [1], [13], [14], there are also alternatives which make use of the time-domain sparsity of mmWave channel. For instance, in [16], the structured compressive sensing (SCS) framework [17], [18] together with training sequence optimization was investigated for mmWave channel estimation, and promising performance was obtained.

In this paper, we aim to design fast yet accurate channel estimation for mmWave communications by exploiting the spacial sparsity, which can be used for both the analog and hybrid beamforming/combining structures, under the constant-modulus (CM) constraint with normal angle resolution. To this purpose, we propose a multipath decomposition and recovery (MDR) approach in this paper. The proposed approach can be divided into two stages. In the first stage, instead of searching the real MPCs, we decompose each real MPC into several *virtual* MPCs, and we exploit the hierarchical search to acquire the *virtual* MPCs based on a normal-resolution codebook. Then in the second stage, the real MPCs are reconstructed from the virtual MPCs acquired in the first stage, which is shown to be a sparse reconstruction problem. By using the results of the virtual multipath acquisition, the size of the dictionary matrix of the sparse reconstruction problem can be greatly reduced, and the problem can be efficiently solved by using the classic orthogonal matching pursuit (OMP) method. Moreover, to make the proposed approach applicable for both analog and hybrid beamforming/combining devices with strict CM constraint, we particularly design a hierarchical codebook for the hierarchical search in the first stage by using an enhanced sub-array technique, and the codebook is also applicable in other hierarchical search methods. It is verified by the simulation results that the proposed approach achieves a superior tradeoff between estimation performance and training overhead over the alternatives. It should be emphasized that one major contribution of this paper is the idea of MDR, which enables the combination of the hierarchical search and CS, and thus fast yet accurate mmWave channel estimation, and the other contribution is the design of the hierarchical codebook under strict CM constraint, which has not been found yet in the literature to the best of our knowledge.

The rest of this paper is organized as follows. The system and channel models are introduced in Section II, where the background of mmWave channel estimation is also presented. The MDR approach is proposed in Section III, and the corresponding codebook design is given in Section IV. Performance evaluation and comparison are conducted in Section V. Lastly, the conclusions are drawn in Section VI.

Symbol Notations: a, $\mathbf{a}$, $\mathbf{A}$, and $\mathcal{A}$ denote a scalar variable, a vector, a matrix, and a set, respectively. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote conjugate, transpose and conjugate transpose, respectively. $|\cdot|$, $\|\cdot\|_0$, $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ denote the absolute value, zero-norm, two-norm and Frobenius-norm, respectively. $\mathbb{E}(\cdot)$ denotes the expectation operation. $[\mathbf{a}]_i$ and $[\mathbf{A}]_{ij}$ denote the i-th entry of $\mathbf{a}$ and the i-row and j-column entry of $\mathbf{A}$, respectively.

II. System Model and Background of mmWave Channel Estimation

A. System Model

We adopt an analog beamforming/combining model, and the received symbol is expressed as

$$y = \sqrt{P}\mathbf{w}_{\mathrm{r}}^{\mathrm{H}}\mathbf{H}\mathbf{w}_{\mathrm{t}}s + z, \tag{1}$$

where s is a transmitted symbol, P is the average transmission power, $\mathbf{w}_{\mathrm{r}}$ and $\mathbf{w}_{\mathrm{t}}$ are Rx/Tx antenna weight vectors (AWVs), respectively, $\mathbf{H}$ is the channel matrix, and z is the Gaussian white noise. Let N_{r} and N_{t} denote the numbers of antennas at Rx and Tx, respectively. Then $\mathbf{w}_{\mathrm{r}}$ and $\mathbf{w}_{\mathrm{t}}$ are $N_{\mathrm{r}} \times 1$ and $N_{\mathrm{t}} \times 1$ vectors, respectively, with constant modulus and unit 2-norm, i.e., $|\mathbf{w}_{\mathrm{r}}| = \mathbf{1}/\sqrt{N_{\mathrm{r}}}$ and $|\mathbf{w}_{\mathrm{t}}| = \mathbf{1}/\sqrt{N_{\mathrm{t}}}$. In the case that a hybrid beamforming/combining structure is adopted at Rx and Tx, $\mathbf{w}_{\mathrm{r}}$ and $\mathbf{w}_{\mathrm{t}}$ will be the product of a digital beamforming/combining vector and an analog precoding/combining matrix with constant modulus [10].

It is known that mmWave channels have limited scattering, and MPCs are mainly generated by reflections [3], [19], [20]. Different MPCs have different angles of departure (AoDs) and angles of arrival (AoAs). Without loss of generality, we adopt the directional mmWave channel model assuming a uniform linear array (ULA) with a half-wavelength antenna space. Then an mmWave channel can be expressed as [10], [12], [21]–[24]

$$\mathbf{H} = \sum_{\ell=1}^{L} \lambda_\ell \mathbf{a}_{\mathrm{r}}(\theta_\ell)\mathbf{a}_{\mathrm{t}}^{\mathrm{H}}(\psi_\ell), \tag{2}$$

where λ_ℓ is the complex coefficient of the ℓ-th path and $\mathbb{E}(\sum_{\ell=1}^{L}|\lambda_\ell|^2) = N_{\mathrm{r}}N_{\mathrm{t}}$, L is the number of MPCs, $\mathbf{a}_{\mathrm{r}}(\cdot)$ and $\mathbf{a}_{\mathrm{t}}(\cdot)$ are Rx/Tx steering vector functions defined as

$$\begin{aligned}\mathbf{a}_{\mathrm{r}}(\theta) &= \frac{1}{\sqrt{N_{\mathrm{r}}}}[e^{j\pi 0\theta}, e^{j\pi 1\theta}, ..., e^{j\pi(N_{\mathrm{r}}-1)\theta}]^{\mathrm{T}},\\ \mathbf{a}_{\mathrm{t}}(\psi) &= \frac{1}{\sqrt{N_{\mathrm{t}}}}[e^{j\pi 0\psi}, e^{j\pi 1\psi}, ..., e^{j\pi(N_{\mathrm{t}}-1)\psi}]^{\mathrm{T}},\end{aligned} \tag{3}$$

which depend on the array geometry, θ_ℓ and ψ_ℓ are cos(AoA) and cos(AoD) of the ℓ-th path, respectively. Let $\bar{\theta}_\ell$ and $\bar{\psi}_\ell$ denote the real AoA and AoD of the ℓ-th path, respectively; then we have $\theta_\ell = \cos(\bar{\theta}_\ell)$ and $\psi_\ell = \cos(\bar{\psi}_\ell)$. Therefore, θ_ℓ and ψ_ℓ are within the range $[-1, 1]$. For convenience and without loss of generality, in the rest of this paper, θ_ℓ and ψ_ℓ are also called AoAs and AoDs, respectively.

B. Background of mmWave Channel Estimation

To estimate the channel, we need to make some measurements based on the signal model shown in (1). In each measurement, Tx sets its AWV (Tx AWV) and transmits a training symbol[1], while Rx sets its AWV (Rx AWV) and receives the training symbol. If the conventional least-square (LS) method is adopted to estimate the channel, we need at least $N_{\mathrm{r}}N_{\mathrm{t}}$ measurements, which is unaffordable in mmWave communications where N_{r} and N_{t} are large in general. To reduce the pilot overhead, there are two main candidate branches: the CS based method and

[1]In practice, a training sequence may be used instead to achieve a higher received signal-to-noise ratio.

the hierarchical search method. We briefly introduce them for comparison with the proposed approach.

1) The Hierarchial Search Method: As there are L MPCs to be found, it is natural to search them one by one, and the hierarchical search method can be used to reduce the overhead. When using the hierarchical search method to search a single MPC [9], [11], [12], a multi-resolution/hierarchical codebook may be predefined, where a coarse sub-codebook may be defined with a small number of coarse sectors (or low-resolution beams) covering the intended angle range, while a fine sub-codebook may be defined with a large number of fine (or high-resolution) beams covering the same intended angle range. A coarse sector may have the same coverage as that of multiple fine beams together. A divide-and-conquer search may then be carried out across the hierarchical codebook, by finding the best sector first on the low-resolution codebook level, and then finding the best beam on the high-resolution codebook level, while the best high-resolution beam is encapsulated in the best sector.

Different from the single-path search, for multi-path search, the contribution of the former acquired MPCs should be subtracted from the received signal during the search of each MPC. For instance, suppose that we have estimated the coefficients, AoAs and AoDs of the first $L_{\rm f}$ MPCs, denoted by $\hat{\lambda}_i$, $\hat{\theta}_i$, $\hat{\psi}_i$ for $i=1,2,...,L_{\rm f}$. Then in the search of the $(L_{\rm f}+1)$-th MPC, in each measurement we can compute the decision variable as

$$\begin{aligned}\bar{y} =& \underbrace{\sqrt{P}\mathbf{w}_{\rm r}^{\rm H}\mathbf{H}\mathbf{w}_{\rm t}+z}_{\rm Measured} - \underbrace{\sqrt{P}\mathbf{w}_{\rm r}^{\rm H}\left(\sum_{i=1}^{L_{\rm f}}\hat{\lambda}_i\mathbf{a}_{\rm r}(\hat{\theta}_i)\mathbf{a}_{\rm t}^{\rm H}(\hat{\psi}_i)\right)\mathbf{w}_{\rm t}}_{\rm Former\ Contribution}\\ =& \sqrt{P}\mathbf{w}_{\rm r}^{\rm H}\left(\sum_{i=L_{\rm f}+1}^{L}\lambda_i\mathbf{a}_{\rm r}(\theta_i)\mathbf{a}_{\rm t}^{\rm H}(\psi_i)\right)\mathbf{w}_{\rm t}+z\\ &+\sqrt{P}\mathbf{w}_{\rm r}^{\rm H}\left(\sum_{i=1}^{L_{\rm f}}\Big(\lambda_i\mathbf{a}_{\rm r}(\theta_i)\mathbf{a}_{\rm t}^{\rm H}(\psi_i)-\hat{\lambda}_i\mathbf{a}_{\rm r}(\hat{\theta}_i)\mathbf{a}_{\rm t}^{\rm H}(\hat{\psi}_i)\Big)\right)\mathbf{w}_{\rm t}\\ \triangleq& \sqrt{P}\mathbf{w}_{\rm r}^{\rm H}\left(\sum_{i=L_{\rm f}+1}^{L}\lambda_i\mathbf{a}_{\rm r}(\theta_i)\mathbf{a}_{\rm t}^{\rm H}(\psi_i)\right)\mathbf{w}_{\rm t}+z+I_{\rm res},\qquad(4)\end{aligned}$$

where $I_{\rm res}$ is the residual interference. If the AoAs and AoDs are accurately estimated, the coefficients will also be accurately estimated, and $I_{\rm res}$ will be small. Dropping the noise component, we have

$$\bar{y}\approx\sqrt{P}\mathbf{w}_{\rm r}^{\rm H}\left(\sum_{i=L_{\rm f}+1}^{L}\lambda_i\mathbf{a}_{\rm r}(\theta_i)\mathbf{a}_{\rm t}^{\rm H}(\psi_i)\right)\mathbf{w}_{\rm t},\qquad(5)$$

which means the $(L_{\rm f}+1)$-th MPC can be normally found by using the hierarchical search method. However, if the AoAs and AoDs are not accurately estimated, there would be significant coefficient errors, and $I_{\rm res}$ will be large. In such a case, the search of the $(L_{\rm f}+1)$-th MPC will be dramatically affected by the residual interference. Note that usually earlier paths have larger gains than later paths. Small error in an earlier path could lead to significant errors for the later path. That is to say, accurate estimation of the AoAs and AoDs are critical for the hierarchical search method.

Usually, the angle resolution of a codebook for a ULA with $N_{\rm A}$ antennas is normally $2/N_{\rm A}$, because a steering vector has an angle resolution of $2/N_{\rm A}$ [11], [12], [25], [26]. Hence, in this paper we call a codebook with a resolution of $2/N_{\rm A}$ a normal-resolution codebook, and one with a resolution better than $2/N_{\rm A}$ a high-resolution codebook. As shown later (see Figs. 5 and 6), the performance of hierarchical search with a normal-resolution codebook is poor. Although it is possible to design a high-resolution so as to accurately estimate the AoAs and AoDs, it will be at the cost of larger training overhead and computational complexity. Furthermore, a finer resolution codebook also calls for finer resolution at each phase shifter, which may not be feasible with current mmWave RF technology.

2) The CS Approach: Based on the signal model in (1), we may make multiple measurements with different Tx AWVs $[\mathbf{w}_{{\rm t}1},\mathbf{w}_{{\rm t}2},...,\mathbf{w}_{{\rm t}k_{\rm t}}]\triangleq\mathbf{W}_{\rm t}$ and Rx AWVs $[\mathbf{w}_{{\rm r}1},\mathbf{w}_{{\rm r}2},...,\mathbf{w}_{{\rm r}k_{\rm r}}]\triangleq\mathbf{W}_{\rm r}$, and we assume, without loss of generality, $s=1$, then we observe the measurements

$$\mathbf{Y}=\sqrt{P}\mathbf{W}_{\rm r}^{\rm H}\mathbf{H}\mathbf{W}_{\rm t}+\mathbf{Z},\qquad(6)$$

where $\mathbf{Z}$ is the noise matrix. By sampling the AoA/AoD domains with sufficiently high resolution δ, we can obtain $\mathbf{A}_{\rm r}=[\mathbf{a}_{\rm r}(-1+\frac{\delta}{2}),\mathbf{a}_{\rm r}(-1+\frac{3\delta}{2}),...]$ and $\mathbf{A}_{\rm t}=[\mathbf{a}_{\rm t}(-1+\frac{\delta}{2}),\mathbf{a}_{\rm t}(-1+\frac{3\delta}{2}),...]$. Then $\mathbf{H}$ can be approximately expressed as $\mathbf{H}=\mathbf{A}_{\rm r}\mathbf{\Sigma}\mathbf{A}_{\rm t}^{\rm H}$, where $\mathbf{\Sigma}$ is a diagonal and sparse matrix with the diagonal entries corresponding to the channel coefficients λ_ℓ. Substituting $\mathbf{H}$ in (6) with this expression and vectorizing $\mathbf{Y}$, we have [13], [14]

$$\begin{aligned}{\rm vec}(\mathbf{Y})&={\rm vec}(\sqrt{P}\mathbf{W}_{\rm r}^{\rm H}\mathbf{A}_{\rm r}\mathbf{\Sigma}\mathbf{A}_{\rm t}^{\rm H}\mathbf{W}_{\rm t}+\mathbf{Z})\\ &=\sqrt{P}\left((\mathbf{A}_{\rm t}^{\rm H}\mathbf{W}_{\rm t})^{\rm T}\otimes(\mathbf{W}_{\rm r}^{\rm H}\mathbf{A}_{\rm r})\right){\rm vec}(\mathbf{\Sigma})+{\rm vec}(\mathbf{Z})\\ &\triangleq\sqrt{P}\mathbf{Q}{\rm vec}(\mathbf{\Sigma})+{\rm vec}(\mathbf{Z}),\qquad(7)\end{aligned}$$

where $\otimes$ is the Kronecker product. Since $\|{\rm vec}(\mathbf{\Sigma})\|_0=L\ll N_{\rm r}N_{\rm t}$, sparse recovery tools can be adopted to estimate $\mathbf{H}$, where the dictionary matrix $\mathbf{Q}$ can be obtained by randomly setting the Tx/Rx training AWVs in each measurement as $[\mathbf{w}_{\rm r}]_k\in\{e^{j\theta}/\sqrt{N_{\rm r}}\}$ and $[\mathbf{w}_{\rm t}]_m\in\{e^{j\theta}/\sqrt{N_{\rm t}}\}$ with uniformly distributed phase θ. Note that as the number of candidate vectors, i.e., the number of columns of $\mathbf{Q}$, is large, the computational complexity of the CS approach is high. In addition, the total number of measurements is $T_{\rm CS}=k_{\rm r}k_{\rm t}$. It was shown (and will also be shown later) that when $T_{\rm CS}$ is not large enough, the performance of the CS approach is not satisfactory[2] [13].

In summary, the hierarchical search method can efficiently estimate multiple MPCs one by one, but the problem is that significant residual error may appear due to the limited angle resolution of a codebook, which degrades the search performance. The CS method may achieve good performance, but only when the number of measurements is large enough. In this paper, we propose a combined method in estimating multiple paths while sticking to a normal resolution codebook.

III. Channel Estimation With Multipath Decomposition and Recovery

The proposed MDR method has two stages, namely virtual multipath acquisition (VMA) based on multipath decomposition for the first stage and multipath recovery for the second stage. We start from the first stage.

[2]The ACS scheme proposed in [10] can reduce the training overhead to some extent, but multiple RF chains are required to guarantee satisfactory performance.

A. The First Stage: Virtual Multipath Acquisition

1) Decomposition of a Real MPC: First, we need to decompose a real MPC into several virtual MPC. In particular, we approximate a real MPC to the summation of 4 virtual MPC. Without loss of generality, we will take the ℓ-th MPC in the model (2) for example to show the approximation process, and the corresponding response can be expressed as

$$\mathbf{H}_\ell = \lambda_\ell \mathbf{a}_{\mathrm{r}}(\theta_\ell)\mathbf{a}_{\mathrm{t}}^{\mathrm{H}}(\psi_\ell). \tag{8}$$

Before the approximation process, we show a property of the Tx/Rx steering vectors. We sample the AoA and AoD domains with angle resolutions $2/N_{\mathrm{r}}$ and $2/N_{\mathrm{t}}$, respectively, and obtain two sets of steering vectors:

$$\mathbf{U} = \left[\mathbf{a}_{\mathrm{r}}\left(-1+\frac{1}{N_{\mathrm{r}}}\right), \mathbf{a}_{\mathrm{r}}\left(-1+\frac{3}{N_{\mathrm{r}}}\right), \ldots \mathbf{a}_{\mathrm{r}}\left(-1+\frac{2N_{\mathrm{r}}-1}{N_{\mathrm{r}}}\right)\right],$$
$$\mathbf{V} = \left[\mathbf{a}_{\mathrm{t}}\left(-1+\frac{1}{N_{\mathrm{t}}}\right), \mathbf{a}_{\mathrm{t}}\left(-1+\frac{3}{N_{\mathrm{t}}}\right), \ldots \mathbf{a}_{\mathrm{t}}\left(-1+\frac{2N_{\mathrm{t}}-1}{N_{\mathrm{t}}}\right)\right]. \tag{9}$$

It is easy to verify that $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}_{N_{\mathrm{r}}}$ and $\mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{I}_{N_{\mathrm{t}}}$, which means that $\{\mathbf{a}_{\mathrm{r}}(-1+\frac{2k-1}{N_{\mathrm{r}}})\}_{k=1,2,\ldots,N_{\mathrm{r}}}$ and $\{\mathbf{a}_{\mathrm{t}}(-1+\frac{2k-1}{N_{\mathrm{t}}})\}_{k=1,2,\ldots,N_{\mathrm{t}}}$ constitute orthogonal bases of $\mathbb{C}^{N_{\mathrm{r}}}$ and $\mathbb{C}^{N_{\mathrm{t}}}$, respectively. Hence, for arbitrary θ_ℓ and ψ_ℓ, we can express $\mathbf{a}_{\mathrm{r}}(\theta_\ell)$ and $\mathbf{a}_{\mathrm{t}}(\psi_\ell)$ as linear combinations of the two bases, respectively, i.e.,

$$\mathbf{a}_{\mathrm{r}}(\theta_\ell) = \sum_{k=1}^{N_{\mathrm{r}}} \alpha_{k,\ell}\mathbf{a}_{\mathrm{r}}\left(-1+\frac{2k-1}{N_{\mathrm{r}}}\right),$$
$$\mathbf{a}_{\mathrm{t}}(\psi_\ell) = \sum_{k=1}^{N_{\mathrm{t}}} \beta_{k,\ell}\mathbf{a}_{\mathrm{t}}\left(-1+\frac{2k-1}{N_{\mathrm{t}}}\right), \tag{10}$$

where the coefficients $\alpha_{k,\ell}$ and $\beta_{k,\ell}$ are the projections of $\mathbf{a}_{\mathrm{r}}(\theta_\ell)$ on $\mathbf{a}_{\mathrm{r}}(-1+\frac{2k-1}{N_{\mathrm{r}}})$ and $\mathbf{a}_{\mathrm{t}}(\psi_\ell)$ on $\mathbf{a}_{\mathrm{t}}(-1+\frac{2k-1}{N_{\mathrm{t}}})$, respectively, and are computed as

$$\alpha_{k,\ell} = \mathbf{a}_{\mathrm{r}}^{\mathrm{H}}\left(-1+\frac{2k-1}{N_{\mathrm{r}}}\right)\mathbf{a}_{\mathrm{r}}(\theta_\ell),$$
$$\beta_{k,\ell} = \mathbf{a}_{\mathrm{t}}^{\mathrm{H}}\left(-1+\frac{2k-1}{N_{\mathrm{t}}}\right)\mathbf{a}_{\mathrm{t}}(\psi_\ell). \tag{11}$$

Next, we will show that although $\mathbf{a}_{\mathrm{r}}(\theta_\ell)$ can be expressed as a linear combination of N_{r} steering vectors, there are only two steering vectors, whose steering angles are the closest to θ_ℓ, carrying significant coefficients; the others carry very small coefficients and can be neglected in general. To demonstrate this, we need to use the property of Fejér kernel function. Let $f_{\mathrm{r}}(\varrho, x) = |\mathbf{a}_{\mathrm{r}}^{\mathrm{H}}(\varrho)\mathbf{a}_{\mathrm{r}}(\varrho + x)|$, i.e., a Fejér kernel function. It is easy to verify that $f_{\mathrm{r}}(\varrho, x)$ does not depend on ϱ [25]. That is why we adopt the cosine AoA/AoD domain to decompose the multipath instead of the real AoA/AoD domain. Besides, $f_{\mathrm{r}}(x) \triangleq f_{\mathrm{r}}(\varrho, x)$ goes to zero quickly by increasing $|x|$ [25]. It is known that $f_{\mathrm{r}}(0) = 1$, and when $|x| > 1/N_{\mathrm{r}}$, the kernel is far smaller than 1 given that N_{r} is large. According to this property, there are only 2 elements in $\{\alpha_{k,\ell}\}_{k=1,2,\ldots,N_{\mathrm{r}}}$ that may have a significant absolute value. The AoA indices (i.e., k) of the two elements can be obtained by finding k to minimize $|-1+\frac{2k-1}{N_{\mathrm{r}}} - \theta_\ell|$:

$$I_\ell^+ = \lceil (N_{\mathrm{r}}(\theta_\ell + 1) + 1)/2 \rceil,$$
$$I_\ell^- = \lfloor (N_{\mathrm{r}}(\theta_\ell + 1) + 1)/2 \rfloor, \tag{12}$$

where $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ denote ceil and floor integer operations, respectively. Note that the AoA indices should be positive integers no larger than N_{r}. Hence, necessary modulus operation is required, i.e., I_ℓ^+ is reset to 1 if $I_\ell^+ = N_{\mathrm{r}} + 1$ while I_ℓ^- is reset to N_{r} if $I_\ell^- = 0$. In brief, $\mathbf{a}_{\mathrm{r}}(\theta_\ell)$ can be approximated as

$$\begin{aligned}\mathbf{a}_{\mathrm{r}}(\theta_\ell) &\approx \alpha_{I_\ell^+,\ell}\mathbf{a}_{\mathrm{r}}\left(-1+\frac{2I_\ell^+ - 1}{N_{\mathrm{r}}}\right) \\ &\quad + \alpha_{I_\ell^-,\ell}\mathbf{a}_{\mathrm{r}}\left(-1+\frac{2I_\ell^- - 1}{N_{\mathrm{r}}}\right) \\ &\triangleq \alpha_{I_\ell^+,\ell}\mathbf{a}_{\mathrm{r}}(\theta_{\ell+}) + \alpha_{I_\ell^-,\ell}\mathbf{a}_{\mathrm{r}}(\theta_{\ell-}).\end{aligned} \tag{13}$$

Analogously, $\mathbf{a}_{\mathrm{t}}(\psi_\ell)$ can be approximated as

$$\begin{aligned}\mathbf{a}_{\mathrm{t}}(\psi_\ell) &\approx \beta_{J_\ell^+,\ell}\mathbf{a}_{\mathrm{t}}\left(-1+\frac{2J_\ell^+ - 1}{N_{\mathrm{t}}}\right) \\ &\quad + \beta_{J_\ell^-,\ell}\mathbf{a}_{\mathrm{t}}\left(-1+\frac{2J_\ell^- - 1}{N_{\mathrm{t}}}\right) \\ &\triangleq \beta_{J_\ell^+,\ell}\mathbf{a}_{\mathrm{t}}(\psi_{\ell+}) + \beta_{J_\ell^-,\ell}\mathbf{a}_{\mathrm{t}}(\psi_{\ell-}).\end{aligned} \tag{14}$$

where

$$J_\ell^+ = \lceil (N_{\mathrm{t}}(\psi_\ell + 1) + 1)/2 \rceil,$$
$$J_\ell^- = \lfloor (N_{\mathrm{t}}(\psi_\ell + 1) + 1)/2 \rfloor. \tag{15}$$

Consequently, we have

$$\begin{aligned}\mathbf{H}_\ell &\approx \lambda_\ell\left[\alpha_{I_\ell^+,\ell}\mathbf{a}_{\mathrm{r}}(\theta_{\ell+}) + \alpha_{I_\ell^-,\ell}\mathbf{a}_{\mathrm{r}}(\theta_{\ell-})\right] \\ &\quad \times \left[\beta_{J_\ell^+,\ell}\mathbf{a}_{\mathrm{t}}(\psi_{\ell+}) + \beta_{J_\ell^-,\ell}\mathbf{a}_{\mathrm{t}}(\psi_{\ell-})\right]^{\mathrm{H}}.\end{aligned} \tag{16}$$

The full channel matrix $\mathbf{H}$ can also be approximated by

$$\begin{aligned}\mathbf{H} &\approx \sum_{\ell=1}^{L}\lambda_\ell\left[\alpha_{I_\ell^+,\ell}\mathbf{a}_{\mathrm{r}}(\theta_{\ell+}) + \alpha_{I_\ell^-,\ell}\mathbf{a}_{\mathrm{r}}(\theta_{\ell-})\right] \\ &\quad \times \left[\beta_{J_\ell^+,\ell}\mathbf{a}_{\mathrm{t}}(\psi_{\ell+}) + \beta_{J_\ell^-,\ell}\mathbf{a}_{\mathrm{t}}(\psi_{\ell-})\right]^{\mathrm{H}}.\end{aligned} \tag{17}$$

As we can see each real MPC is approximated by the summation of 4 *virtual* MPC. Moreover, it is emphasized that the Rx/Tx steering vectors of the 4 virtual MPCs are two *adjacent* basis vectors within $\mathbf{U}$ and $\mathbf{V}$, respectively. Fig. 1 illustrates these properties of an MPC with an AoA of 0 (in the cosine angle domain, which means the real AoA is $\pi/2$ or $3\pi/2$) and an AoD of 1/32, where $N_{\mathrm{t}} = N_{\mathrm{r}} = 16$. The steering angles of the orthogonal basis at the Rx side is $-15/16, -13/16, \ldots, -1/16, 1/16, \ldots, 15/16$; thus the two closest steering angles to the AoA are $I_\ell^+ = +1/N_{\mathrm{r}}$ and $I_\ell^- = -1/N_{\mathrm{r}}$. Similarly, the two closest steering angles to the AoD are $J_\ell^+ = +1/N_{\mathrm{t}}$ and $J_\ell^- = -1/N_{\mathrm{t}}$. The absolute values of α_{1-}, α_{1+}, β_{1-} and β_{1+} are also shown in the figure.

2) Hierarchical Search of the Virtual MPCs: Note that the real MPCs have arbitrary AoAs and AoDs, which means that we may not make an accurate estimation of them with a resolution-limited codebook. Different from the real MPCs, the virtual

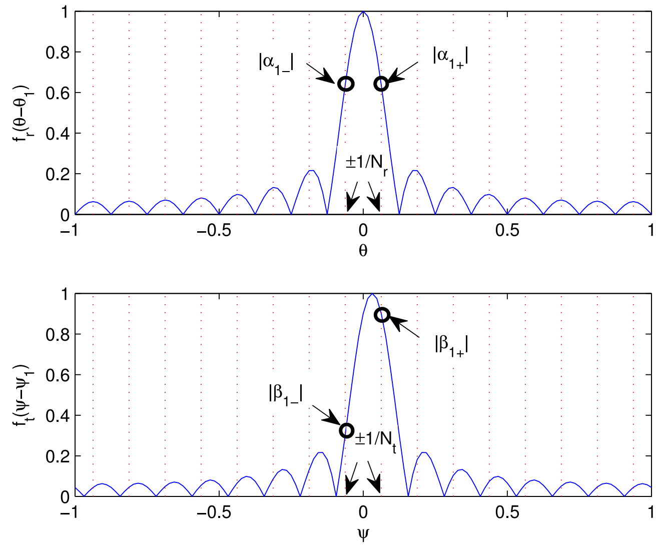


Fig. 1. Illustration of the Fejér kernel functions for an MPC with an AoA of 0 and an AoD of 1/32, where $N_{\rm t} = N_{\rm r} = 16$.

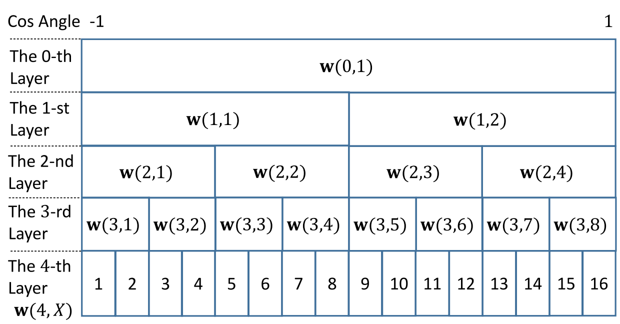


Fig. 2. Beam coverage illustration of a hierarchical codebook with 16 antennas.

MPCs have *fixed discrete* AoAs and AoDs, which means that they can be accurately estimated with a normal-resolution codebook containing the AWVs in $\mathbf{U}$ and $\mathbf{V}$. Thus, instead of directly estimating the L real MPCs, we may estimate the $4L$ virtual MPCs based on a normal-resolution codebook. Afterwards, we can reconstruct the original L real MPCs based on the $4L$ virtual MPCs. Additionally, although the number of virtual MPCs is $4L$, 4 times of that of the real MPCs, the 4 virtual MPCs corresponding to a real MPC are adjacent to each other, which means that once we acquire anyone of them, we can directly acquire the others without launching a new hierarchical search. Thus, the virtual MPCs can be also efficiently acquired by using the hierarchical search method, which will be introduced in detail.

Since the AoAs and AoDs of the virtual MPCs are within $\{-1+\frac{2k-1}{N_{\rm r}}\}_{k=1,2,...,N_{\rm r}}$ and $\{-1+\frac{2k-1}{N_{\rm t}}\}_{k=1,2,...,N_{\rm t}}$, respectively, the virtual MPCs can be accurately estimated by hierarchical search with a normal-resolution codebook. For instance, Fig. 2 shows a typical normal-resolution codebook $\mathcal{F}$ with $N=16$ antennas. The codebook has $(\log_2 N+1)$ layers. In the k-th layer, $k=0,1,2,...,\log_2 N$, there are 2^k codewords of the same beam width with different steering angles and collectively covering the entire search space in the cosine angle domain. Note that a codeword is actually an AWV used for beamforming. Let $\mathbf{w}(k,n)$ denote the n-th codeword in the k-th layer,

Algorithm 1: Hierarchical Search of the Virtual MPCs.

1) Initialization:
$S=\log_2 N_{\rm r}$. /*Assume $N_{\rm r}=N_{\rm t}$*/.
$S_0=2$. /*The initial layer index.*/
$\mathbf{H}_{\rm fd}=\mathbf{0}$. /*The already found MPCs.*/

2) Iteration:
for $\ell=1:L$ **do**
  /*Search for the initial Tx/Rx codewords.*/
  for $m=1:2^{S_0}$ **do**
    for $n=1:2^{S_0}$ **do**
      $y(m,n)=\sqrt{P}\mathbf{w}_{\rm r}(S_0,n)^{\rm H}\mathbf{H}\mathbf{w}_{\rm t}(S_0,m)+z-\mathbf{w}_{\rm r}(S_0,n)^{\rm H}\mathbf{H}_{\rm fd}\mathbf{w}_{\rm t}(S_0,m)$
  $(m_{\rm t}\ n_{\rm r})=\arg\max_{(m,n)}|y(m,n)|$
  /*Hierarchical search.*/
  for $s=(S_0+1):S$ **do**
    for $m=1,2$ **do**
      for $n=1,2$ **do**
        $y(m,n)=\sqrt{P}\mathbf{w}_{\rm r}(s,2(n_{\rm r}-1)+n)^{\rm H}\mathbf{H}\mathbf{w}_{\rm t}(s,2(m_{\rm t}-1)+m)+z-\mathbf{w}_{\rm r}(s,2(n_{\rm r}-1)+n)^{\rm H}\mathbf{H}_{\rm fd}\mathbf{w}_{\rm t}(s,2(m_{\rm t}-1)+m)$
    $(a\ b)=\arg\max_{(m,n)}|y(m,n)|$
    $m_{\rm t}=2(m_{\rm t}-1)+a;\ n_{\rm r}=2(n_{\rm r}-1)+b;$
  /*Collection of the virtual MPCs.*/
  for $m=-1,0,1$ **do**
    for $n=-1,0,1$ **do**
      $y=\sqrt{P}\mathbf{w}_{\rm r}(S,n_{\rm r}+n)^{\rm H}\mathbf{H}\mathbf{w}_{\rm t}(S,m_{\rm t}+m)+z$
      $\mathbf{H}_{\rm fd}=\mathbf{H}_{\rm fd}+y\mathbf{w}_{\rm r}(S,n_{\rm r}+n)\mathbf{w}_{\rm t}(S,m_{\rm t}+m)^{\rm H}$

3) Results:
Return $\mathbf{H}_{\rm fd}$

$n=1,...,2^k$. Then the beam coverage of $\mathbf{w}(k,n)$ is approximately the union of the beam coverage of the 2 codewords in the $(k+1)$-th layer, i.e., $\{\mathbf{w}(k+1,2(n-1)+m)\}_{m=1,2}$. There are different methods to design such a codebook [10]–[12], and all the hierarchical codebooks designed by different methods can be used to search the virtual MPCs, provided that the last-layer codewords are steering vectors towards $\{-1+\frac{2k-1}{N}\}_{k=1,2,...,N}$.

Based on the hierarchical codebook, we next introduce the proposed hierarchical search algorithm to acquire the virtual MPCs, which is shown in Algorithm 1, where $\mathbf{w}_{\rm t}(\cdot)$ and $\mathbf{w}_{\rm r}(\cdot)$ represent the Tx and Rx codewords, respectively. Thus, the virtual MPCs can also be acquired efficiently. There are L iterations in the search process, and the virtual MPCs corresponding to a single MPC are acquired in each iteration. The algorithm is briefly illustrated as follows.

- *Search for the initial Tx/Rx codewords:* As in mmWave communication the transmission power is generally limited, the beamforming training may not start from the 0-th layer, where the codeword is omni-directional and the gain is low. Instead, the beamforming training may need to start from a higher layer, e.g., the S_0-th layer, to provide sufficient start-up beamforming gain. In this process, there are 2^{S_0} candidate codewords at both Tx and Rx. Thus, an exhaustive search over all BS/MS codeword pairs is adopted to search the best Tx/Rx codeword pair, which are treated as the parent codewords for the following search.

- *Hierarchical search:* In this process, a layered search is performed to refine the beam width step by step, until the most significant virtual MPC is acquired at the last layer (the S-th layer).
- *Collection of the virtual MPCs:* After the hierarchical search, the most significant virtual MPC is acquired. Since the other virtual MPCs are adjacent to the already acquired virtual MPC, we collect all the virtual MPCs with adjacent AoAs/AoDs to the acquired virtual MPC. With this operation, 9 instead of 4 virtual MPCs are actually acquired. However, the other 5 virtual MPCs have much smaller strength compared with the 4 desired virtual MPCs, and thus affect little on the results.

Note that in Algorithm 1 we implicitly assume that there exists a feedback channel, through which the index of the best Tx codeword (say, m_t) is transmitted back to the Tx. Since only the indices need to be sent back, the required rate of the feedback channel is small. In the case of time division duplex (TDD) where channel reciprocity is available, the hierarchical search method in [11] can be used instead, where the two nodes transmit training symbols alternatively, such that feedback is not needed. Additionally, when there are multiple users in a mmWave cellular scenario, if spatial-division multiple access (SDMA) is used, a hybrid beamforming structure with multiple RF chains [10] should be equipped at the BS to support multiple users at the same time slot.

B. The Second Stage: Multipath Recovery

As we have estimated the virtual channel $\mathbf{H}_{\rm fd}$, we have the following relation:

$$\mathbf{H} = \sum_{\ell=1}^{L} \lambda_\ell \mathbf{a}_{\rm r}(\theta_\ell)\mathbf{a}_{\rm t}^{\rm H}(\psi_\ell) \approx \mathbf{H}_{\rm fd}. \tag{18}$$

To reconstruct the original channel $\mathbf{H}$, we need to estimate λ_ℓ, θ_ℓ and ψ_ℓ. Hence, we formulate the following problem:

$$\underset{\lambda_\ell,\theta_\ell,\psi_\ell}{\text{minimize}} \;\; \|\mathbf{H}_{\rm fd} - \sum_{\ell=1}^{L} \lambda_\ell \mathbf{a}_{\rm r}(\theta_\ell)\mathbf{a}_{\rm t}^{\rm H}(\psi_\ell)\|_{\rm F}. \tag{19}$$

Then, analogous to the pure CS approach, we could sample the AoA and AoD domain with a high resolution, i.e., an angle interval $2/(KN_{\rm r})$ at the Rx and $2/(KN_{\rm t})$ at the Tx, where K is the over-sampling factor, and we could obtain $\mathbf{A}_{\rm r} = [\mathbf{a}_{\rm r}(-1+\frac{1}{KN_{\rm r}}), \mathbf{a}_{\rm r}(-1+\frac{3}{KN_{\rm r}}), ..., \mathbf{a}_{\rm r}(-1+\frac{2KN_{\rm r}-1}{KN_{\rm r}})]$ and $\mathbf{A}_{\rm t} = [\mathbf{a}_{\rm t}(-1+\frac{1}{KN_{\rm t}}), \mathbf{a}_{\rm t}(-1+\frac{3}{KN_{\rm t}}), ..., \mathbf{a}_{\rm t}(-1+\frac{2KN_{\rm t}-1}{KN_{\rm t}})]$. This manipulation is applicable, but at the cost of a high computational complexity. In fact, by exploiting the search results in Algorithm 1, we can significantly reduce the number of the Rx and Tx candidate AWVs. Concretely, since the ℓ-th estimated AoA of the first stage is $\hat{\theta}_\ell = -1 + \frac{2n_{{\rm r}\ell}-1}{N_{\rm r}}$, the uncertainty range of the ℓ-th AoA should be $[\hat{\theta}_\ell - \frac{2}{N_{\rm r}}, \hat{\theta}_\ell + \frac{2}{N_{\rm r}}]$, which means that the candidate AoAs are the angle set obtained by sampling the angle range $[\hat{\theta}_\ell - \frac{2}{N_{\rm r}}, \hat{\theta}_\ell + \frac{2}{N_{\rm r}}]$ with an interval $2/(KN_{\rm r})$. Consequently, the reduced candidate Rx and Tx candidate AWVs are

$$\begin{aligned}\bar{\mathbf{A}}_{\rm r} = \Bigg[&\left[\mathbf{a}_{\rm r}\left(\hat{\theta}_1 - \frac{2}{N_{\rm r}} + \frac{2k}{KN_{\rm r}}\right)\right]_{k=0,1,\ldots,2K},\\ &\left[\mathbf{a}_{\rm r}\left(\hat{\theta}_2 - \frac{2}{N_{\rm r}} + \frac{2k}{KN_{\rm r}}\right)\right]_{k=0,1,\ldots,2K}, \ldots\\ &\left[\mathbf{a}_{\rm r}\left(\hat{\theta}_L - \frac{2}{N_{\rm r}} + \frac{2k}{KN_{\rm r}}\right)\right]_{k=0,1,\ldots,2K}\Bigg],\end{aligned} \tag{20}$$

and

$$\begin{aligned}\bar{\mathbf{A}}_{\rm t} = \Bigg[&\left[\mathbf{a}_{\rm t}\left(\hat{\psi}_1 - \frac{2}{N_{\rm t}} + \frac{2k}{KN_{\rm t}}\right)\right]_{k=0,1,\ldots,2K},\\ &\left[\mathbf{a}_{\rm t}\left(\hat{\psi}_2 - \frac{2}{N_{\rm t}} + \frac{2k}{KN_{\rm t}}\right)\right]_{k=0,1,\ldots,2K}, \ldots\\ &\left[\mathbf{a}_{\rm t}\left(\hat{\psi}_L - \frac{2}{N_{\rm t}} + \frac{2k}{KN_{\rm t}}\right)\right]_{k=0,1,\ldots,2K}\Bigg],\end{aligned} \tag{21}$$

respectively, where $\hat{\psi}_\ell = -1 + \frac{2m_{{\rm t}\ell}-1}{N_{\rm t}}$. Then $\mathbf{H}$ can be approximately expressed as $\mathbf{H} = \bar{\mathbf{A}}_{\rm r}\mathbf{\Sigma}\bar{\mathbf{A}}_{\rm t}^{\rm H}$, where $\mathbf{\Sigma}$ is a diagonal and sparse matrix with the diagonal entries corresponding to the channel coefficients λ_ℓ, i.e., $\|{\rm vec}(\mathbf{\Sigma})\|_0 = L$. In a sequel,

$$\begin{aligned}&\|\mathbf{H}_{\rm fd} - \sum_{\ell=1}^{L} \lambda_\ell \mathbf{a}_{\rm r}(\theta_\ell)\mathbf{a}_{\rm t}(\psi_\ell)^{\rm H}\|_{\rm F}\\ &= \|{\rm vec}\left(\mathbf{H}_{\rm fd}\right) - \left(\bar{\mathbf{A}}_{\rm t}^* \otimes \bar{\mathbf{A}}_{\rm r}\right){\rm vec}\left(\Sigma\right)\|_2\\ &\triangleq \|{\rm vec}\left(\mathbf{H}_{\rm fd}\right) - \bar{\mathbf{Q}}{\rm vec}\left(\Sigma\right)\|_2\end{aligned} \tag{22}$$

Hence, the problem (28) becomes

$$\begin{aligned}&\underset{\lambda_\ell,\theta_\ell,\psi_\ell}{\text{minimize}} \;\; \|{\rm vec}\left(\mathbf{H}_{\rm fd}\right) - \bar{\mathbf{Q}}{\rm vec}\left(\Sigma\right)\|_2\\ &\text{subject to} \;\; \|{\rm vec}(\mathbf{\Sigma})\|_0 = L\end{aligned} \tag{23}$$

which is a standard sparse reconstruction problem and can be effectively solved by exploiting the OMP algorithm [14]. Note that an intrinsic difference between the problem shown in (23) and the one shown in (7) is that $\mathbf{Y}$ in (7) is measured by using random Tx/Rx AWVs, while $\mathbf{H}_{\rm fd}$ in (23) is obtained by using Algorithm 1 based on a hierarchical codebook.

In practice, the number of MPCs (i.e., L) is not known a priori. Besides, in some cases, it is not necessary to estimate all of the MPCs. In such cases, the number of MPCs in the proposed MDR approach, in both of the two stages, is set to $L = L_{\rm d}$, the desired number of MPCs. For instance, if we want to realize a 2-stream transmission, we only need to estimate $L_{\rm d} = 2$ MPCs, no matter how many MPCs the channel really has.

Since the sparse reconstruction stage does not need measurement, and the required feedback rate is small, the total number of measurements of the proposed MDR method is

$$T_{\rm MDR} = L(4^{S_0} + 2(\log_2(N_{\rm r}) + \log_2(N_{\rm t}) - 2S_0) + 9). \tag{24}$$

Note that this is the training overhead for an analog beamforming/combining structure. In the case of a hybrid structure, where parallel transmission of multiple-stream training sequences are available, the overhead will be further reduced.

It is noteworthy that compared with the existing methods for instantaneous mmWave channel estimation, MDR requires lower training overhead, as will be shown in Section V. However, the training overheads of these methods depend on the number of desired MPCs (L). When L is small, MDR appears efficient. For instance, when $L=1, S_0=1$, and $N_{\mathrm{t}}=N_{\mathrm{r}}=16$, $T_{\mathrm{MDR}}=4\log_2(16)+9=25$, which is even smaller than the required number of measurements in [27] (about 50 measurements) for spatial covariance estimation of a mmWave channel. However, when L is not small enough, e.g., when $L=5$, $T_{\mathrm{MDR}}=125$, which is much greater than the overheads of the spatial covariance estimation methods in [27]. This means that in the case of fast-varying channel, MDR can only support to estimate a small number of MPCs. When the desired number of MPCs is large, the spatial covariance estimation methods proposed in [27] can be used instead.

IV. Hierarchical Codebook Design

As we can see, codebook design is crucial for the proposed approach. Although different codebooks with a hierarchical structure shown in Fig. 2 are all applicable for the proposed approach, the existing codebooks in the literature have different limitations. For instance, the codebooks designed in [10] and [26] require multiple RF chains; they are not suitable for analog beamforming/combining devices, because wide beams cannot be well shaped. In addition, although the codebooks designed in [12] and [11] can be used for analog beamforming/combining devices[3], part of the antennas need to be switched off for some codewords, i.e., a strict CM constraint is not satisfied. This not only reduces the maximal transmission power, but also requires an analog switch for each antenna element path, leading to additional cost and power consumption [28]. In this section, we propose an enhanced sub-array scheme to design a hierarchical codebook as shown in Fig. 2, with a single RF chain under a strict CM constraint. The designed codebook can be used for both analog and hybrid beamforming/combining devices.

A. Preliminaries of Codebook Design

Let $A(\mathbf{w},\Omega)$ denote the beam gain of an arbitrary codeword $\mathbf{w}$ along angle Ω, which is defined as

$$A(\mathbf{w},\Omega)=\sqrt{N}\mathbf{a}(N,\Omega)^{\mathrm{H}}\mathbf{w}=\sum_{n=1}^{N}[\mathbf{w}]_n e^{-\mathrm{j}\pi(n-1)\Omega}, \quad (25)$$

where N is the number of elements of $\mathbf{w}$, and the steering vector function $\mathbf{a}(\cdot)$ is defined as

$$\mathbf{a}(N,\Omega)=\frac{1}{\sqrt{N}}[e^{j\pi 0\Omega},\, e^{j\pi 1\Omega},...,e^{j\pi(N-1)\Omega}]^{\mathrm{T}}. \quad (26)$$

Let $\mathcal{CV}(\mathbf{w})$ denote the beam coverage of an AWV $\mathbf{w}$ in the cosine angle domain. According to Fig. 2, we have

$$\mathcal{CV}(\mathbf{w}(k,n))=\left[-1+\frac{2n-2}{2^k},-1+\frac{2n}{2^k}\right],$$
$$k=0,1,2,...,\log_2 N,\ n=1,2,3,...,2^k. \quad (27)$$

The task of codebook design is to design all the codewords $\mathbf{w}(k,n)$ satisfying the beam coverage in (27) under the CM constraint. The key is to design codewords with relatively wide beams. Suppose the desired beam coverage is $[\Omega_l,\,\Omega_u]$; then a good codeword should satisfy that the beam gain along the angles within the coverage is as high as possible, while that along the angles out of the coverage is as low as possible. Hence, an optimization problem can be formulated to design a codeword:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}}\ \ & \varepsilon \\ \text{subject to}\ \ & |A(\mathbf{w},\Omega)|>1,\ \Omega\in[\Omega_l,\,\Omega_u], \\ & |A(\mathbf{w},\Omega)|<\varepsilon,\ \Omega\notin[\Omega_l,\,\Omega_u], \\ & |[\mathbf{w}]_1|=|[\mathbf{w}]_2|=...=|[\mathbf{w}]_N|. \end{aligned} \quad (28)$$

As there are numerous constraints for continuous Ω, one can sample the angle domain $[-1,\,1]$ to obtain discrete Ω, such that an optimization problem with limited number of constraints is formulated. However, since most of the constraints are non-convex, the problem is difficult to solve. An optimal solution of this problem can be hardly found even with the exhaustive search method, because the size of $\mathbf{w}$, i.e., N, is too large in general. For instance, if we search over the possible phases of all the weight elements with a step of $\pi/18$, we need $(2\pi/(\pi/18))^N=36^N$ tests, which is prohibitively high even for off-line computation. In such a case, sub-optimal heuristic methods are usually adopted to design appropriate codewords [10]–[12], [26]. In this paper, we propose an improved heuristic method for codebook design under strict CM constraint.

We first define the beam width of a steering vector with N antennas:

$$\mathcal{CV}(\mathbf{a}(N,\Omega))=\left[\Omega-\frac{1}{N},\Omega+\frac{1}{N}\right], \quad (29)$$

which means that the steering vectors have a beam width $2/N$ centering at the steering angle [25].

Besides, we present the following beam rotation lemma for codebook design. The proof is not difficult and can be found in [11].

Lemma 1: Given the first codeword in the k-th layer $\mathbf{w}(k,1)$, all the other codewords in the k-th layer can be found through rotating $\mathbf{w}(k,1)$ by $\frac{2n-2}{2^k}$, $n=2,3,...,2^k$, respectively, i.e., $\mathbf{w}(k,n)=\mathbf{w}(k,1)\circ\sqrt{N}\mathbf{a}(N,\frac{2n-2}{2^k})$, where $\circ$ is the entry-wise product.

B. The Enhanced Sub-Array Scheme

Based on the preliminaries, we propose an enhanced sub-array scheme to design the codeword $\mathbf{w}(k,1)$ with beam coverage $[-1,-1+\frac{2}{2^k}]$ in this subsection. The other codewords in the k-th layer, i.e., $\{\mathbf{w}(k,n)\,|\,n=2,3,...,2^k\}$ can be obtained by using Lemma 1 with $\mathbf{w}(k,1)$. The idea of the scheme is to divide the large array into several virtual sub-arrays, and let the sub-arrays steer to evenly-spaced angles within the beam coverage. A key difference of the enhanced sub-array scheme from the joint sub-array and deactivation method in [11] is that beam overlap is allowed in the enhanced scheme, while in [11] the steering angles must be sufficiently spaced. Beam overlap increases the mutual influence between adjacent sub-arrays, and thus calls for more delicate weight setting and optimization. In the following, we will introduce the enhanced sub-array scheme in detail.

[3]Codebooks that can be used for analog beamforming/combining devices can be surely used for hybrid analog beamforming/combining devices.

1) The Number of Sub-Arrays: Firstly, we decide the number of sub-arrays that we need to use. Let an N-element antenna array be divided into S sub-arrays. Then each sub-array has N/S antennas, and the beam width of each sub-array is $2/(N/S) = 2S/N$. If the steering directions of the S sub-arrays are spaced by $2S/N$ in the cosine angle domain, the total beam width of the sub-arrays is $2S/N * S = 2S^2/N$ [11], which is S^2 times of the beam width of a steering vector with N antennas. Hence, the broadening factor of the sub-array technique is S^2.

According to (27), the targeted beam width of the k-th layer codewords is $2/2^k = 2^{1-k}$. Hence, the number of sub-arrays for the k-th layer codewords satisfies $S_k = \sqrt{\frac{2^{1-k}}{2/N}} = \sqrt{2^{-k}N}$. A Problem is that $\sqrt{2^{-k}N}$ is not necessarily an integer, and even $\sqrt{2^{-k}N}$ is an integer, it does not necessarily hold that N is an integer multiple of S. To address this issue, we can overlap the beam coverage of the sub-arrays, i.e., the angle space between adjacent sub-arrays can be less than $2S/N$. In such a case, the broadening factor may be less than S^2, and thus $S_k \geq \sqrt{2^{-k}N}$. Furthermore, we assume that both S_k and N are an integer powers of 2. With this assumption, it can be obtained that the number of sub-arrays for the k-th layer codewords is

$$S_k = 2^{\lceil(\log_2 N - k)/2\rceil}, \tag{30}$$

where $\lceil\cdot\rceil$ is the ceiling integer operation. It can be observed that the numbers of sub-arrays for the $\log_2 N$, $(\log_2 N - 1)$, $(\log_2 N - 2)$, $(\log_2 N - 3)$, $(\log_2 N - 4)$,...-th layer codewords are $1, 2, 2, 4, 4, \ldots$ With this setting, it is assured that $S_k \geq \sqrt{2^{-k}N}$ and N is an integer times of S_k.

2) The Weight Settings of the Sub-Arrays: Next, we need to set the AWVs of the S_k sub-arrays. For $\mathbf{w}(k,1)$, the steering angle space between adjacent sub-arrays is $\Delta = 2^{1-k}/S_k$, and the steering angles of the sub-arrays are

$$\omega_m = -1 + \frac{2m-1}{2}\Delta,\ m = 1, 2, \ldots, S_k. \tag{31}$$

Let $\mathbf{f}_m$ denote the AWV of the m-th sub-array. Considering the CM constraint, $\mathbf{f}_m$ can be expressed as

$$\mathbf{f}_m = \sqrt{\frac{N_S}{N}} e^{j\rho_m} \mathbf{a}(N_\mathrm{S}, \omega_m), \tag{32}$$

where $N_S = N/S_k$ is the number of antennas of each sub-array, $e^{j\rho_m}$ is a co-phase factor between different sub-arrays. With these notations, $\mathbf{w}(k,1)$ is set as

$$[\mathbf{w}(k,1)]_{(m-1)N_S+1:mN_S} = \mathbf{f}_m,\ m = 1, 2, \ldots, S_k. \tag{33}$$

It is clear that with the setting in (33) $\mathbf{w}(k,1)$ obeys the CM constraint and has a unit 2-norm. The remaining issue is to determine the co-phases ρ_m.

3) Co-Phase Optimization of the Sub-Arrays: To optimize the co-phases is challenging because there are two objectives for the codeword design. The first one is to maximize the beam gain along the main lobe direction, and the other is to minimize the gain fluctuation within the beam coverage. Even we can formulate an optimization problem, the number of variables can be large when S_k is large, which means that the numerical search method may be of high computational complexity. In this subsection, we propose an intuitive approach to formulate an optimization problem, and we find a suboptimal solution with closed form for the problem.

With the setting in (33) for $\mathbf{w}(k,1)$, it is guaranteed that the main power of the antenna array is within the beam coverage $[-1, -1 + \frac{2}{2^k}]$. Since the sub-arrays steer along ω_m, $m = 1, 2, \ldots, N_S$, the beam gains along ω_m would be large. To reduce the gain fluctuation, we hope that the intersection points in the angle domain of the beam regions of the sub-arrays, i.e., $\Omega_\ell = -1 + \ell\Delta$, $\ell = 1, 2, \ldots, N_S - 1$, also have high beam gains. As a result, we can formulate the following optimization problems

$$\underset{\rho_m}{\text{maximize}}\ |A(\mathbf{w}(k,1), \Omega_\ell)|,\ \ell = 1, 2, \ldots, N_S - 1, \tag{34}$$

where $\mathbf{w}(k,1)$ is shown in (33), and $\Omega_\ell = -1 + \ell\Delta$. Note that ρ_m are involved in all the $(N_S - 1)$ optimization problems; thus it is almost impossible to find an optimal solution for all these problems. Fortunately, with some manipulations we are able to find a suboptimal solution with closed form. Details are shown in Appendix A, and the final solution is,

$$\rho_m = -\pi m(N_\mathrm{S} - 1)\Delta/2 - \pi N_\mathrm{S} m(m-1)\Delta/2, \tag{35}$$

where $\Delta = 2^{1-k}/S_k$.

4) Codebook Generation: In this part, we summary the codebook generation with the proposed enhanced sub-array technique.

Recall that we need to design $\mathbf{w}(k,n)$ with beam widths $2/2^k$ in the k-th layer. For $k = 0, 1, 2, \ldots, \log_2 N$, we follow the following procedures to compute $\mathbf{w}(k,n)$:

- Separate $\mathbf{w}(k,1)$ into $S_k = 2^{\lceil(\log_2 N - k)/2\rceil}$ sub-arrays; thus each sub-array has $N_S = N/S_k$ antennas;
- Set the AWVs of the S_k sub-arrays: for $m = 1, 2, \ldots, S_k$, set $[\mathbf{w}(k,1)]_{(m-1)N_S+1:mN_S} = \sqrt{\frac{N_S}{N}} e^{j\rho_m} \mathbf{a}(N_\mathrm{S}, \omega_m)$, where ρ_m is shown in (35), and $\mathbf{a}(N_\mathrm{S}, \omega_m)$ is shown in (26);
- According to Lemma 1, we have $\mathbf{w}(k,n) = \mathbf{w}(k,1) \circ \sqrt{N}\mathbf{a}(N, \frac{2(n-1)}{N})$, $n = 2, 3, \ldots, 2^k$, where $\circ$ is the entry-wise product.

It is clear that there is no deactivation operation for all the codewords. Thus, unlike the deactivation method in [12] and the joint sub-array and deactivation method in [11], the proposed codebook does not require an on-off switch in each antenna branch and increases the maximal total transmission power. Fig. 3 shows the beam pattern comparison between the proposed enhanced sub-array scheme (Proposed) and the joint sub-array and deactivation method in [11] (JOINT), where we can find that for the 1st and the 3rd layer codewords, the enhanced scheme can achieve a significantly higher beam gain, due to no deactivation operation. Meanwhile, we can find that for the proposed scheme, the beam width of $\mathbf{w}(1,1)$ is indeed roughly 2 times of that of $\mathbf{w}(2,1)$, and 4 times of that of $\mathbf{w}(3,1)$, which are in accordance with that in Fig. 2.

V. Performance Comparisons

In this section, we evaluate the performance of the proposed MDR approach, as well as the designed codebook. In the simulations, both line-of-sight (LOS) LOS and non-LOS (NLOS) channel models are considered based on (2). For LOS channel, the first MPC has a constant coefficient and random AoD and AoA, while the other NLOS MPCs have complex Gaussian-distributed coefficients and random AoDs and AoAs [11], [23]. The LOS MPC is generally much stronger than the NLOS MPCs. For NLOS channel, all the MPCs have complex Gaussian-distributed coefficients with the same variance

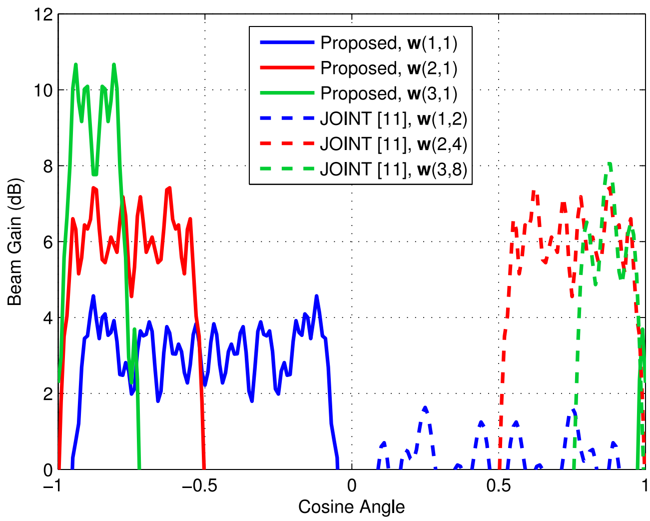


Fig. 3. Beam Comparison between the proposed enhanced sub-array scheme (Proposed) and the joint sub-array and deactivation method in [11] (JOINT) with a ULA, where $N = 64$.

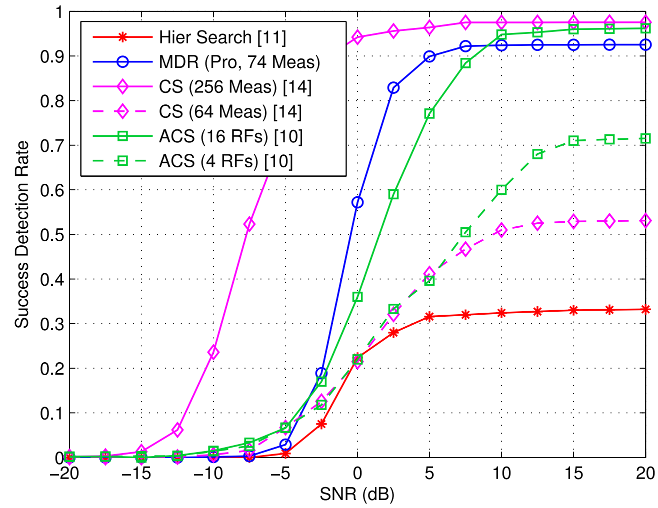


Fig. 5. Comparison of the detection performance between different approaches, where $K = L = 2$. The success detection rate here means the rate to successfully detect all the MPCs. An NLOS channel model is adopted.

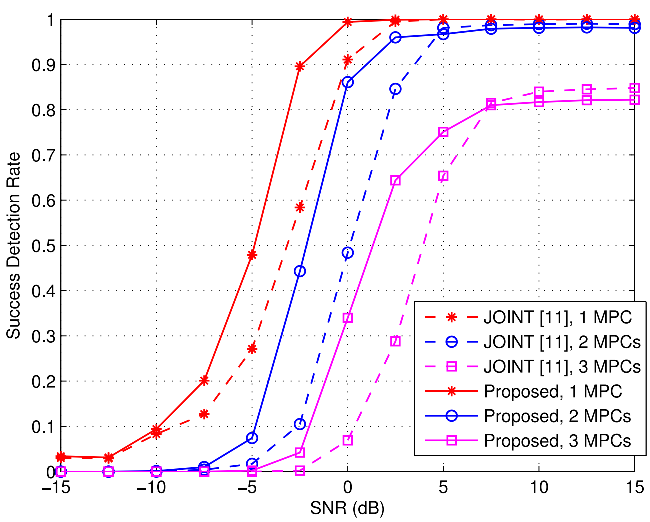


Fig. 4. Comparison of success detection rate of varying numbers of MPCs between the MDR approach with the proposed codebook and with the JOINT codebook designed in [11], where $K = 2$, $L = 3$. An NLOS channel model is adopted.

and random AoDs and AoAs [10], [11], [23]. Both the LOS and NLOS channels are sparse in the angle domain, because the number of MPCs is much smaller than the numbers of the Tx/Rx antennas [10], [11], [23]. Besides, $N_{\mathrm{r}} = N_{\mathrm{t}} = 32$ for all the simulations. The results are based on the average performance of 10^3 channel realizations.

First, we compare the performance of the proposed codebook with the JOINT codebook designed in [11][4]. Fig. 4 shows the comparison of success detection rate of varying numbers of MPCs between the MDR approach with the proposed codebook and with the JOINT codebook, where $K = 2$, $L = 3$. It is shown that the proposed codebook can achieve a significant signal-to-noise ratio (SNR) gain at the mediate to low SNR range. This benefit over JOINT is due to that there is no deactivation operation on the antennas for all the codewords in the proposed codebook. In contrast, some codewords of the JOINT codebook need to turn off half of the antennas, which results in a reduction of the maximal transmission power. However, at the high SNR regime, JOINT behaves slightly better than the proposed codebook, because the steering angle space between adjacent sub-arrays of the proposed codebook is smaller than that of JOINT, which leads to severer fluctuation on the beam pattern. Overall, the proposed codebook may be more attractive in mmWave communications, because the SNR is typically mediate to low, and a strict CM helps to reduce the hardware complexity.

Next, we compare the performance of the proposed MDR approach with the other alternatives in terms of success detection rate of MPCs and mean square error (MSE) of channel estimation, which is defined by $\mathrm{MSE} = \mathbb{E}(\|\mathbf{H} - \hat{\mathbf{H}}\|_{\mathrm{F}}^2)/(N_{\mathrm{r}} N_{\mathrm{t}})$, where $\hat{\mathbf{H}}$ is the estimated channel matrix. The involved methods include the hierarchical search method in [11], the CS method in [14], the ACS method in [10], and the LS method introduced in the beginning of Section II-B. The hierarchical search method is to directly search multiple real MPCs using the same normal-resolution codebook designed in this paper, and the ACS method exploits the codebook designed in [10].

Figs. 5 and 6 show the comparison results of success detection rate of MPCs and MSE of channel estimation, respectively, where $K = L = 2$. From these two figures we can find that the conventional hierarchical search method achieves poor performance. That is because the estimation error of AoAs and AoDs of the MPCs is significant due to the limited angle resolution of the codebook, which results in significant residual interference as defined in (4). Additionally, performances of the CS approach are highly dependent on the number of measurements, while those of the ACS approach are highly dependent on the number of RF chains. Only when the number of measurements and the number of RF chains are large enough, the CS and the ACS

[4]The comparison between the JOINT codebook and the alternatives can be found in [11], where it is shown that the JOINT codebook is superior than the alternatives.

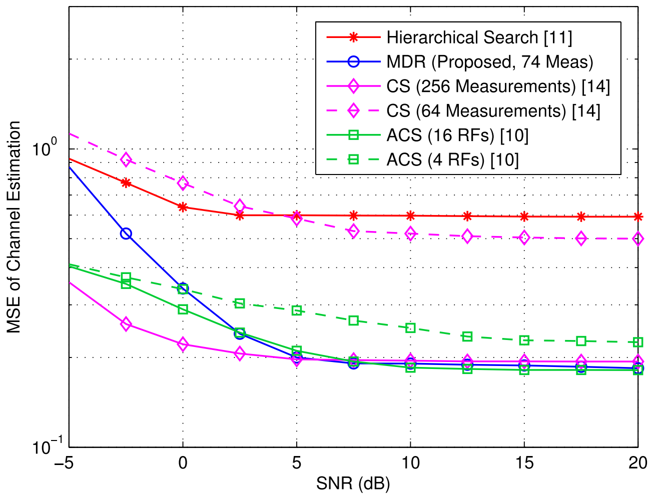


Fig. 6. Comparison of the MSE performance between different approaches, where $K = L = 2$. An NLOS channel model is adopted.

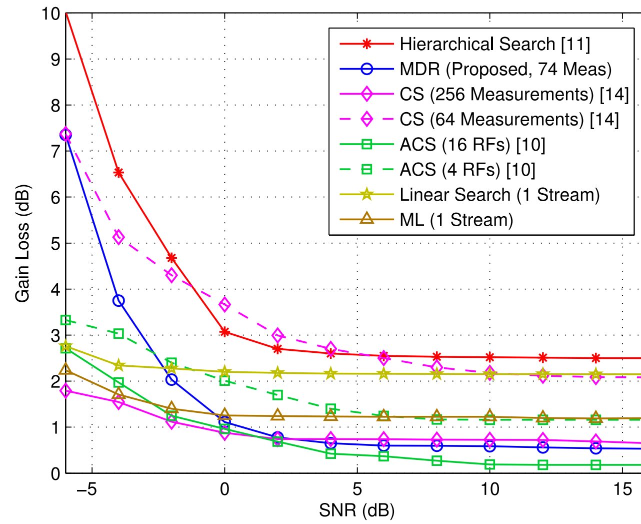


Fig. 7. Comparison of the gain-loss performance between different approaches, where $K = L = 2$. The ML (1 Stream) method exploits 256 measurements. An NLOS channel model is adopted.

approaches, respectively, can achieve satisfactory performances; otherwise their performances will be not satisfactory. In comparison, the proposed MDR approach can achieve promising success detection rate and MSE performances, with only one single RF chain and a smaller number of measurements.

In addition to the MSE performance, a more direct metric to evaluate the performance of mmWave channel estimation is the relative gain loss of beamforming. For each method, the beamforming vectors at the Tx and Rx are set to the right and left singular vectors of the estimated channel matrices, respectively, and then a practical beam gain can be obtained with the Tx/Rx beamforming vectors. The relative gain loss for a method is defined as the ideal beam gain subtracting the obtained practical beam gain in dB. Except the involved typical mmWave channel estimation methods, there are also some baseline 1-stream beamforming methods, which do not need to estimate a full channel. For instance, the line search method is to measure the channel on all AoA/AoD pairs ($N_\mathrm{r} N_\mathrm{t}$ pairs in total) and select the one with the maximum received energy. Another approach is that we make some measurements with random CM Tx/Rx AWVs, just the same as the CS method, and then we approximate the channel as a single path channel with a single AoA/AoD such that we can estimate them by using the maximal likelihood (ML) method.

Fig. 7 shows the comparison results of gain loss, where $K = L = 2$. Again we observe that the performance of the CS approach is highly dependent on the number of measurements, while those of the ACS approach are highly dependent on the number of RF chains. Only when the number of measurements and the number of RF chains are large enough, the CS and the ACS approaches can achieve satisfactory performances; otherwise their performances will be not satisfactory. In comparison, the proposed MDR approach can achieve promising gain-loss performances, with only one single RF chain and a smaller number of measurements. These results show a well agreement with those from Fig. 6. Moreover, we can find that the two 1-stream methods, i.e., Line Search (1 Stream) and ML (1 Stream), behave poorer than MDR with high SNR while better with low SNR, this is because with low SNR MDR induces additional noise when estimating more MPCs, while with high SNR MDR can make more accurate estimation of the other MPCs; thus it obtains more channel energy by estimating more MPCs.

The above comparisons exploit an NOLS channel model, where all the MPCs have the same average power and a uniformly distributed AoA/AoD within $[-1, 1]$. In practice there may be LOS channel where the LOS component is dominant over the NLOS MPCs, and clustering channel where each MPC has angular spread at Tx/Rx rather than only a single AoA/AoD. Moreover, the angle range of AoA/AoD may be a subset of $[-1, 1]$. To make sufficient performance evaluations, we also compare the gain-loss performance of MDR under different types of channels, as shown in Fig. 8, where $K = L = 2$. In the comparison, the non-clustering channel model refers to the one in (2), while the clustering channel model uses the one in [22], where the number of clusters is $N_\mathrm{cl} = L$, the number of rays within a cluster is $N_\mathrm{ray} = 4$, and the standard deviation of angular spread is $1/N_\mathrm{t}/5$ for AoD and $1/N_\mathrm{r}/5$ for AoA. The smaller angle range model also exploits the one in (2), but the angle ranges of all AoAs/AoDs are $[0, 1]$ rather than $[-1, 1]$. For the LOS channel, the power of the LOS components is 10 dB higher than the NLOS components. From the comparison results we can observe that the angle range affects little on the gain-loss performance. In addition, MDR has less gain loss at low SNR when exploiting a clustering model, which means that it is easier for MDR to collect channel energy under a clustering channel than a non-clustering channel. Moreover, comparing the upper sub-figure with the bottom sub-figure, we can find that MDR works well under both NLOS and LOS channels, and it achieves slightly better performance under a LOS channel.

Lastly, we evaluate the training overhead that the proposed MDR approach requires and compare it with the other approaches. For a fair comparison, we consider only one-stream transmission, and do not count the overhead reduction that may be achieved via multi-stream transmission with multiple RF chains. Moreover, the training overhead is measured by the number of measurements for channel estimation, and for

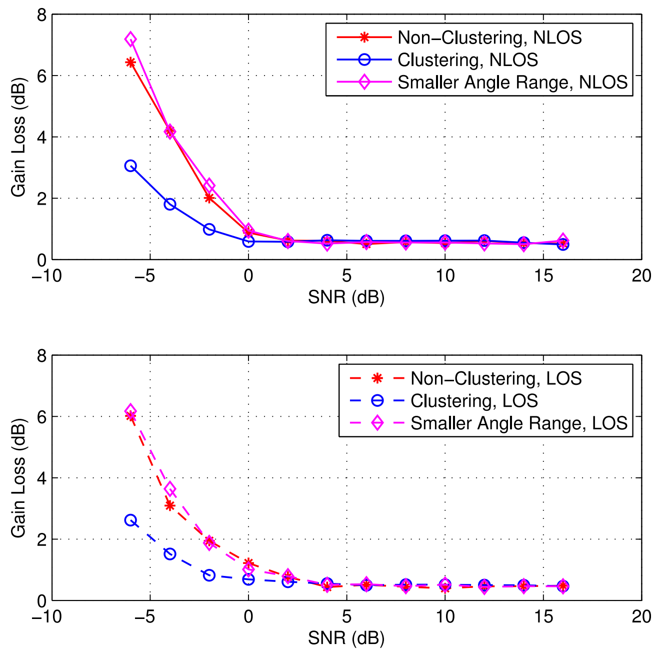


Fig. 8. Comparison of the gain-loss performance of MDR between different channels, where $K = L = 2$.

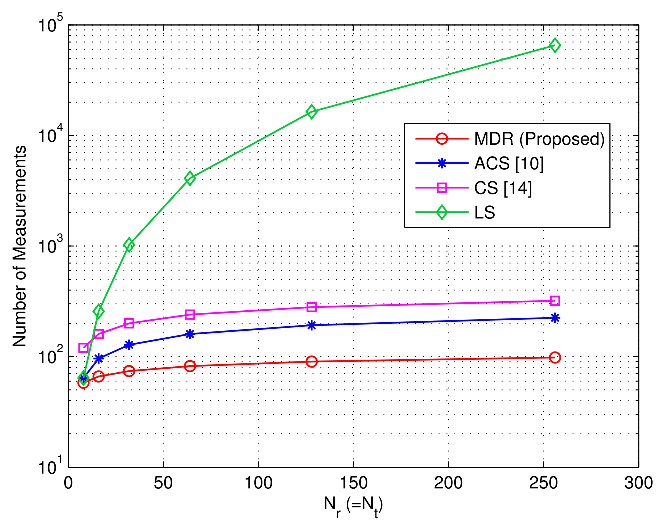


Fig. 9. The comparison of required training overhead between different approaches.

convenience, we assume $N_{\rm r} = N_{\rm t} = N$ and $L = 2$. As we know, the conventional LS requires at least $N_{\rm r}N_{\rm t} = N^2$ measurements. The proposed MDR approach requires $T_{\rm MDR}$ measurements as shown in (24), where S_0 can be typically set to 2. According to [10], the ACS approach requires $2^2L^3\log_2(N/L)$ measurements. For the pure CS approach, the required number of measurements is typically $\rho\log_2(N_{\rm seq})$ [29], where $N_{\rm seq}$ is the length of the candidate row vectors in the dictionary matrix. In the context of this paper, $N_{\rm seq} = N_{\rm r}N_{\rm t}$. To guarantee a satisfactory performance, we set $\rho = 20$ here. Fig. 9 shows the comparison result, where we can find that the proposed MDR approach requires the least training overhead.

VI. Conclusion

To enable fast and accurate channel estimation for mmWave communications, we have proposed an MDR approach in this paper. The proposed approach has two stages. In the first stage, a real MPC is decomposed into several virtual MPCs, and the virtual MPCs are acquired by using the hierarchical search method based on a normal-resolution codebook. Then in the second stage, the real MPCs are recovered from the virtual MPCs acquired in the first stage, and can be reconstructed using a sparse compressive sensing approach. By exploiting the results of the virtual multipath acquisition, the size of the dictionary matrix is greatly reduced, and thus the channel recovery problem can be efficiently solved by the classic OMP method.

Moreover, to make the proposed approach applicable for both analog and hybrid beamforming/combining devices, we have particularly designed a codebook for the hierarchical search by using an enhanced sub-array technique. The codebook satisfies strict CM constraint; thus can reduce hardware complexity compared with JOINT, a state-of-the-art hierarchical codebook. Performance comparisons show that the designed codebook is superior than JOINT at mediate to low SNR regime. Moreover, compared with the other candidate approaches, the proposed MDR approach achieves promising MPC detection and MSE performances, while requires less training overhead.

Appendix A
A Solution of Problem (34)

Recalling that we have divided $\mathbf{w}(k,1)$ into S_k virtual sub-arrays, and each sub-array has N_S antennas. The steering angle space between adjacent sub-arrays is $\Delta = 2^{1-k}/S_k$, and the steering angles of the sub-arrays are $\omega_m = -1 + \frac{2m-1}{2}\Delta$, $m = 1, 2, ..., S_k$. In addition, $\mathbf{w}(k,n)$ is set as (33). According to the definition of beam gain as shown in (25), we have

$$\begin{aligned} A(\mathbf{w}(k,1),\Omega_\ell) &= \sum_{n=1}^{N}[\mathbf{w}(k,1)]_n e^{-j\pi(n-1)\Omega_\ell} \\ &= \sum_{m=1}^{S_k}\sum_{n=1}^{N_{\rm S}}[\mathbf{w}(k,1)]_{(m-1)N_{\rm S}+n} e^{-j\pi((m-1)N_{\rm S}+n-1)\Omega_\ell} \\ &= \frac{N_S}{\sqrt{N}}\sum_{m=1}^{S_k} e^{-j\pi(m-1)N_{\rm S}\Omega_\ell} e^{j\rho_m}\mathbf{a}(N_{\rm S},\Omega_\ell)^{\rm H}\mathbf{a}(N_{\rm S},\omega_m). \end{aligned} \tag{36}$$

It is clear that it is still complicated to determine ρ_m by optimizing the absolute beam gain in (36). Notice that $|\mathbf{a}(N_{\rm S},\omega_1)^{\rm H}\mathbf{a}(N_{\rm S},\omega_2)|$ becomes smaller when $|\omega_1 - \omega_2|$ becomes greater from 0 to $2/N_{\rm S}$, and can be neglected when $|\omega_1 - \omega_2| > 2/N_{\rm S}$. This means that the two sub-arrays with steering angles closest to Ω_ℓ have the most significant effects on the beam gain along Ω_ℓ, while the sub-arrays with steering angles far from Ω_ℓ have much smaller effect on the beam gain along Ω_ℓ. This motivates us to consider only the two close sub-arrays when optimizing the beam gain for simplicity. Since $\Omega_\ell = -1 + \ell\Delta$, the two close steering angles are ω_ℓ and $\omega_{\ell+1}$.

Consequently, we have

$$\begin{aligned}&A(\mathbf{w}(k,1),\Omega_\ell)\\&\approx \frac{N_S}{\sqrt{N}}e^{-j\pi(\ell-1)N_\mathrm{S}\Omega_\ell}e^{j\rho_\ell}\mathbf{a}(N_\mathrm{S},\Omega_\ell)^\mathrm{H}\mathbf{a}(N_\mathrm{S},\omega_\ell)\\&\quad+\frac{N_S}{\sqrt{N}}e^{-j\pi\ell N_\mathrm{S}\Omega_{\ell+1}}e^{j\rho_{\ell+1}}\mathbf{a}(N_\mathrm{S},\Omega_\ell)^\mathrm{H}\mathbf{a}(N_\mathrm{S},\omega_{\ell+1})\\&=\frac{1}{\sqrt{N}}e^{-j\pi(\ell-1)N_\mathrm{S}\Omega_\ell}e^{j\rho_\ell}\\&\quad\times\left(\sum_{i=1}^{N_\mathrm{S}}e^{-j\pi(i-1)\Delta/2}+e^{j\pi N_\mathrm{S}\Omega_\ell}e^{j(\rho_{\ell+1}-\rho_\ell)}\sum_{i=1}^{N_\mathrm{S}}e^{j\pi(i-1)\Delta/2}\right).\end{aligned}\tag{37}$$

Thus, we further obtain

$$\begin{aligned}&A(\mathbf{w}(k,1),\Omega_\ell)\\&=\frac{1}{\sqrt{N}}e^{-j\pi(\ell-1)N_\mathrm{S}\Omega_\ell}e^{j\rho_\ell}\times\Bigg(e^{-j\pi(N_\mathrm{S}-1)\Delta/4}\frac{\sin(-N_\mathrm{S}\pi\Delta/4)}{\sin(-\pi\Delta/4)}\\&\quad+e^{j\pi N_\mathrm{S}\Omega_\ell}e^{j(\rho_{\ell+1}-\rho_\ell)}e^{j\pi(N_\mathrm{S}-1)\Delta/4}\frac{\sin(N_\mathrm{S}\pi\Delta/4)}{\sin(\pi\Delta/4)}\Bigg)\\&=\frac{1}{\sqrt{N}}e^{-j\pi(\ell-1)N_\mathrm{S}\Omega_\ell}e^{j\rho_\ell}\times\frac{\sin(N_\mathrm{S}\pi\Delta/4)}{\sin(\pi\Delta/4)}\\&\quad\times\left(e^{-j\pi(N_\mathrm{S}-1)\Delta/4}+e^{j\pi N_\mathrm{S}\Omega_\ell}e^{j(\rho_{\ell+1}-\rho_\ell)}e^{j\pi(N_\mathrm{S}-1)\Delta/4}\right).\end{aligned}\tag{38}$$

It is clear that to optimize $|A(\mathbf{w}(k,1)|$, it should hold that

$$2n\pi-\pi(N_\mathrm{S}-1)\Delta/4=\pi N_\mathrm{S}\Omega_\ell+(\rho_{\ell+1}-\rho_\ell)+\pi(N_\mathrm{S}-1)\Delta/4,\tag{39}$$

where n is an arbitrary integer. Without loss of generality, we set $n=0$. As N_S is also an integer power of 2, we have

$$\rho_{\ell+1}-\rho_\ell=-\pi(N_\mathrm{S}-1)\Delta/2-\pi N_\mathrm{S}\ell\Delta.\tag{40}$$

Finally, we can find a solution

$$\rho_\ell=-\pi\ell(N_\mathrm{S}-1)\Delta/2-\pi N_\mathrm{S}\ell(\ell-1)\Delta/2.\tag{41}$$

References

[1] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[2] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[3] S. Sun, T. S. Rappaport, R. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[4] Z. Xiao, P. Xia, and X.-G. Xia, "Full-duplex millimeter-wave communication," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 136–143, Dec. 2017.

[5] Y. Niu, C. Gao, Y. Li, D. Jin, L. Su, and D. Wu, "Boosting spatial reuse via multiple-path multihop scheduling for directional mmwave WPANs," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6614–6627, Aug. 2016.

[6] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, Apr. 2015.

[7] Y. Niu *et al.*, "Energy-efficient scheduling for mmWave backhauling of small cells in heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2674–2687, Mar. 2017.

[8] J. Qiao, X. Shen, J. W. Mark, and Y. He, "MAC-layer concurrent beamforming protocol for indoor millimeter-wave networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 327–338, Jan. 2015.

[9] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[10] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[11] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[12] T. He and Z. Xiao, "Suboptimal beam search algorithm and codebook design for millimeter-wave communications," *Mobile Netw. Appl.*, vol. 20, no. 1, pp. 86–97, Jan. 2015.

[13] A. Alkhateeby, G. Leusz, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 2909–2913.

[14] J. Lee, G.-T. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. IEEE Global Commun. Conf.*, 2014, pp. 3326–3331.

[15] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[16] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, "Design and optimization on training sequence for mmWave communications: A new approach for sparse channel estimation in massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1486–1497, Jul. 2017.

[17] X. Ma, F. Yang, S. Liu, W. Ding, and J. Song, "Structured compressive sensing-based channel estimation for time frequency training OFDM systems over doubly selective channel," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 266–269, Apr. 2017.

[18] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, "Sparse channel estimation for MIMO-OFDM systems in high-mobility situations," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 6113–6124, Jul. 2018.

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

[20] T. S. Rappaport, Y. Qiao, J. I. Tamir, J. N. Murdock, and E. Ben-Dor, "Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems," in *Proc. IEEE Radio Wireless Symp.*, Jan. 2012, pp. 151–154.

[21] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X. G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, May 2018.

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

[23] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[24] J. Nsenga, W. Van Thillo, F. Horlin, V. Ramon, A. Bourdoux, and R. Lauwereins, "Joint transmit and receive analog beamforming in 60 GHz MIMO multipath channels," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009, pp. 1–5.

[25] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.

[26] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook based beamforming sequence design in millimeter-wave systems," in *Proc. IEEE Global Telecommun. Conf.*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

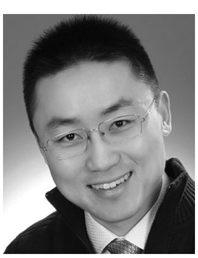
[27] P. A. Eliasi, S. Rangan, and T. S. Rappaport, "Low-rank spatial channel estimation for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2748–2759, May 2017.

[28] E. Adabi Firouzjaei, "mm-wave phase shifters and switches," Ph.D. dissertation, Elect. Eng. Comput. Sci. Dept., Univ. California Berkeley, Berkeley, CA, USA, 2010.

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

Zhenyu Xiao (M'11–SM'17) received the B.E. degree from Huazhong University of Science and Technology, Wuhan, China, in 2006, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2011. From 2011 to 2013, he held a postdoctoral position with the Electronic Engineering Department, Tsinghua University. From 2013 to 2016, he was a Lecturer with the Department of Electronic and Information Engineering, Beihang University, Beijing, China, where he is currently an Associate Professor. He has authored or coauthored more than 50 papers and was a Reviewer for the IEEE Transactions on Signal Processing, IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, IEEE Communications Letters, etc. He was a TPC member of the IEEE GLOBECOM'12, IEEE WCSP'12, IEEE ICC'15, etc. His research interests include communication signal processing and practical system implementation for wideband communication systems, which cover synchronization, multipath signal processing, diversity, multiple antenna technology, etc. His current research focuses on millimeter-wave 5G and airborne communications.

Hang Dong received the B.S. degree from the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China, in 2017. He is currently working toward the M.S. degree with the Department of Electronic and Information Engineering, Beihang University, Beijing, China. His current research focuses on millimeter-wave communications and array signal processing.



Lin Bai (M'13–SM'17) received the B.Sc. degree in electronics and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 2004, the M.Sc. (with distinction) degree in communication systems from the University of Wales, Swansea, U.K., in 2007, and the Ph.D. degree in advanced telecommunications from the School of Engineering, Swansea University, Swansea, U.K., in 2010. Since 2011, he has been an Associate Professor/Ph.D. Supervisor with the School of Electronics and Information Engineering, Beihang University (Beijing University of Aeronautics and Astronautics), Beijing, China. He has authored two books namely *Low Complexity MIMO Detection* (Springer, 2012) and *Low Complexity MIMO Receivers* (Springer, 2014). His research interests include signal processing of wireless communications, particularly multiple-input multiple-output systems, array/smart antenna, lattice-based approaches, and nonorthogonal multiple access.

Dr. Bai is currently the Symposium Cochair for the 2019 IEEE Global Communications Conference. He is currently a Lead Guest Editor for the IEEE Wireless Communication Magazine and a Guest Editor for the IEEE Internet of Things Journal. He is an Editor/Associate Editor for several academic journals including the IEEE Wireless Communication Letters, IEEE Access, IET Communications, and *KSII Transactions on Internet and Information Systems*, and a Managing Editor for the *Journal of Communications and Information Networks*. He was a Guest Editor for the *International Journal of Distributed Sensor Networks* from 2012 to 2014. He was the recipient of the IEEE Communications Letters Exemplary Reviewers Certificate for 2012 and a co-recipient of the Best Student Paper Award from ICNS 2013 (conference).

Pengfei Xia (S'03–M'05–SM'10) received the Ph.D. degree from University of Minnesota, Minneapolis, MN, USA, in 2005. He is currently a Full Chair Professor with the College of Electronics and Information, Tongji University, Shanghai, China. He has coauthored more than 70 journal and conference papers and has more than 100 international patents (including patent applications). His research interests include the intersection of wireless communications and signal processing, including mm-wave communications, transceiver beamforming, signal/codebook designs, compressive sensing, 4G/5G cellular systems, and IEEE 802.11 WLANs. He was the corecipient of the IEEE Signal Processing Society Best Paper Award in 2011 and coedited one of the first book on 60 GHz millimeter wave communication systems. He is currently an Associate Editor for the IEEE Transactions on Signal Processing and the IEEE Communications Magazine.

Xiang-Gen Xia (M'97–SM'00–F'09) received the B.S. degree in mathematics from Nanjing Normal University, Nanjing, China, the M.S. degree in mathematics from Nankai University, Tianjin, China, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1983, 1986, and 1992, respectively.

During 1995–1996, he was a Senior/Research Staff Member with Hughes Research Laboratories, Malibu, CA, USA. In September 1996, he joined the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA, where he is currently the Charles Black Evans Professor. He has authored the book *Modulated Coding for Intersymbol Interference Channels* (Marcel Dekker, 2000). His research interests include space-time coding, MIMO and OFDM systems, digital signal processing, and SAR and ISAR imagings.

Dr. Xia is currently serving and has served as an Associate Editor for numerous international journals including the IEEE Wireless Communications Letters, IEEE Transactions on Signal Processing, IEEE Transactions on Wireless Communications, IEEE Transactions on Mobile Computing, and IEEE Transactions on Vehicular Technology. He is the Technical Program Chair for the Signal Processing Symposium, Globecom 2007 in Washington D.C. and the General Co-Chair of ICASSP 2005 in Philadelphia. He was the recipient of the National Science Foundation Faculty Early Career Development Program Award in 1997, the Office of Naval Research Young Investigator Award in 1998, and the Outstanding Overseas Young Investigator Award from the National Nature Science Foundation of China in 2001. He was also the recipient of the Outstanding Junior Faculty Award of the Engineering School of the University of Delaware in 2001.