Harmonic Retrieval Based Baseband Channel Estimation for Millimeter Wave OFDM Systems

Ziyuan Sha [ID], Zhaocheng Wang [ID], *Senior Member, IEEE*, and Sheng Chen [ID], *Fellow, IEEE*

Abstract—For massive multi-input multi-output (MIMO) enhanced millimeter wave (mmWave) frequency-division multiplexing (OFDM) systems, channel estimation (CE) is challenging. In this paper, the baseband CE of mmWave based MIMO-OFDM systems after beam searching is formulated as a harmonic retrieval (HR) problem, where each path of the channel represents a harmonic with its frequency and strength to be estimated. We propose two methods, a windowed orthogonal matching pursuit (window-OMP) method and a windowed discrete Fourier transform (window-DFT) method, to approximately acquire the maximum likelihood (ML) estimate of the baseband CE. The window-OMP method is capable of approximating the ML estimator with high accuracy, while the window-DFT method has a lower complexity and is shown to acquire approximate ML estimate under the assumption that the frequencies of harmonics are sufficiently separated. Theoretical analysis is performed to derive a closed-form Cramér–Rao lower bound as well as to investigate the effect of wrong paths to the estimation accuracy. A simulation study is conducted to investigate the performance of the proposed methods, and the results obtained verify that our methods outperform the existing estimation of signal parameters by rotational invariant techniques based HR method and the conventional interpolation method.

Index Terms—Millimeter wave, massive MIMO, OFDM, baseband channel estimation, orthogonal matching pursuit, discrete Fourier transform, windowing.

I. INTRODUCTION

IN THE next generation cellular communication system commonly known as 5G, millimeter wave (mmWave) communication and massive multi-input multi-output (MIMO) [1]–[6] will be two key technologies to meet the explosive growth of mobile traffic. MmWave usually represents the frequency band from 30 GHz to 300 GHz. While most of the available bandwidths in lower frequency bands have already been occupied [1], the mmWave band offers abundant bandwidth to support bandwidth-hungry high-data-rate applications. However, the propagation pathloss is extremely high in the mmWave band. Fortunately, the size of antenna is proportional to the wavelength, and this makes it feasible to employ very large-scale antenna arrays at both transmitter and receiver of mmWave communication systems. According to [7], the benefit of high directional power gain of beamforming can sufficiently compensate the high pathloss in the mmWave band, which guarantees a high signal-to-noise ratio (SNR) at the receiver to support extremely high data rate transmission with large bandwidth.

Hardware constraints also impose a great challenge in mmWave communication [2]. In particular, it is impossible to apply powerful full digital transceiver design in the mmWave band. This is because such a full digital design would require an equally large number of radio frequency (RF) chains to match the large number of antennas employed, which would impose too high hardware complexity and consume too much power. Therefore, the hybrid analog-digital transceiver architecture [2], [8]–[10] is more practical, where beamforming is accomplished by analog phase shifters or switches. Thus, one RF chain is connected to several antennas via digitally controlled phase shifters, and the number of RF chains needed is much smaller than the number of antennas employed.

In the analog domain, beam searching technique provides high directional power gain with low complexity [2], [11], [12]. Specifically, the phase shifters are controlled by beam steering vectors which determine the directions of beams. According to [11], in single RF chain scenario, the optimal pair of beam steering vectors are the antenna response vectors corresponding to the angle of arrival (AOA) and angle of departure (AOD) of the strongest path. However, for practical consideration, the phase resolution of phase shifter is finite and the beam steering vector is usually selected from a codebook, e.g., the columns of a discrete Fourier transform (DFT) matrix. The study [13] has verified that the SNR loss is slight (maybe less than 0.25 dB) due to the finite resolution of phase shifter. A beam searching procedure [12] needs to be conducted to select the optimal beam steering vector pair at the transmitter and receiver from their codebooks, which may play a key role in the initial access of 5G in the future.

Like lower-frequency band based MIMO, where accurate channel estimation (CE) is crucial to realize the full MIMO

Z. Sha is with the Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: shazy17@mails.tsinghua.edu.cn).

Z. Wang is with the Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, and also with the Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China (e-mail: zcwang@tsinghua.edu.cn).

S. Chen is with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K., and also with the King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: sqc@ecs.soton.ac.uk).

Digital Object Identifier 10.1109/TVT.2019.2895611

performance potential [14]–[16], CE is critically important to perform precoding and combining for mmWave MIMO [17]. Most existing works on mmWave CE exploit the sparsity of mmWave channel to design low complexity algorithms for CE [8], [18]–[25]. More specifically, in [8], the channel is measured by the pilot signal transmitted via different beam steering vectors and the sparsity of propagation path is utilized to design compressed sensing (CS) based algorithm. In the works [18]–[21], the sparsity in angular domain is utilized to estimate the AOA, where CS based algorithms are proposed in [18], [20], [21] and a DFT based method is proposed in [19]. Likewise, the time-domain sparsity of paths is utilized in [22]. Furthermore, many methods are proposed to exploit the sparsity of mmWave channel and most of them have focused on the estimation of the large channel matrix, or equivalently on estimating the channel parameters to acquire the large channel matrix, with widely adopted CS based methods. Additionally, the minimum mean square error (MMSE) criterion is adopted for mmWave CE in [24] and [25], where the CE methods proposed are applicable to both spatially uncorrelated channels and spatially correlated mmWave channels. In [25], the angular domain sparsity is also exploited to improve mmWave CE by incorporating CS method.

Most of the aforementioned approaches estimate the channel matrix before beam searching. Provided that the accurate CSI can be acquired, the approach of estimating the channel matrix before beam searching offers two advantages. First, the steering vectors can then be selected based on the estimated channel matrix, which are more accurate than those acquired by beam searching, especially for multiple beam searching. Second, the non-codebook based beamforming can be supported based on the estimated channel matrix, which achieves better performance than the codebook based beamforming. However, the SNR before beam searching may be too low to guarantee a sufficiently accurate CE due to the high path loss. Furthermore, the computational complexity and storage requirements imposed by the CE before beam searching to process large-dimensional matrices associated with large-scale antenna arrays may be excessively high. Consequently, the benefits of the CE before beam searching are limited.

In mmWave orthogonal frequency-division multiplexing (OFDM) systems, the complexity constraint becomes even more strict. This is because the number of subcarriers in mmWave OFDM is much larger than that of lower-frequency band OFDM counterpart, because of the huge bandwidth of mmWave systems. The channels on different subcarriers are different in frequency-selective fading channel, and they must be estimated and stored. Clearly, estimating the CSI of large-scale antenna array based mmWave OFDM systems before beam searching becomes challenging. Therefore, it is much more practical to estimate the equivalent digital baseband channel matrix after beam searching, which is a much smaller matrix whose size is determined by the number of RF chains rather than the number of antennas.

Against the above background, in this paper, we focus on the baseband CE of mmWave OFDM systems after beam searching, which overcomes the aforementioned two drawbacks of CE be-

fore beam searching. In this case, the conventional CE methods, such as the least squares (LS) estimator and the MMSE estimator, for small scale MIMO [26] can be adopted to estimate the frequency-domain (FD) CSI of pilot subcarriers, while the FD CSI of other subcarriers can be estimated via interpolation based method [27]. However, the sparsity of mmWave channels is not exploited in the conventional CE methods. In this paper, we show that the baseband CE problem of the widely adopted multi-path or multi-cluster channel can be formulated as a sparse complex harmonic retrieval (HR) problem, where each path or cluster represents a harmonic with its frequency and strength to be estimated. From the estimation theory perspective, the maximum likelihood (ML) estimator is asymptotically the unbiased efficient estimator [28], reaching the Cramér-Rao lower bound (CRLB). We propose two methods, a windowed orthogonal matching pursuit (window-OMP) method and a windowed discrete Fourier transform (window-DFT) method, to approximately acquire the ML estimate. Specifically, to solve the ML estimation problem for the baseband CE, the OMP algorithm derives a solution with relatively high accuracy, while the DFT method derives an approximate solution under the assumption that the frequencies of harmonics are sufficiently separated. The DFT method suffers an accuracy loss compared to the OMP method, while imposing lower complexity. Furthermore, window function is applied to mitigate the frequency leakage problem in both DFT and OMP based algorithms.

HR problems have been investigated by many existing works, where the subspace methods are employed widely, such as the multiple signal classification (MUSIC) [29] and the estimation of signal parameters by rotational invariant techniques (ESPRIT) [30], [31]. In particular, the works [32], [33] investigated a similar HR problem for single-input single-out (SISO) OFDM systems using the ESPRIT algorithm. The methods proposed in this paper are different from the existing works. Moreover, our methods outperform the ESPRIT method at high SNR region, as verified in the simulation study. Our main contributions are summarized as follows.

- We formulate the mmWave baseband CE as a HR problem for both multi-path and multi-cluster channel models. The sparsity information of the mmWave channel is effectively exploited to achieve better CE performance than conventional interpolation method.
- A window-OMP method and a window-DFT method are proposed. OMP and DFT method are both capable of acquiring approximate ML estimate, and window function is applied to further improve their accuracy. The proposed methods outperform the existing ESPRIT method.
- In our theoretical analysis, a closed-form approximate CRLB for the parameter estimate in the HR problem is derived. Then, we prove that the effect of wrong paths (harmonics) on the CE accuracy is very limited, which ensures the robustness of our proposed methods.
- Our proposed estimator reduces pilot overhead significantly, since the pilot pattern can be sparser and does not need to cover the whole bandwidth, which enables a much more flexible and effective as well as sparser pilot pattern design.

The rest of this paper is organized as follows. Section II introduces our system model, including the channel model and hybrid analog-digital architecture. Section III formulates the baseband CE after beam searching as a HR problem. In Section IV, we propose our window-OMP and window-DFT methods. Theoretical analysis is presented in Section V, while in Section VI, we conduct a simulation study to investigate the performances of our proposed methods. Section VII draws our conclusions.

In this paper, normal-face lower-case letters denote scalars, while boldface lower-case and upper-case symbols denote column vector and matrix, respectively. Matrix transpose, conjugate, conjugate transpose, inverse and pseudo-inverse are denoted by $(\cdot)^{\mathrm{T}}$, $(\cdot)^{*}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{-1}$ and $(\cdot)^{\dagger}$, respectively, while $\mathrm{range}(\cdot)$, $\mathrm{rank}(\cdot)$ and $\mathrm{tr}(\cdot)$ denote the range, rank, and trace of a matrix. The dimension of a subspace is denoted by $\dim(\cdot)$, and $\mathsf{E}[\cdot]$ is the expectation operator, while $\|\cdot\|_2$ represents the l_2-norm. $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, $\boldsymbol{0}_N$ is the N-dimensional zero vector, and $\boldsymbol{0}_{N \times N}$ is the $N \times N$ zero matrix, while $\mathrm{diag}\{a_1, \ldots, a_m\} = \mathrm{diag}\{a_i\}_{i=1}^m$ is the diagonal matrix with $a_1, \ldots, a_m$ at its diagonal entries.

II. SYSTEM MODEL

In this section, the mmWave massive MIMO channel model is introduced first, and the time-domain (TD) sparsity after beam searching is illustrated according to the channel model. Next, the hybrid analog-digital architecture and its signal model are presented. The transmitter has N_t antennas and $N_{RF,t}$ RF chains, while the receiver is equipped with N_r antennas and $N_{RF,r}$ RF chains. The number of RF chains is much smaller than the number of antennas due to hardware constraints, i.e., $N_{RF,t} \ll N_t$ and $N_{RF,r} \ll N_r$. In the context of cellular system, the transmitter represents a base station (BS) and the receiver represents a user equipment (UE) in downlink transmission, and vice versa in uplink transmission.

A. Channel Model

A widely adopted multi-path two-dimensional (2D) channel model [2], [8] is used in this paper, where the antenna arrays at transmitter and receiver are uniform linear arrays (ULAs). Assume that the antenna spacing is half of the wavelength. Then the normalized response vector of a ULA with N antennas at the AOA or AOD θ is expressed by

$$\boldsymbol{a}(\theta) = \frac{1}{\sqrt{N}} \left[1 \; e^{\mathrm{j}\pi \sin(\theta)} \cdots e^{\mathrm{j}(N-1)\pi \sin(\theta)} \right]^{\mathrm{T}} \in \mathbb{C}^N. \quad (1)$$

Further assume that there exist L paths in the mmWave channel. Let the lth path's complex gain and delay be α_l and τ_l, while $\theta_{r,l}$ and $\theta_{t,l}$ are the AOA and AOD of the lth path. The TD channel matrix can then be formulated as [20], [22]

$$\boldsymbol{C}(t) = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l \boldsymbol{a}_r(\theta_{r,l}) \boldsymbol{a}_t^{\mathrm{H}}(\theta_{t,l}) \delta(t - \tau_l), \quad (2)$$

where $\delta(t)$ is the Dirac delta function, $\boldsymbol{a}_r(\theta_{r,l}) \in \mathbb{C}^{N_r}$ is the receiver antenna response vector at the AOA $\theta_{r,l}$ and $\boldsymbol{a}_t(\theta_{t,l}) \in$
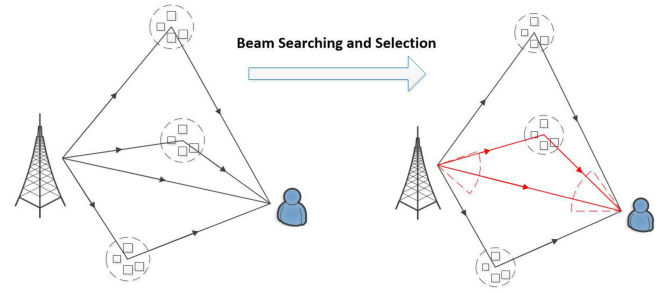

Fig. 1. The mmWave channel exhibits the sparsity in the propagation paths, and the sparsity becomes even stronger after beam searching.
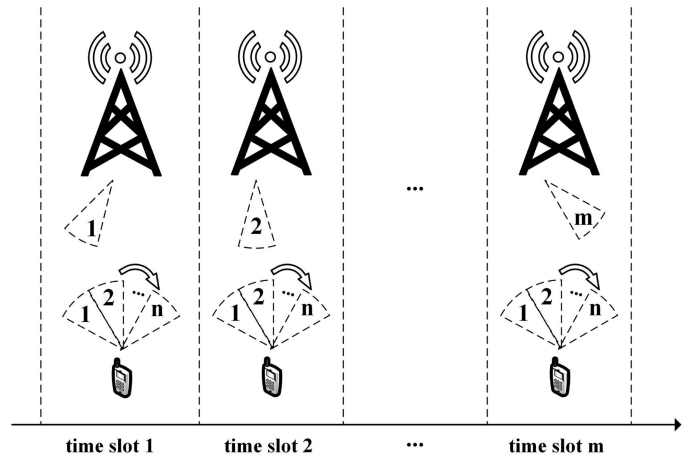

Fig. 2. Illustration of downlink beam searching procedure.

$\mathbb{C}^{N_t}$ is the transmitter antenna response vector at the AOD $\theta_{t,l}$. The fading is assumed to be sufficiently slow so that the channel is time-invariant. Based on (2), the FD multi-path channel, i.e., the channel frequency response $\boldsymbol{H}(f) \in \mathbb{C}^{N_r \times N_t}$, is given by

$$\boldsymbol{H}(f) = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l \boldsymbol{a}_r(\theta_{r,l}) \boldsymbol{a}_t^{\mathrm{H}}(\theta_{t,l}) e^{-\mathrm{j}2\pi \tau_l f}. \quad (3)$$

Because of the inherent sparse property of the mmWave channel [2], [17], the number of propagation paths L is usually small. Moreover, after beam searching, the transmitter and receiver will select the directions of beams to maximize the directional power gain as illustrated in Fig. 1. Since the paths whose AODs are outside the transmitter beam or whose AOAs are outside the receiver beam are filtered out, only fewer paths will actually influence the channel and, therefore, the effective L becomes even smaller. Sometimes, only the line-of-sight (LOS) path will be preserved. Consequently, the effective mmWave channel becomes much sparser after beam searching. Moreover, the beamwidth becomes narrower as the size of ULA increases, and as a result, there exist less paths whose AODs or AOAs are inside the beam of transmitter or receiver, which also indicates that the effective L becomes smaller. Therefore, the channel become sparser as the number of antennas increases.

Fig. 2 illustrates the downlink beam searching procedure with m beams at BS and n beams at UE. We discuss the exhaustive
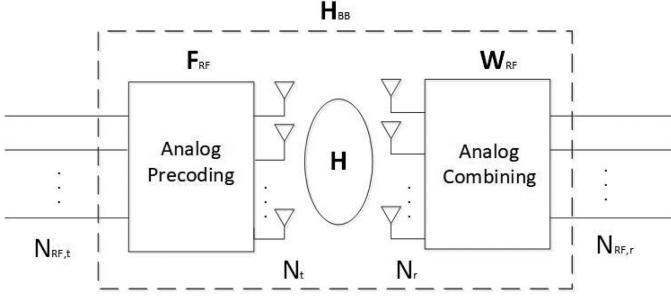
Fig. 3. Hybrid analog-digital transceiver model, where the baseband equivalent channel consists of the original channel as well as the analog precoding at transmitter and the analog combining at receiver. The digital precoding, which does not influence this baseband channel, is omitted.

beam searching method here. Specifically, the BS uses the beams 1 to m from the time slots 1 to m, respectively. In every time slot, the UE sequentially uses the beams 1 to n, and it measures the received powers of all the mn pairs of steering vectors at BS and UE. Then, the steering vector pair with the strongest power gain is fed back to the BS, and it is selected. More advanced beam searching methods can be found for example in [12].

B. Hybrid Analog-Digital Architecture and Signal Model

A hybrid analog-digital transceiver architecture is given in Fig. 3, where the digital precoding is omitted since it is usually performed after the baseband CE and it does not influence the baseband equivalent channel. Before the transmit antennas, analog precoding (transmit beamforming) is conducted to the signal from transmit RF chains. Similarly, after the receive antennas, analog combining (receive beamforming) is performed.

The baseband transmit and receive vectors are denoted as $x \in \mathbb{C}^{N_{RF,t}}$, $y \in \mathbb{C}^{N_{RF,r}}$ respectively. As shown in Fig. 3, the signal model can be expressed as

$$y = W_{RF}^{H} H F_{RF} x + W_{RF}^{H} n, \qquad (4)$$

where $F_{RF} = \begin{bmatrix} f_1 f_2 \cdots f_{N_{RF,t}} \end{bmatrix} \in \mathbb{C}^{N_t \times N_{RF,t}}$ and $W_{RF} = \begin{bmatrix} w_1 w_2 \cdots w_{N_{RF,r}} \end{bmatrix} \in \mathbb{C}^{N_r \times N_{RF,r}}$ are the analog precoding and combining matrices, respectively, and $n \sim \mathcal{CN}(\mathbf{0}_{N_r}, \sigma_n^2 I_{N_r})$ is the receiver complex additive white Gaussian noise (AWGN) vector. The elements of F_{RF} and W_{RF} meet the constant modulus constraints $|f_{i,j}| = \frac{1}{\sqrt{N_t}}$, for $1 \leq i \leq N_t$ and $1 \leq j \leq N_{RF,t}$, as well as $|w_{i,j}| = \frac{1}{\sqrt{N_r}}$, for $1 \leq i \leq N_r$ and $1 \leq j \leq N_{RF,r}$, where $f_{i,j}$ and $w_{i,j}$ denote the ith-row and jth-column elements of F_{RF} and W_{RF}, respectively. If W_{RF} is selected from a DFT codebook, $W_{RF}^{H} n$ remains to be an AWGN vector with the distribution $W_{RF}^{H} n \sim \mathcal{CN}(\mathbf{0}_{N_{RF,r}}, \sigma_n^2 I_{N_{RF,r}})$, since the DFT matrix is an unitary matrix.

From (4), the equivalent baseband channel after beam searching can be derived as

$$H_{BB} = W_{RF}^{H} H F_{RF} \in \mathbb{C}^{N_{RF,r} \times N_{RF,t}}. \qquad (5)$$

Since $N_{RF,t}$ and $N_{RF,r}$ are typically very small, the baseband channel matrix

$$H_{BB} = \begin{bmatrix} w_1^{H} H f_1 & w_1^{H} H f_2 & \cdots & w_1^{H} H f_{N_{RF,t}} \\ w_2^{H} H f_1 & w_2^{H} H f_2 & \cdots & w_2^{H} H f_{N_{RF,t}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N_{RF,r}}^{H} H f_1 & w_{N_{RF,r}}^{H} H f_2 & \cdots & w_{N_{RF,r}}^{H} H f_{N_{RF,t}} \end{bmatrix}, \qquad (6)$$

is small size. This is because owing to mmWave hardware and power consumption constraints, only very few RF chains can be deployed, especially at UE. Thus, $N_{RF,t}$ and $N_{RF,r}$ are typically very small. In the sequel, we will consider the case of $N_{RF,t} = N_{RF,r} = 1$. In this case, the precoding matrix F_{RF} reduces to the vector $f \in \mathbb{C}^{N_t}$, while the combining matrix W_{RF} reduces to the vector $w \in \mathbb{C}^{N_r}$. Hence, the baseband channel matrix H_{BB} becomes a scalar

$$H_{BB} = w^{H} H f. \qquad (7)$$

Note that it is sufficient to derive the baseband CE for the single-RF-chain case, since the proposed estimator can be adopted to estimate each element of H_{BB} in the same way.

III. PROBLEM FORMULATION IN OFDM SYSTEMS

We formulate the baseband CE after beam searching as a HR problem. We focus on single-user scenario, but our approach is equally applicable to the multi-user scenario.[1]

A. Multipath Model Problem Formulation

For OFDM systems in frequency-selective channels, the FD channel matrix varies with the frequency as depicted in (3). Let us consider the OFDM system with K subcarriers and a subcarrier spacing of Δf. Then according to (7) and (3), the baseband channel at the kth subcarrier is expressed as

$$H_{BB}(k) = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l w^{H} a_r(\theta_{r,l}) a_t^{H}(\theta_{t,l}) f e^{-j2\pi \tau_l k \Delta f}. \qquad (8)$$

Since the beam steering vectors w and f are determined in beam searching, they are fixed and known vectors in (8).

Let us define the equivalent path strength parameter β_l and the delay parameter Δ_l for the lth path respectively as

$$\beta_l = \sqrt{\frac{N_t N_r}{L}} \alpha_l w^{H} a_r(\theta_{r,l}) a_t^{H}(\theta_{t,l}) f, \qquad (9)$$

$$\Delta_l = -2\pi \tau_l \Delta f, \qquad (10)$$

---

[1]For the multi-user scenario, the channel associated with each UE can be independently estimated using the method proposed in this paper. This is because in downlink transmission, the system allocates different bandwidths, each consisting of several subcarriers, to different UEs, and each UE can estimate its channel via the pilot distributed on its own bandwidth. In uplink transmission, the BS can similarly estimate each UE's uplink channel.

which are independent of the subcarrier index k. Then the FD baseband channel at the subcarrier k can be written as

$$H_{BB}(k) = \sum_{l=1}^{L} \beta_l e^{\mathrm{j}\Delta_l k}, \ 0 \le k \le K - 1. \tag{11}$$

It is readily seen that by estimating β_l and Δ_l for $1 \le l \le L$, we can acquire the FD CSI of all the subcarriers. Further observe from (11) that $H_{BB}(k)$ is a superposition of the L harmonics (sinusoids) $e^{\mathrm{j}\Delta_l k}$ with the frequencies Δ_l for $1 \le l \le L$. Moreover, owing to the strong sparsity of the mmWave baseband channel, the number of paths L is extremely small and, therefore, we arrive at a very sparse complex HR problem to estimate β_l and Δ_l for $1 \le l \le L$.

Remark 1: Our proposed baseband CE approach can acquire all the FD coefficients for all the K subcarriers by estimating an extremely small number of parameters $\{\Delta_l, \beta_l\}_{l=1}^{L}$. This should be contrast with the conventional OFDM estimation approach based on interpolation, which needs to estimate all the FD coefficients $H_{BB}(k)$ for all the K subcarriers. Since K is very large, the number of parameters to be estimated is very large by the conventional method. Moreover, the conventional approach can only acquire $H_{BB}(k)$ for pilot subcarriers, and it has to use interpolation to estimate $H_{BB}(k)$ for other subcarriers. The FD pilot density must be sufficiently high to ensure a sufficient interpolation accuracy.

Compared with other existing method which estimates the channel matrix $\boldsymbol{H}$ before beam seaching, our approach also offers considerable advantage. In the context of CE before beam seaching, the FD channel matrix at the kth subcarrier can be expressed as

$$\boldsymbol{H}(k) = \sqrt{\frac{N_t N_r}{\bar{L}}} \sum_{l=1}^{\bar{L}} \alpha_l \boldsymbol{a}_r(\theta_{r,l}) \boldsymbol{a}_t^{\mathrm{H}}(\theta_{t,l}) e^{-\mathrm{j}2\pi \tau_l k \Delta f}$$

$$= \sqrt{\frac{N_t N_r}{\bar{L}}} \sum_{l=1}^{\bar{L}} \bar{\boldsymbol{H}}_l e^{\mathrm{j}\Delta_l k}, \ 0 \le k \le K - 1, \tag{12}$$

where we use $\bar{L}$ to denote the number of propagation paths before beam searching, and clearly $\bar{L} > L$. By estimating $\{\bar{\boldsymbol{H}}_l, \Delta_l\}_{l=1}^{\bar{L}}$, the FD CSI of all the subcarriers, i.e., $\boldsymbol{H}(k)$ for $0 \le k \le K - 1$, can be acquired. However, it is apparent that the parameters to be estimated in $\{\bar{\boldsymbol{H}}_l, \Delta_l\}_{l=1}^{\bar{L}}$, which might include the path gains, delays, AOAs and AODs of $\bar{L}$ paths, are much more than the parameters to be estimated in $\{\beta_l, \Delta_l\}_{l=1}^{L}$ of L paths. Consequently, our approach offers advantage in terms of training pilot requirement and computational complexity, compared with these existing methods.

B. Extension to Multicluster Model

Although the model (3) is widely adopted, it is a simplified model which represents the clusters of the mmWave channel by the paths. In fact, each cluster of the mmWave channel always consists of multiple rays or sub-paths with the distribution on delay, power, AOA and AOD [2], [34]. Therefore, a more

accurate channel model is the multi-cluster model [17]

$$\boldsymbol{H}(f) = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{l_1=1}^{N_{cl}} \sum_{l_2=1}^{N_{ray}} \alpha_{l_1,l_2} \boldsymbol{a}_r(\theta_{r,l_1,l_2})$$

$$\times \boldsymbol{a}_t^{\mathrm{H}}(\theta_{t,l_1,l_2}) e^{-\mathrm{j}2\pi \tau_{l_1,l_2} f}, \tag{13}$$

where N_{cl} is the number of clusters and N_{ray} is the number of rays in each cluster, α_{l_1,l_2} and τ_{l_1,l_2} are the complex gain and delay of the l_2th ray in the l_1th cluster, respectively, while θ_{r,l_1,l_2} and θ_{t,l_1,l_2} are the AOA and AOD of the l_2th ray in the l_1th cluster, respectively.

Even though N_{cl} is extremely small, the total number of rays $N_{cl} N_{ray}$ may be large, e.g., $N_{ray} = 20$ is specified in [34], and the sparsity property of paths may no longer be valid. However, we can still formulate a sparse HR problem with a reasonable assumption that the delay spread (DS) of the rays in each cluster can be neglected. Actually, the DS within each cluster is very small, e.g., in the 3GPP model [34], the DS within cluster in urban micro scenario is about 5 ns for the LOS and 11 ns for non-LOS. Moreover, the work [35] has verified that the DS will be even smaller after beam searching. Therefore, it is reasonable to neglect the DS within each cluster by assuming $\tau_{l_1,l_2} = \tau_{l_1}$ for $1 \le l_2 \le N_{ray}$. Thus, similar to (9) and (10) for the simplified model (3), we can define β_{l_1} and Δ_{l_1} for $1 \le l_1 \le N_{cl}$ as

$$\beta_{l_1} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{l_2=1}^{N_{ray}} \alpha_{l_1,l_2} \boldsymbol{w}^{\mathrm{H}} \boldsymbol{a}_r(\theta_{r,l_1,l_2}) \boldsymbol{a}_t^{\mathrm{H}}(\theta_{t,l_1,l_2}) \boldsymbol{f},$$
$$\tag{14}$$

$$\Delta_{l_1} = -2\pi \tau_{l_1} \Delta f. \tag{15}$$

It can be seen that in (14), the rays in each cluster are combined together because they share the same delay. Thus the baseband multi-cluster model (13) can be simplified to

$$H_{BB}(k) = \sum_{l_1=1}^{N_{cl}} \beta_{l_1} e^{\mathrm{j}\Delta_{l_1} k}, \tag{16}$$

which has an extremely small number of parameters to be estimated, since N_{cl} is extremely small, especially after beam searching. It is readily seen that (16) is identical to (11).

IV. PROPOSED CHANNEL ESTIMATION METHODS

The pilot transmission model is first introduced. Then, the ML estimation is formulated for the HR problem presented in Section III. We show that the OMP and DFT methods are capable of acquiring approximate ML estimates, but the latter requires that the frequencies of the harmonics are sufficiently separated. The window function is applied to further improve the accuracy of the OMP and DFT methods, and we derive our proposed window-OMP and window-DFT methods.

A. Pilot Transmission Model

Let us denote the indexes of the K subcarriers from $k = 0$ to $K - 1$. Assume that the pilot interval is K_p and the number of total pilot subcarriers is K_s. Let the minimum index of the pilot

subcarriers be k_0, where k_0 can be arbitrarily selected as long as $k_0 + (K_s - 1)K_p \leq (K - 1)$. Therefore, the indexes of all the pilot subcarriers can be written as

$$k_0 + k'K_p, \ k' = 0, 1, \ldots, (K_s - 1). \tag{17}$$

The received pilot signal of the k'th pilot subcarrier is

$$y(k_0 + k'K_p) = H_{BB}(k_0 + k'K_p) s + \boldsymbol{w}^{\mathrm{H}} \boldsymbol{n}(k_0 + k'K_p), \tag{18}$$

where s denotes the transmitted pilot signal with the unit modulus $|s| = 1$, and $\boldsymbol{n}(k_0 + k'K_p)$ is the AWGN vector at the subcarrier $k_0 + k'K_p$. Thus, the CSI of pilot subcarriers can be estimated based on the LS criterion as

$$\begin{aligned} \widehat{H}_{BB}(k_0 + k'K_p) &= \frac{y(k_0 + k'K_p)}{s} \\ &= H_{BB}(k_0 + k'K_p) + \boldsymbol{w}^{\mathrm{H}} \boldsymbol{n}(k_0 + k'K_p), \end{aligned} \tag{19}$$

where we have substituted $\frac{\boldsymbol{w}^{\mathrm{H}} \boldsymbol{n}(k)}{s}$ by $\boldsymbol{w}^{\mathrm{H}} \boldsymbol{n}(k)$, since both have the identical distribution.

Then, substituting (11) into (19) yields

$$\widehat{H}_{BB}(k_0 + k'K_p) = \sum_{l=1}^{L} \beta_l e^{\mathrm{j}\Delta_l (k_0 + k'K_p)} + \boldsymbol{w}^{\mathrm{H}} \boldsymbol{n}(k_0 + k'K_p). \tag{20}$$

It can be seen that $\widehat{H}_{BB}(k_0 + k'K_p)$ contains the sum of the L harmonics $\beta_l e^{\mathrm{j}\Delta_l (k_0 + k'K_p)}$, $1 \leq l \leq L$, and the lth harmonic has the complex gain $\beta_l e^{\mathrm{j}\Delta_l k_0}$ and the frequency $K_p \Delta_l$. Therefore, the baseband CE can be formulated as a HR problem.

B. Approximate ML Estimation Based on OMP Method

Collect the estimated CSI of the pilot subcarriers in a vector

$$\widehat{\boldsymbol{h}} = \left[\widehat{H}_{BB}(k_0) \cdots \widehat{H}_{BB}(k_0 + (K_s - 1)K_p) \right]^{\mathrm{T}} \in \mathbb{C}^{K_s}. \tag{21}$$

According to (20), $\widehat{\boldsymbol{h}}$ can be expressed as

$$\widehat{\boldsymbol{h}} = \boldsymbol{A}\boldsymbol{s} + \widetilde{\boldsymbol{n}}, \tag{22}$$

where the AWGN vector

$$\widetilde{\boldsymbol{n}} = \left[\boldsymbol{w}^{\mathrm{H}} \boldsymbol{n}(k_0) \cdots \boldsymbol{w}^{\mathrm{H}} \boldsymbol{n}(k_0 + (K_s - 1)K_p) \right]^{\mathrm{T}} \in \mathbb{C}^{K_s} \tag{23}$$

follows the distribution $\widetilde{\boldsymbol{n}} \sim \mathcal{CN}\left(\boldsymbol{0}_{K_s}, \sigma_n^2 \boldsymbol{I}_{K_s} \right)$, $\boldsymbol{A} \in \mathbb{C}^{K_s \times L}$ and $\boldsymbol{s} \in \mathbb{C}^L$ with

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{\mathrm{j}K_p \Delta_1} & e^{\mathrm{j}K_p \Delta_2} & \cdots & e^{\mathrm{j}K_p \Delta_L} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\mathrm{j}K_p \Delta_1 (K_s-1)} & e^{\mathrm{j}K_p \Delta_2 (K_s-1)} & \cdots & e^{\mathrm{j}K_p \Delta_L (K_s-1)} \end{bmatrix}, \tag{24}$$

$$\boldsymbol{s} = \left[\beta_1 e^{\mathrm{j}\Delta_1 k_0} \ \beta_2 e^{\mathrm{j}\Delta_2 k_0} \cdots \beta_L e^{\mathrm{j}\Delta_L k_0} \right]^{\mathrm{T}}. \tag{25}$$

Further define

$$\boldsymbol{\Delta} = \left[\widetilde{\Delta}_1 \cdots \widetilde{\Delta}_L \right]^{\mathrm{T}} = \left[K_p \Delta_1 \cdots K_p \Delta_L \right]^{\mathrm{T}} \tag{26}$$

to represent the frequencies of the harmonics. Then, $\boldsymbol{s}$, $\boldsymbol{\Delta}$ and L are the parameters to be estimated.

The probability density function (PDF) of $\widehat{\boldsymbol{h}}$ is

$$p(\widehat{\boldsymbol{h}}; \boldsymbol{s}, \boldsymbol{\Delta}, L) = \frac{1}{(\pi \sigma_n^2)^{K_s}} e^{-\frac{1}{\sigma_n^2} (\widehat{\boldsymbol{h}} - \boldsymbol{A}\boldsymbol{s})^{\mathrm{H}} (\widehat{\boldsymbol{h}} - \boldsymbol{A}\boldsymbol{s})}. \tag{27}$$

According to the ML estimation principle and the PDF (27), the ML estimator of $\boldsymbol{s}$, $\boldsymbol{\Delta}$ and L can be expressed as

$$\begin{aligned} \{\widehat{\boldsymbol{s}}, \widehat{\boldsymbol{\Delta}}, \widehat{L}\} &= \arg \max_{\{\boldsymbol{s}, \boldsymbol{\Delta}, L\}} p(\widehat{\boldsymbol{h}}; \boldsymbol{s}, \boldsymbol{\Delta}, L) \\ &= \arg \min_{\{\boldsymbol{s}, \boldsymbol{\Delta}, L\}} \|\widehat{\boldsymbol{h}} - \boldsymbol{A}\boldsymbol{s}\|_2^2. \end{aligned} \tag{28}$$

The number of the harmonics is assumed to be L_p-sparse, i.e., $L \leq L_p$. Note that the L columns of $\boldsymbol{A}$ can be chosen from the columns of a partial DFT matrix $\boldsymbol{\Phi} \in \mathbb{C}^{K_s \times N_{dic}}$, which is the first K_s rows of a N_{dic}-dimensional DFT matrix. Here N_{dic} is a sufficiently large integer to ensure that the quantization error of $\boldsymbol{\Delta}$ is sufficiently small. Thus, the solution of (28) can be derived from the following CS problem with $\boldsymbol{\Phi}$ as the dictionary matrix

$$\begin{aligned} \min_{\widetilde{\boldsymbol{s}} \in \mathbb{C}^{N_{dic}}, L} & \ \|\widehat{\boldsymbol{h}} - \boldsymbol{\Phi}\widetilde{\boldsymbol{s}}\|_2^2, \\ \text{s.t.} & \ \left| \mathrm{supp}(\widetilde{\boldsymbol{s}}) \right| \leq L_p. \end{aligned} \tag{29}$$

Here $\mathrm{supp}(\widetilde{\boldsymbol{s}})$ represents the support set of $\widetilde{\boldsymbol{s}}$, namely, the set of non-zero elements in $\widetilde{\boldsymbol{s}}$.

Since the partial DFT matrix satisfies the restricted isometry property (RIP) condition [36], the OMP algorithm is chosen to solve the CS problem (29) due to its high efficiency and simplicity. The detailed description of our proposed window-OMP method is given in Section IV-D.

C. Approximation ML Estimation Based on DFT Method

For the optimization problem (28), when the parameters $\{\boldsymbol{\Delta}, L\}$ are fixed, it becomes a linear LS problem. Hence, the optimal solution of $\boldsymbol{s}$ for given $\{\boldsymbol{\Delta}, L\}$ is

$$\boldsymbol{s}^{\star} = \boldsymbol{A}^{\dagger} \widehat{\boldsymbol{h}}, \tag{30}$$

where $\boldsymbol{A}^{\dagger} = (\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A})^{-1} \boldsymbol{A}^{\mathrm{H}}$ since $\boldsymbol{A}$ is a Vandermonde matrix with $\mathrm{rank}(\boldsymbol{A}) = L$. Therefore, using $\boldsymbol{s}^{\star}$ of (30) in the ML problem (28), it becomes equivalently to

$$\begin{aligned} \{\widehat{\boldsymbol{\Delta}}, \widehat{L}\} &= \arg \min_{\boldsymbol{\Delta}, L} \|\widehat{\boldsymbol{h}} - \boldsymbol{A}\boldsymbol{A}^{\dagger} \widehat{\boldsymbol{h}}\|_2^2 \\ &= \arg \max_{\boldsymbol{\Delta}, L} \left(\widehat{\boldsymbol{h}}^{\mathrm{H}} \boldsymbol{A} (\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A})^{-1} \boldsymbol{A}^{\mathrm{H}} \widehat{\boldsymbol{h}} \right). \end{aligned} \tag{31}$$

According to [28], when ω is not close to 0 and N is large, we have the following approximate equalities

$$\begin{cases} \frac{1}{N^{k+1}} \sum_{n=1}^{N} n^k \sin(\omega n + \phi) \approx 0, \ k \geq 0, \\ \frac{1}{N^{k+1}} \sum_{n=1}^{N} n^k \cos(\omega n + \phi) \approx 0, \ k \geq 0. \end{cases} \tag{32}$$

Consider the case where the frequencies in $\boldsymbol{\Delta}$ are sufficiently separated. In this case, based on (32) with $k = 0$, the element in the mth-row and nth-column of $\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}$ can be written as

$$\left[\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A} \right]_{m,n} = \sum_{k=0}^{K_s-1} e^{\mathrm{j}(\widetilde{\Delta}_n - \widetilde{\Delta}_m)} \approx \begin{cases} K_s, & m = n, \\ 0, & m \neq n. \end{cases} \tag{33}$$

Thus, $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ can be approximated as a diagonal matrix, i.e., $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} \approx K_s \boldsymbol{I}_L$. Therefore, under the condition that the frequencies in $\boldsymbol{\Delta}$ are sufficiently separated, the ML problem (31) can be approximated as

$$\left\{\widehat{\boldsymbol{\Delta}}, \widehat{L}\right\} \approx \arg\max_{\boldsymbol{\Delta}, L}\left(\widehat{\boldsymbol{h}}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\widehat{\boldsymbol{h}}\right) = \arg\max_{\boldsymbol{\Delta}, L}\left\|\boldsymbol{A}^{\mathrm{H}}\widehat{\boldsymbol{h}}\right\|_2^2, \tag{34}$$

where the elements in $\boldsymbol{A}^{\mathrm{H}}\widehat{\boldsymbol{h}}$ can be regarded as the elements chosen from vector $\boldsymbol{\Phi}^{\mathrm{H}}\widehat{\boldsymbol{h}}$, which is equivalent to the DFT of $\widehat{\boldsymbol{h}}$ after appending zeros, i.e., the DFT of

$$\widetilde{\boldsymbol{h}} = \left[\widehat{\boldsymbol{h}}^{\mathrm{T}}\ \boldsymbol{0}_{N_{dic}-K_s}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N_{dic}}. \tag{35}$$

Since $L \le L_p$, the optimal solution of (34) corresponds to the L_p maximum elements of $\left\|\boldsymbol{\Phi}^{\mathrm{H}}\widehat{\boldsymbol{h}}\right\|_2^2$. Hence, under the assumption of the sufficiently separated frequencies, the optimal solution of (34) is also corresponding to the L_p highest peaks of the frequency spectrum of $\widetilde{\boldsymbol{h}}$.

Therefore, it can be seen that the conventional DFT method can also be used to acquire an approximate ML estimate as long as the assumption of the sufficiently separated frequencies is satisfied. Typically, we can set $N_{dic} = 2^N$ and apply the fast Fourier transform (FFT) algorithm to reduce the complexity.

Remark 2: Since we usually have $L < L_p$, the number of the peaks in the frequency spectrum can be less than L_p. To estimate L correctly, a threshold should be set for peak searching. It can easily seen that the noise power in the DFT spectrum is upper bounded by $\sigma_n^2 K_s$ or equivalently the noise standard deviation (STD) is upper bounded by $\sigma_n \sqrt{K_s}$. We can set the threshold to the two STDs of the noise in the spectrum or 5% of the highest peak $p_{\max}$ in the spectrum, depending on which one is larger, that is,

$$threshold = \max\{\gamma \sigma_n \sqrt{K_s}, 0.05 p_{\max}\}, \tag{36}$$

where $\gamma = 1.5 < 2$ is used to take into account that $\sigma_n \sqrt{K_s}$ is an upper bound for the STD of the noise.

D. Proposed Windowed DFT and OMP Algorithms

In the DFT method, appending zeros to $\widehat{\boldsymbol{h}}$ to obtain the N_{dic}-length signal $\widetilde{\boldsymbol{h}}$ is equivalent to placing a rectangular window on the original N_{dic}-length signal, which leads to frequency leaking effect on the DFT spectrum. Hence, an appropriate window function is utilized to mitigate the side-lobe effect of appending zeros, and the signal before DFT can be written as

$$\widetilde{\boldsymbol{h}}_w = \left[\bar{w}(0)\widehat{H}_{BB}(k_0)\cdots\bar{w}(K_s-1)\widehat{H}_{BB}(k_0+(K_s-1)K_p)\right.$$
$$\left.0\cdots0\right]^{\mathrm{T}} \in \mathbb{C}^{N_{dic}}, \tag{37}$$

where $\bar{w}(\cdot)$ is the chosen window function. A comparison between the original DFT spectrum without windowing and the Hamming windowed DFT spectrum is shown in Fig. 4 for the case of the path number $L = 1$ and SNR $= 10\log_{10}(\frac{1}{\sigma_n^2}) = -10$ dB, where a 1024-point FFT is applied. The peak in the spectrum represents the single harmonic whose frequency is proportional to its DFT index. Observe that although the amplitude of the main lobe is weaker after adding the window,
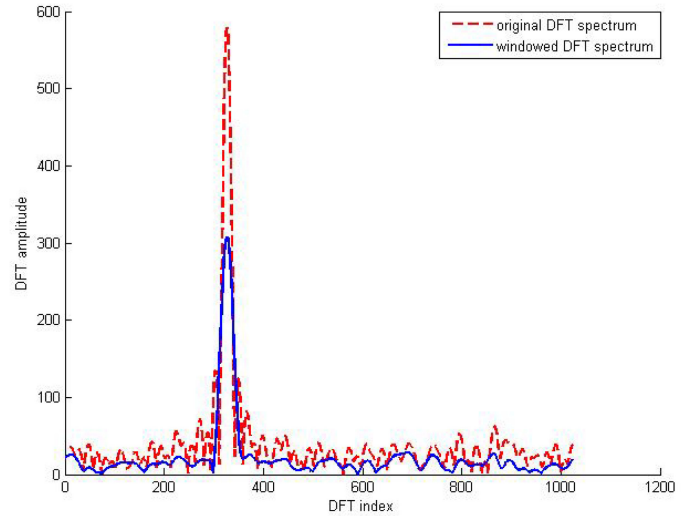

Fig. 4. Comparison of the original DFT and the Hamming windowed DFT for the case of $L = 1$ and SNR $= -10$ dB, where 1024-point FFT is applied.

Algorithm 1: Window-DFT Method.

Input: $\widehat{\boldsymbol{h}}$, K_s, N_{dic}, L_p, and σ_n;

1: Append zeros and apply window function to $\widehat{\boldsymbol{h}}$ to acquire $\widetilde{\boldsymbol{h}}_w$;
2: Perform N_{dic}-FFT on $\widetilde{\boldsymbol{h}}_w$;
3: Search the highest peak, record its FFT index $\widehat{k}_1$ and peak amplitude A_1; set the peak index set $\mathcal{K} = \{\widehat{k}_1\}$, path number $\widehat{L} = 1$; // FFT index is from 1 to N_{dic}
4: **for** $i = 2 : L_p$ **do**
5:     Search the ith highest peak, record its FFT index $\widehat{k}_i$ and peak amplitude A_i;
6:     **if** $A_i > \max\{1.5\sigma_n\sqrt{K_s}, 0.05A_1\}$ **then**
7:         $\widehat{L} = \widehat{L} + 1$, $\mathcal{K} = \mathcal{K}\bigcup\{\widehat{k}_i\}$;
8:     **else**
9:         Break;
10:     **end if**
11: **end for**
12: $\widehat{\widetilde{\Delta}}_l = 2\pi\frac{(\widehat{k}_i-1)}{N_{dic}}$, $\widehat{k}_l \in \mathcal{K}$, $l = 1, \ldots, \widehat{L}$; $\widehat{\boldsymbol{s}} = \widehat{\boldsymbol{A}}^{\dagger}\widehat{\boldsymbol{h}}$;
13: $\widehat{\Delta}_l = \frac{(\widehat{\widetilde{\Delta}}_l - 2\pi)}{K_p}$, $\widehat{\beta}_l = \widehat{s}_l e^{-\mathrm{j}\widehat{\Delta}_l k_0}$, $l = 1, \ldots, \widehat{L}$;

---

the side-lobes are significantly suppressed. Reducing the frequency leaking is particularly important in the case of multiple harmonics to suppress the effect of one harmonic on other harmonics. The proposed window-DFT method is summarized in Algorithm 1.

Similarly, the proposed window-OMP method is summarized in Algorithm 2. Different from the original OMP method, a window function is added to the residual vector $\boldsymbol{r}$ before the matching step (lines 4-5). In the matching step of the original OMP, the inner product vector of the dictionary matrix and residual vector is computed as $\boldsymbol{c} = \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{r}$, and the index corresponding to the element with the maximum amplitude in $\boldsymbol{c}$ is picked. It is obvious that this matching step is equivalent to finding the

---

Algorithm 2: Window-OMP Method.

Input: $\widehat{h}$, K_s, N_{dic}, and L_p;

1: Generate dictionary matrix $\boldsymbol{\Phi}$; set peak index set $\mathcal{K} = \emptyset$, path number $\widehat{L} = 0$, residual vector $\boldsymbol{r} = \widehat{h}$;
2: **for** $i = 1 : L_p$ **do**
3:     Apply window function to $\boldsymbol{r}$ to acquire $\boldsymbol{r}_w$;
4:     $\boldsymbol{c} = \boldsymbol{\Phi}^{\mathrm{H}} \boldsymbol{r}_w$;
5:     $\widehat{k}_i = \arg\max_k\{|c_k|\}$, $\mathcal{K}_{new} = \mathcal{K} \bigcup \{\widehat{k}_i\}$; // c_k is the kth element of $\boldsymbol{c}$.
6:     $\widehat{s} = \boldsymbol{\Phi}^{\dagger}_{\mathcal{K}_{new}} \boldsymbol{r}$, $\boldsymbol{r}_{new} = \widehat{h} - \boldsymbol{\Phi}_{\mathcal{K}_{new}} \widehat{s}$;
7:     **if** $\|\boldsymbol{r}\|_2^2 - \|\boldsymbol{r}_{new}\|_2^2 > 0.05\|\boldsymbol{r}\|_2^2$ **then**
8:         $\boldsymbol{r} = \boldsymbol{r}_{new}, \mathcal{K} = \mathcal{K}_{new}, \widehat{L} = \widehat{L} + 1$;
9:     **else**
10:        Break;
11:     **end if**
12: **end for**
13: $\widehat{\widetilde{\Delta}}_l = 2\pi\frac{(\widehat{k}_i - 1)}{N_{dic}}$, $\widehat{k}_l \in \mathcal{K}$, $l = 1, \ldots, \widehat{L}$; $\widehat{s} = \widehat{A}^{\dagger}\widehat{h}$;
14: $\widehat{\Delta}_l = \frac{(\widehat{\widetilde{\Delta}}_l - 2\pi)}{K_p}$, $\widehat{\beta}_l = \widehat{s}_l e^{-j\widehat{\Delta}_l k_0}$, $l = 1, \ldots, \widehat{L}$;

---

highest peak in the frequency spectrum of $\boldsymbol{r}$ after appending zeros. Hence, a window function can similarly be utilized to mitigate the side-lobe effect, and the windowed residual vector is given by

$$\boldsymbol{r}_w = \left[\bar{w}(0)r_1 \cdots \bar{w}(K_s - 1)r_{K_s}\right]^{\mathrm{T}} \in \mathbb{C}^{K_s}, \quad (38)$$

where r_k is the kth element in $\boldsymbol{r}$. In addition, in line 6,

$$\boldsymbol{\Phi}_{\mathcal{K}_{new}} = \left[\boldsymbol{\phi}_{\widehat{k}_1} \cdots \boldsymbol{\phi}_{\widehat{k}_{|\mathcal{K}_{new}|}}\right], \quad (39)$$

with $\boldsymbol{\phi}_k$ denoting the kth column of $\boldsymbol{\Phi}$. Also as depicted in line 7, the stop criterion for the window-OMP algorithm is that the reduction of the l_2-norm of the residual vector is less than 5% after an iteration.

E. Pilot Density Comparison With Conventional Interpolation

In our proposed method, the estimated $\widehat{\Delta}_l$ is in the range of $\left[-\frac{2\pi}{K_p}, 0\right]$. Then according to (15), we have

$$K_p\Delta f \leq \frac{1}{\tau_l}, \quad (40)$$

for each path. This indicates that the pilot interval $K_p\Delta f$ [Hz] should not exceed $\frac{1}{\sigma_{\max}}$ in our method, where $\sigma_{\max}$ is the maximum DS of the channel. By contrast, the K_p of the conventional interpolation method must be sufficiently small to guarantee the accuracy of the CSI interpolation. Furthermore, the pilot subcarriers in the conventional interpolation method must be distributed over the entire OFDM band of the K subcarriers, while our method does not need this criterion.

More specifically, in the cell-specific reference signal of LTE-Advanced [37], where the interpolation method is adopted for the CE of OFDM systems, the pilot interval $K_p = 3$ and the subcarrier spacing $\Delta f = 15$ kHz are specified. Thus the pilot interval in Hz is $K_p\Delta f = 45$ kHz for the LTE-Advanced

pilot pattern. Typically, the maximum DS $\sigma_{\max}$ of the channel is not more than 5 times of the root-mean-square DS σ_{rms} of the channel, i.e., $\sigma_{\max} \leq 5\sigma_{\mathrm{rms}}$. Without loss of generality, consider the urban macro scenario [34], where $\sigma_{\mathrm{rms}} = 650$ ns. The corresponding pilot interval in Hz for our proposed method can be as large as $\frac{1}{\sigma_{\max}} \geq 308$ kHz. Hence, compared to the LTE-Advanced pilot pattern for the conventional method, the pilot pattern for our proposed method can be more than 6 times sparser.

V. THEORETICAL ANALYSIS

We first derive the closed-form approximate CRLB for the unbiased estimator of the parameters in (22) given L. Then, we show that the effect of wrong estimated paths (harmonics) is very limited by projection analysis of the LS method.

A. CRLB

In (22), the elements of the complex parameter vector $\boldsymbol{s}$ can be expressed as $s_l = \nu_l + j\gamma_l$, $1 \leq l \leq L$, where ν_l and γ_l are the real and imaginary parts of s_l, respectively, and L is a known number. Then, the parameters to be estimated can be written in the vector form

$$\boldsymbol{\theta} = \left[\boldsymbol{\nu}^{\mathrm{T}} \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\Delta}^{\mathrm{T}}\right]^{\mathrm{T}} = \left[\theta_1 \cdots \theta_{3L}\right]^{\mathrm{T}} \in \mathbb{R}^{3L}, \quad (41)$$

where $\boldsymbol{\nu} = [\nu_1 \cdots \nu_L]^{\mathrm{T}}$ and $\boldsymbol{\gamma} = [\gamma_1 \cdots \gamma_L]^{\mathrm{T}}$. Denote the expectation of $\widehat{h}$ by $\boldsymbol{\mu} = E[\widehat{h}] = \boldsymbol{A}\boldsymbol{s}$. Then, we have

$$\frac{\partial\boldsymbol{\mu}}{\partial\nu_k} = \boldsymbol{a}_k, \ 1 \leq k \leq L, \quad (42)$$

$$\frac{\partial\boldsymbol{\mu}}{\partial\gamma_k} = j\boldsymbol{a}_k, \ 1 \leq k \leq L, \quad (43)$$

$$\frac{\partial\boldsymbol{\mu}}{\partial\widetilde{\Delta}_k} = j(\nu_k + j\gamma_k)\boldsymbol{\Lambda}\boldsymbol{a}_k, \ 1 \leq k \leq L, \quad (44)$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\{0, 1, \ldots, K_s - 1\}$ and $\boldsymbol{a}_k$ is the kth column of $\boldsymbol{A}$. The pth-row and qth-column element of Fisher information matrix $\boldsymbol{F}(\boldsymbol{\theta}) \in \mathbb{R}^{3L \times 3L}$ is given by [28]

$$[\boldsymbol{F}(\boldsymbol{\theta})]_{p,q} = \frac{2}{\sigma_n^2}\Re\left\{\frac{\partial\boldsymbol{\mu}^{\mathrm{H}}}{\partial\theta_p}\frac{\partial\boldsymbol{\mu}}{\partial\theta_q}\right\}, \quad (45)$$

where $\Re\{\cdot\}$ denotes the real part. Partition $\boldsymbol{F}(\boldsymbol{\theta})$ into the 9 $L \times L$ sub-matrices as

$$\boldsymbol{F}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\nu},\boldsymbol{\nu}} & \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\nu},\boldsymbol{\gamma}} & \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\nu},\boldsymbol{\Delta}} \\ \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\gamma},\boldsymbol{\nu}} & \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\gamma},\boldsymbol{\gamma}} & \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\gamma},\boldsymbol{\Delta}} \\ \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\Delta},\boldsymbol{\nu}} & \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\Delta},\boldsymbol{\gamma}} & \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\Delta},\boldsymbol{\Delta}} \end{bmatrix}. \quad (46)$$

Note that $\boldsymbol{F}^{\mathrm{T}}(\boldsymbol{\theta}) = \boldsymbol{F}(\boldsymbol{\theta})$. According to (42) to (45) and based on (32), we can derive the following approximations

$$\boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\nu},\boldsymbol{\nu}} \approx \frac{2K_s}{\sigma_n^2}\boldsymbol{I}_L, \quad (47)$$

$$\boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\gamma},\boldsymbol{\gamma}} \approx \frac{2K_s}{\sigma_n^2}\boldsymbol{I}_L, \quad (48)$$

$$\boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\Delta},\boldsymbol{\Delta}} \approx \frac{(K_s - 1)K_s(2K_s - 1)}{3\sigma_n^2}\mathrm{diag}\{\nu_l^2 + \gamma_l^2\}_{l=1}^L, \quad (49)$$

$$\boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\nu},\boldsymbol{\gamma}} = \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\gamma},\boldsymbol{\nu}}^{\mathrm{T}} \approx \boldsymbol{0}_{L \times L}, \quad (50)$$

$$\boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\nu},\boldsymbol{\Delta}} = \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\Delta},\boldsymbol{\nu}}^{\mathrm{T}} \approx \frac{(K_s - 1)K_s}{\sigma_n^2}\mathrm{diag}\{-\gamma_l\}_{l=1}^L, \quad (51)$$

$$\boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\gamma},\boldsymbol{\Delta}} = \boldsymbol{F}(\boldsymbol{\theta})_{\boldsymbol{\Delta},\boldsymbol{\gamma}}^{\mathrm{T}} \approx \frac{(K_s - 1)K_s}{\sigma_n^2}\mathrm{diag}\{\nu_l\}_{l=1}^L. \quad (52)$$

All the sub-matrices in $\boldsymbol{F}(\boldsymbol{\theta})$ are approximately diagonal matrices.

The inversion of $\boldsymbol{F}(\boldsymbol{\theta})$ can be used to derive the following closed-form approximate CRLBs for the corresponding parameters $\{\nu_l, \gamma_l, \widetilde{\Delta}_l\}_{l=1}^L$

$$\mathrm{CRLB}(\nu_l) = \left[\boldsymbol{F}^{-1}(\boldsymbol{\theta})\right]_{l,l} \approx \frac{\sigma_n^2}{K_s}\left(\frac{1}{2} + \frac{(K_s - 1)\gamma_l^2}{4d_l}\right), \quad (53)$$

$$\mathrm{CRLB}(\gamma_l) = \left[\boldsymbol{F}^{-1}(\boldsymbol{\theta})\right]_{L+l,L+l} \approx \frac{\sigma_n^2}{K_s}\left(\frac{1}{2} + \frac{(K_s - 1)\nu_l^2}{4d_l}\right), \quad (54)$$

$$\mathrm{CRLB}(\widetilde{\Delta}_l) = \left[\boldsymbol{F}^{-1}(\boldsymbol{\theta})\right]_{2L+l,2L+l} \approx \frac{\sigma_n^2}{K_s(K_s - 1)d_l}, \quad (55)$$

where

$$d_l = \frac{(K_s + 1)(\nu_l^2 + \gamma_l^2)}{6}. \quad (56)$$

The CRLB represents the minimum achievable mean square error (MSE) of unbiased estimator, and it will be used as the ultimate performance benchmark in our simulation to compare the performance of various estimators.

B. Effect of Wrong Estimated Paths

It is important to set an appropriate threshold for the window-DFT method. A high threshold may cause the problem of some actual paths been undetected, while a low threshold value may result in some wrong estimated paths. An example for wrong estimated path in the window-DFT method is illustrated in Fig. 5, where the wrong estimated path can be caused by the spectrum noise. The same problem also exists in the window-OMP method, and an appropriate choice of the stopping threshold is desired.

We will show that the effect of wrong paths is not critical. Therefore, it is reasonable to set a relatively low threshold in the two methods to reduce the probability of undetected paths.

First consider the estimation error without wrong path. In this case, the signal matrix $\boldsymbol{A}$ includes the L correct paths with $\widetilde{\Delta}_l$, $1 \leq l \leq L$. As will be confirmed in the simulation results that the estimation error of $\widetilde{\Delta}_l$ is small. Therefore, we assume that the delay parameter estimation is perfect, i.e., $\widehat{\boldsymbol{A}} = \boldsymbol{A}$. In both the window-OMP and window-DFT methods, $\widehat{\boldsymbol{s}}$ is estimated in a LS manner

$$\widehat{\boldsymbol{s}} = \boldsymbol{A}^\dagger\widehat{\boldsymbol{h}} = \boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{s} + \boldsymbol{A}^\dagger\widetilde{\boldsymbol{n}} = \boldsymbol{s} + \boldsymbol{A}^\dagger\widetilde{\boldsymbol{n}}, \quad (57)$$

and the estimation error is given by

$$\Delta\boldsymbol{h} = \boldsymbol{A}\widehat{\boldsymbol{s}} - \boldsymbol{A}\boldsymbol{s} = \boldsymbol{P}_{\boldsymbol{A}}\widetilde{\boldsymbol{n}}, \quad (58)$$
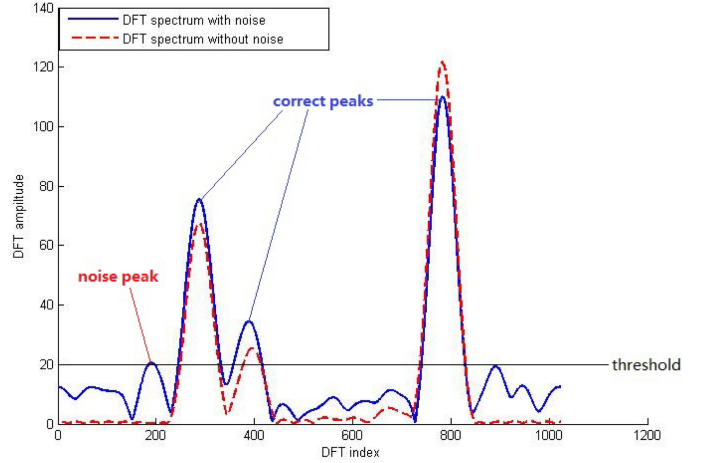

Fig. 5. An example of the appearance of the wrong path caused by the noise in the window-DFT method with SNR = −10 dB.

where $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{A}^\dagger$ is the projection matrix onto range($\boldsymbol{A}$).

Next, assume that we have $\widetilde{L}$ wrong paths, and denote the signal matrix associated with these $\widetilde{L}$ wrong paths by $\widetilde{\boldsymbol{B}} \in \mathbb{C}^{K_s \times \widetilde{L}}$. Then the system's signal matrix becomes $\bar{\boldsymbol{B}} = [\boldsymbol{A}\,\widetilde{\boldsymbol{B}}] \in \mathbb{C}^{K_s \times (L+\widetilde{L})}$, and the true $\widetilde{\boldsymbol{s}} = [\boldsymbol{s}^{\mathrm{T}}\,\boldsymbol{0}_{\widetilde{L}}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{(L+\widetilde{L})}$ is estimated by

$$\widehat{\widetilde{\boldsymbol{s}}} = \bar{\boldsymbol{B}}^\dagger\widehat{\boldsymbol{h}} = \bar{\boldsymbol{B}}^\dagger\boldsymbol{A}\boldsymbol{s} + \bar{\boldsymbol{B}}^\dagger\widetilde{\boldsymbol{n}}, \quad (59)$$

in which the pseudo-inverse of $\bar{\boldsymbol{B}}$ is given by

$$\bar{\boldsymbol{B}}^\dagger = \left(\bar{\boldsymbol{B}}^{\mathrm{H}}\bar{\boldsymbol{B}}\right)^{-1}\bar{\boldsymbol{B}}^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{H}}\widetilde{\boldsymbol{B}} \\ \widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{A} & \widetilde{\boldsymbol{B}}^{\mathrm{H}}\widetilde{\boldsymbol{B}} \end{bmatrix}^{-1}\bar{\boldsymbol{B}}^{\mathrm{H}}$$

$$= \begin{bmatrix} \boldsymbol{D} & -\boldsymbol{A}^\dagger\widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1} \\ -\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\left(\boldsymbol{A}^\dagger\right)^{\mathrm{H}} & \boldsymbol{C}^{-1} \end{bmatrix}\bar{\boldsymbol{B}}^{\mathrm{H}}, \quad (60)$$

where

$$\boldsymbol{C} = \widetilde{\boldsymbol{B}}^{\mathrm{H}}\left(\boldsymbol{I}_{K_s} - \boldsymbol{P}_{\boldsymbol{A}}\right)\widetilde{\boldsymbol{B}} = \widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{P}_{\boldsymbol{A}^\perp}\widetilde{\boldsymbol{B}}, \quad (61)$$

$$\boldsymbol{D} = \left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)^{-1} + \boldsymbol{A}^\dagger\widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\left(\boldsymbol{A}^\dagger\right)^{\mathrm{H}}, \quad (62)$$

and $\boldsymbol{P}_{\boldsymbol{A}^\perp} = \boldsymbol{I}_{K_s} - \boldsymbol{P}_{\boldsymbol{A}}$ is the projection matrix onto range $(\boldsymbol{A})^\perp$. Therefore, $\bar{\boldsymbol{B}}^\dagger\boldsymbol{A}\boldsymbol{s}$ in the right side of (59) is

$$\bar{\boldsymbol{B}}^\dagger\boldsymbol{A}\boldsymbol{s} = \begin{bmatrix} \boldsymbol{s} + \boldsymbol{A}^\dagger\widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{s} - \boldsymbol{A}^\dagger\widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{s} \\ -\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{s} + \boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{s} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{s} \\ \boldsymbol{0}_{\widetilde{L}} \end{bmatrix} = \widetilde{\boldsymbol{s}}, \quad (63)$$

which indicates that the first term in (59) is an accurate estimate of $\widetilde{\boldsymbol{s}}$. Then, the estimated CSI is given by

$$\widehat{\bar{\boldsymbol{h}}} = \bar{\boldsymbol{B}}\widehat{\widetilde{\boldsymbol{s}}} = \boldsymbol{A}\boldsymbol{s} + \bar{\boldsymbol{B}}\bar{\boldsymbol{B}}^\dagger\widetilde{\boldsymbol{n}}. \quad (64)$$

From (60) as well as noting that $\boldsymbol{P_A} = \boldsymbol{AA}^\dagger$ and $\boldsymbol{P_{A^\perp}}$ are Hermitian matrices, we have

$$\boldsymbol{P_{\bar{B}}} = \boldsymbol{\bar{B}}\boldsymbol{\bar{B}}^\dagger = \boldsymbol{P_A} + \boldsymbol{P_A}\widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{P_A} - \widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{P_A}$$
$$- \boldsymbol{P_A}\widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}} + \widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}$$
$$= \boldsymbol{P_A} + \boldsymbol{P_{A^\perp}}\widetilde{\boldsymbol{B}}\boldsymbol{C}^{-1}\widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{P_{A^\perp}}. \tag{65}$$

Further define $\boldsymbol{G} = \boldsymbol{P_{A^\perp}}\widetilde{\boldsymbol{B}}$, which is the projection of $\widetilde{\boldsymbol{B}}$ onto range$(\boldsymbol{A})^\perp$. Note that $\boldsymbol{P_{A^\perp}}$ is also an idempotent matrix, i.e., $\boldsymbol{P_{A^\perp}^2} = \boldsymbol{P_{A^\perp}}$. Hence $\boldsymbol{C}$ can be expressed as

$$\boldsymbol{C} = \widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{P_{A^\perp}}\widetilde{\boldsymbol{B}} = \widetilde{\boldsymbol{B}}^{\mathrm{H}}\boldsymbol{P_{A^\perp}^{\mathrm{H}}}\boldsymbol{P_{A^\perp}}\widetilde{\boldsymbol{B}} = \boldsymbol{G}^{\mathrm{H}}\boldsymbol{G}, \tag{66}$$

and (65) becomes

$$\boldsymbol{P_{\bar{B}}} = \boldsymbol{P_A} + \boldsymbol{G}(\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{H}} = \boldsymbol{P_A} + \boldsymbol{P_G}. \tag{67}$$

Therefore, the estimation error with the $\widetilde{L}$ wrong paths is given by

$$\Delta\widetilde{\boldsymbol{h}} = \widehat{\widetilde{\boldsymbol{h}}} - \boldsymbol{As} = \boldsymbol{P_A}\widetilde{\boldsymbol{n}} + \boldsymbol{P_G}\widetilde{\boldsymbol{n}}. \tag{68}$$

According to (58), the MSE without wrong path can be derived as

$$\mathrm{MSE}_{\boldsymbol{h}} = E\left[\|\boldsymbol{P_A}\widetilde{\boldsymbol{n}}\|_2^2\right] = E\left[\widetilde{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{P_A^{\mathrm{H}}}\boldsymbol{P_A}\widetilde{\boldsymbol{n}}\right]$$
$$= E\left[\widetilde{\boldsymbol{n}}^{\mathrm{H}}\boldsymbol{P_A}\widetilde{\boldsymbol{n}}\right] = \sigma_n^2\mathrm{tr}(\boldsymbol{P_A}) = \sigma_n^2 L, \tag{69}$$

where $\mathrm{tr}(\boldsymbol{P_A})$ is equal to the sum of the eigenvalues of $\boldsymbol{P_A}$ and we have $\mathrm{rank}(\boldsymbol{P_A}) = \mathrm{tr}(\boldsymbol{P_A})$ since all of the eigenvalues of $\boldsymbol{P_A}$ are 0 or 1.

Similarly, from (68), the MSE with the $\widetilde{L}$ wrong paths is given by

$$\mathrm{MSE}_{\widetilde{\boldsymbol{h}}} = E\left[\|\boldsymbol{P_{\bar{B}}}\widetilde{\boldsymbol{n}}\|_2^2\right] = E\left[\|\boldsymbol{P_A}\widetilde{\boldsymbol{n}}\|_2^2 + \|\boldsymbol{P_G}\widetilde{\boldsymbol{n}}\|_2^2\right]$$
$$= \sigma_n^2\left(\mathrm{tr}(\boldsymbol{P_A}) + \mathrm{tr}(\boldsymbol{P_G})\right) = \sigma_n^2(L + \widetilde{L}), \tag{70}$$

where the orthogonality between $\boldsymbol{P_A}$ and $\boldsymbol{P_G}$ are utilized and we have $\mathrm{tr}(\boldsymbol{P_G}) = \mathrm{rank}(\boldsymbol{P_G}) = \widetilde{L}$.

Comparing (69) with (70), it can be seen that the extra MSE resulting from the $\widetilde{L}$ wrong paths is $\sigma_n^2\widetilde{L}$, which is very small since $\widetilde{L}$ is typically very small. From another perspective, the extra error $\boldsymbol{P_G}\widetilde{\boldsymbol{n}}$ can be interpreted as the projection of the noise vector $\widetilde{\boldsymbol{n}}$ onto the range of $\boldsymbol{G}$. Since range$(\boldsymbol{G}) = $ range$(\boldsymbol{P_{A^\perp}}\widetilde{\boldsymbol{B}})$ with $\dim(\mathrm{range}(\boldsymbol{G})) = \mathrm{rank}(\widetilde{\boldsymbol{B}})$, the range of $\boldsymbol{G}$ is a low dimensional space, especially when the number of the wrong paths is small. This indicates that the l_2-norm of the projection of the noise vector is also likely to be small.

VI. SIMULATION STUDY

Monte Carlo simulations are conducted to evaluate the performance of our proposed channel estimation methods. We simulate a mmWave-based massive MIMO-OFDM system with $N_t = 128$ transmit antennas and $N_r = 16$ receive antennas. The carrier frequency is 30 GHz and the system bandwidth
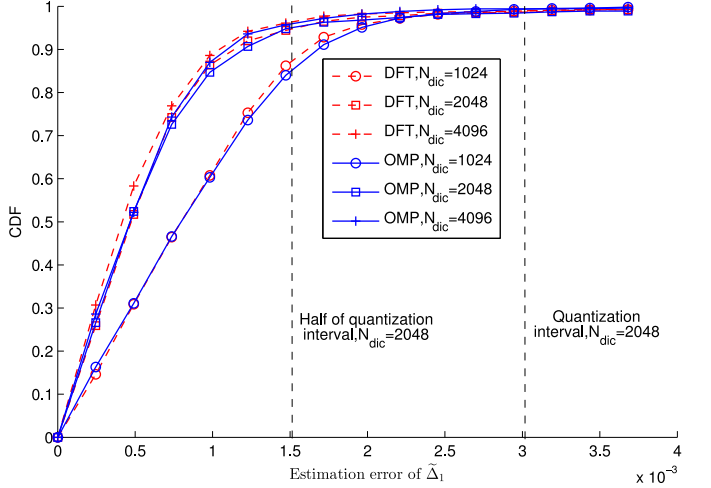

Fig. 6. CDFs of $\left|\widehat{\widetilde{\Delta}}_1 - \widetilde{\Delta}_1\right|$ in the single path senario for different dictionary sizes with SNR$= -10$ dB, obtained by window-DFT and window-OMP.

is 250 MHz which is divided into $K = 1024$ OFDM subcarriers. Unless otherwise specifically stated, the numbers of RF chains at transmitter and receiver are both set to 1. Therefore, the beam steering vectors are selected based on the maximum power gain criterion:

$$\{\boldsymbol{f}^\star, \boldsymbol{w}^\star\} = \arg\max_{\boldsymbol{f},\boldsymbol{w}}\left|\boldsymbol{w}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{f}\right|, \tag{71}$$

where the transmitter and receiver beam steering vectors $\boldsymbol{f}$ and $\boldsymbol{w}$ are searched from the 128-DFT codebook and the 16-DFT codebook, respectively. The channel model (3) is used to generate the MIMO channel matrix. The ESPRIT method proposed in [32] and the conventional interpolation method of [27] as well as the CRLB benchmark are used to compare with our proposed window-OMP and window-DFT methods. For the window-DFT, window-OMP and ESPRIT methods, the pilot number is $K_s = 64$ and the pilot interval is $K_p = 16$.

A. Selection of Dictionary Size N_{dic}

We first investigate the influence of the dictionary size N_{dic} to the harmonic frequency estimation error of the proposed window-DFT and window-OMP methods. The single path senario is considered, and the path delay is uniformly distributed in the range of 50 ns to 200 ns. With the sparse level $L_p = 1$ and three dictionary sizes of $N_{dic} = 1024$, 2048 and 4096, respectively, the cumulative distribution functions (CDFs) of the single path harmonic frequency error, i.e., $\left|\widehat{\widetilde{\Delta}}_1 - \widetilde{\Delta}_1\right|$, obtained by the proposed window-DFT and window-OMP, are depicted in Fig. 6.

In the single-path senario with $L_p = 1$, the window-OMP method performs only one iteration to detect the path and, therefore, it is equivalent to the window-DFT method. This is confirmed by the fact that the CDFs of the window-DFT and window-OMP in Fig. 6 are very similar given the same dictionary size. For the both methods, the estimation error of
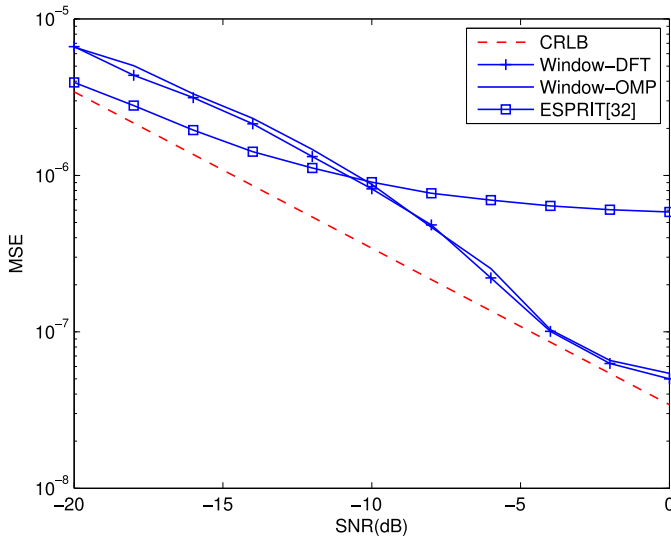
Fig. 7.   MSE of the harmonic frequency estimate for the strongest path.

Fig. 8.   MSE of the harmonic frequency estimate for the weakest path.

$N_{dic} = 2048$ is much lower than that of $N_{dic} = 1024$, which indicates that the resolution to the harmonic frequency of the dictionary matrix with $N_{dic} = 1024$ is insufficient. Furthermore, the estimation error of $N_{dic} = 4096$ is similar to that of $N_{dic} = 2048$. Therefore, it is reasonable to set the dictionary size to $N_{dic} = 2048$.

Note that the quantization interval of the harmonic frequency in the dictionary matrix of $N_{dic} = 2048$ is $\frac{2\pi}{2048} \approx 0.003$. The quantization interval and half quantization interval are marked in Fig. 6, where it can be seen that around 95% of the estimation error is less than half of the interval and almost 100% of the estimation error is less than the quantization interval. This verifies that the harmonic frequency error is typically very small and within the quantization interval.

B. MSE of Harmonic Frequencies Estimation

The number of paths is set to $L = 3$ and it is known. The AOAs and AODs of the three paths are inside the same transmit beam and receive beam. The path delays of the three paths are set to 50 ns, 125 ns and 200 ns so that they are sufficiently separated. The strength parameter of each path is randomly generated according to the distribution $\mathcal{CN}(0, 1)$. The dictionary size $N_{dic} = 2048$ and the sparse level $L_p = 3$ are adopted for both the window-DFT and window-OMP methods. In Figs. 7 and 8, the MSEs of the harmonic frequencies of the strongest path and the weakest path are depicted, respectively, for the window-DFT, window-OMP and existing ESPRIT methods together with the corresponding CRLBs.

For the case of the strongest path shown in Fig. 7, the ESPRIT method achieves a lower MSE than our two proposed methods at the low SNR region, specifically, for SNR < 10 dB. However, at the high SNR region, specifically, for SNR > 10 dB, our two proposed methods outperform the ESPRIT method significantly and, moreover, the MSEs of our window-DFT and window-OMP are very closed to the CRLB when the SNR is
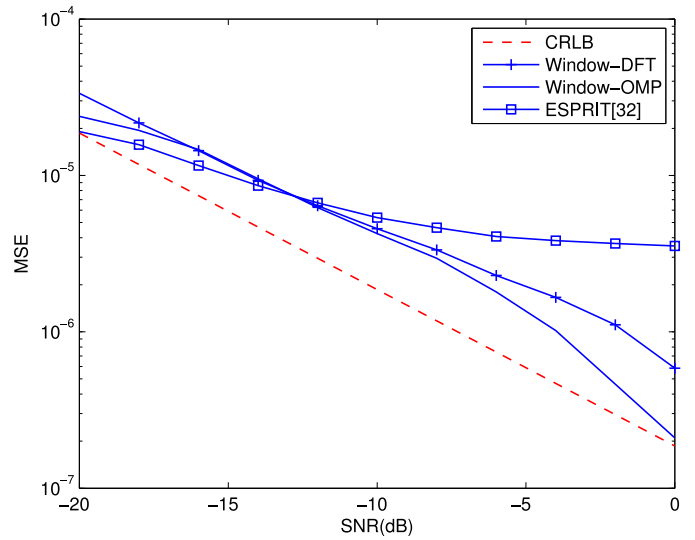
sufficiently high. From Fig. 7, it can also be seen that the performance of the two proposed methods are very similar. This is because the frequency estimation of the strongest harmonic in the window-OMP is performed in its first iteration, which is equivalent to a spectrum peak searching in the window-DFT method.

For the weakest path illustrated in Fig. 8, the ESPRIT method also performs slightly better at the low SNR region, while our two proposed methods are considerably better at the high SNR region. From Fig. 8, it can be seen that at the high SNR region, the window-OMP clearly achieves a lower MSE than the window-DFT. This is because the interference from the strong path to the weak path is canceled in each iteration of the window-OMP method. By contrast, the harmonic frequencies corresponding to the different peaks are estimated simultaneously in the window-DFT method. However, the complexity of the window-OMP is higher than that of the window-DFT method.

C. NMSE of CE for all Subcarriers

Since the beamwidth can be very narrow for the large antenna array, the effective path number after beam searching is usually only $L = 1$. Therefore, we consider a relatively rich-scattering transmission environment in mmWave with a total of 20 paths before beam searching to increase the probability of $L > 1$ after beam searching. Realistically, we consider L as an unknown parameter to be estimated. The AOA and AOD of each path are uniformly distributed in the interval $[0, 2\pi]$, and the path delay of each path is uniformly distributed in the interval of 50 ns to 200 ns. Therefore, the assumption of sufficiently separated frequencies does not hold. The normalized MSE (NMSE) of the CE for all the subcarriers is defined by

$$\text{NMSE} = \frac{\sum_{k=1}^{K} \left\| \boldsymbol{H}_{BB}(k) - \widehat{\boldsymbol{H}}_{BB}(k) \right\|^2}{\sum_{k=1}^{K} \left\| \boldsymbol{H}_{BB}(k) \right\|^2}. \qquad (72)$$
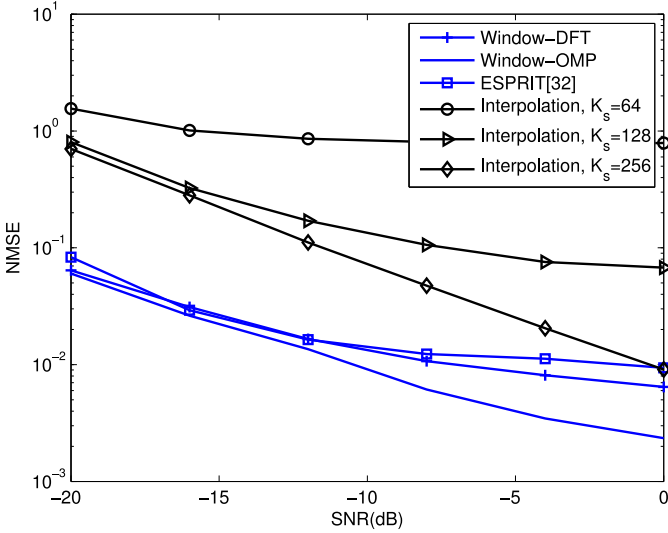
Fig. 9. Comparison of NMSE of the CE for all subcarriers achieved by the conventional interpolation method [27], the ESPRIT method [32], the proposed window-DFT, and window-OMP methods. $K_s = 64$ for the ESPRIT, window-DFT, and window-OMP methods.
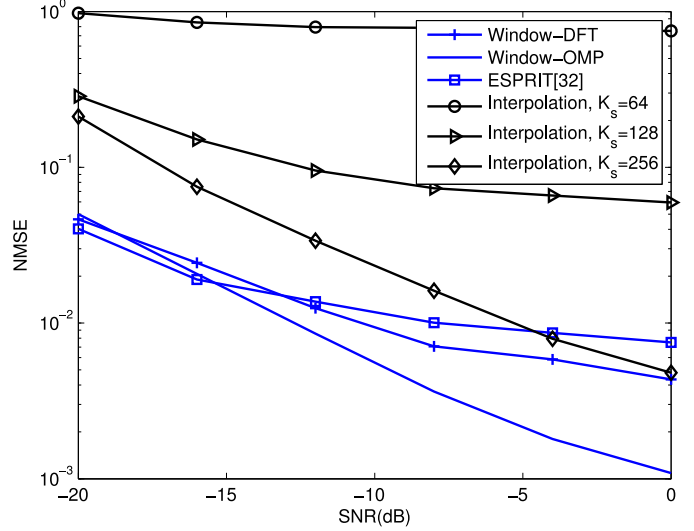

Fig. 10. Comparison of NMSE of the CE for all subcarriers achieved by the conventional interpolation method [27], the ESPRIT method [32], the proposed window-DFT, and window-OMP methods for the 2×2 RF-chains system. $K_s = 64$ for the ESPRIT, window-DFT, and window-OMP methods.

Fig. 9 compares the NMSE performance of the CE for all the subcarriers achieved by the conventional interpolation method [27], the existing ESPRIT method [32] as well as our proposed window-DFT and window-OMP methods.

In Fig. 9, the dictionary size $N_{dic} = 2048$ and the sparse level $L_p = 4$ are used for both the window-DFT and window-OMP methods. Furthermore, we have $K_s = 64$ for the ESPRIT, window-DFT and window-OMP methods. It can be seen that the proposed window-DFT and window-OMP methods generally outperform the ESPRIT method. Observe that the window-OMP method clearly outperform the window-DFT method, particularly at the high SNR region. The reason is that the window-DFT method requires that the frequencies to be estimated are sufficiently separated, but this condition does not hold for this example. The better performance of the window-OMP over the window-DFT is of course achieved at the expense of high computational complexity.

In Fig. 9, we also plot the NMSE performance of the conventional interpolation method with the number of pilots set to $K_s = 64$, 128 and 256, which correspond to the pilot interval $K_p = 16$, 8 and 4, respectively. Obviously, for the interpolation method to work, the pilot density should be high, namely, K_s is large and K_p is small. Even with $K_s = 256$, the NMSE of the conventional interpolation method is considerably worse than the ESPRIT, window-DFT and window-OMP methods, especially when the SNR is low. Also observe from Fig. 9 that our proposed window-OMP method with $K_s = 64$ consistently outperforms the conventional interpolation method with $K_s = 256$ over the entire range of SNR values.

D. Extension to Multiple RF Chains Scenario

To demonstrate that our proposed window-DFT and window-OMP methods are equally applicable to multiple RF-chains

systems, we set the numbers of RF chains to $N_{RF,t} = N_{RF,r} = 2$. Since there are two simultaneously activated beams at both transmitter and receiver, the beam selection criterion is to choose two pairs of transmitter steering vector and receiver steering vector which achieve the strongest and second-strongest power gains, respectively, or equivalently, it can be expressed as

$$
\begin{aligned}
&\{\boldsymbol{f}_1^\star, \boldsymbol{w}_1^\star, \boldsymbol{f}_2^\star, \boldsymbol{w}_2^\star\} \\
&= \arg \max_{\substack{\boldsymbol{f}_1, \boldsymbol{w}_1, \boldsymbol{f}_2, \boldsymbol{w}_2 \\ \{\boldsymbol{f}_1, \boldsymbol{w}_1\} \neq \{\boldsymbol{f}_2, \boldsymbol{w}_2\}}} \left|\boldsymbol{w}_1^{\mathrm{H}} \boldsymbol{H} \boldsymbol{f}_1\right| + \left|\boldsymbol{w}_2^{\mathrm{H}} \boldsymbol{H} \boldsymbol{f}_2\right|. \quad (73)
\end{aligned}
$$

The equivalent baseband channel matrix $\boldsymbol{H}_{BB}$ is a 2×2 matrix with the 4 coefficients, $(\boldsymbol{w}_i^\star)^{\mathrm{H}} \boldsymbol{H} \boldsymbol{f}_j^\star$, $i, j = 1, 2$. Since the two simultaneously used pairs of the steering vectors are usually different at both transmitter and receiver, these 4 coefficients in $\boldsymbol{H}_{BB}$ typically correspond to the different and independent paths. Therefore, we can separately estimate these 4 channel coefficients over all the subcarriers using the conventional interpolation based method and the proposed HR based method.

The corresponding NMSE comparison is demonstrated in Fig. 10. Basically, the performances of all the four CE methods are better, compared to the single RF-chain scenario. This is to be expected, since we now have more RF chains. Similar to Fig. 9 for the single RF-chain system, it can be seen that the proposed window-OMP method attains the best performance, and our window-DFT method achieves slightly better performance than the ESPRIT method. It can also be seen that the ESPRIT method outperforms the conventional interpolation method with $K_s = 256$.

VII. CONCLUSIONS

In this paper, the baseband CE after beam searching for the mmWave OFDM system has been formulated as a harmonic

retrieval problem. We have proposed the window-OMP and window-DFT methods, which are capable of acquiring approximate ML estimate. The window-OMP method generally attains higher estimation accuracy than the window-DFT method, while also imposing higher computational complexity. It has been demonstrated that our two proposed methods achieve higher estimation accuracy than the existing ESPRIT method and the conventional interpolation method. Moreover, compared to the conventional interpolation method, the pilot density of the proposed methods can be significantly reduced and pilots do not have to cover the whole frequency band. Therefore, our proposed methods offer further advantage of much more flexible pilot pattern. Compared with the existing methods of CE before beam searching, our approach also offers considerable advantage, in terms of pilot overhead requirement and processing complexity.

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] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO system," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[3] Q. Li, H. Jung, P. Zong, and G. Wu, "5G millimeter-wave communication channel and technology overview," in *Signal Processing for 5G: Algorithms and Implementations*, F.-L. Luo and C. Zhang Eds. Chichester, U.K.: Wiley, 2016, pp. 354–371.

[4] F. Rusek *et al.,* "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[6] X. Guo, S. Chen, J. Zhang, X. Mu, and L. Hanzo, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7229–7243, Nov. 2016.

[7] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[8] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. 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.

[9] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[10] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analog-digital precoding revisited under realistic RF modeling," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 528–531, Oct. 2016.

[11] O. E. Ayach *et al.*, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. 13th Int. Workshop Signal Process. Adv. Wireless Commun.*, Cesme, Turkey, Jun. 17–20, 2012, pp. 100–104.

[12] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5G mmWave cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 40–47, Nov. 2016.

[13] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

[14] P. Zhang, S. Chen, and L. Hanzo, "Reduced-complexity near-capacity joint channel estimation and three-stage turbo detection for coherent space-time shift keying," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1902–1913, May 2013.

[15] P. Zhang, S. Chen, and L. Hanzo, "Embedded iterative semi-blind channel estimation for three-stage-concatenated MIMO-aided QAM turbo-transceivers," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 439–446, Jan. 2014.

[16] P. Zhang, S. Chen, and L. Hanzo, "Two-tier channel estimation aided near-capcity MIMO transceivers relaying on norm-based joint transmit and receive antenna selection," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 122–137, Jan. 2015.

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

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

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

[20] 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.

[21] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.

[22] 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.

[23] C. Huang, L. Liu, C. Yuen, and S. Sun, "A LSE and sparse message passing-based channel estimation for mmWave MIMO systems," in *Proc. Globecom 2016 Workshops*, Washington, DC, USA, Dec. 4–8, 2016, pp. 1–6.

[24] L. Pan, L. Liang, W. Xu, and X. Dong, "Framework of channel estimation for hybrid analog-and-digital processing enabled massive MIMO communications," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3902–2915, Sep. 2018.

[25] Y. Wang, W. Xu, H. Zhang, and X. You, "Wideband mmWave channel estimation for hybrid massive MIMO with low-precision ADCs," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 285–288, Feb. 2019.

[26] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[27] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

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

[30] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[31] M. C. Vanderveen and A.-J. Vanderveen, A. Paulraj, "Estimation of Multipath Parameters in Wireless Communications," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 682–690, Mar. 1998.

[32] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

[33] S. Liu, F. Wang, R. Zhang, and Y. Liu, "A simplified parametric channel estimation scheme for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5082–5090, Dec. 2008.

[34] *Study on Channel Model for Frequency Spectrum Above 6 GHz*, 3GPP TR 38.900 (Release 15), 2018.

[35] M. Wu *et al.*, "On OFDM and SC-FDE transmissions in millimeter wave channels with beamforming," in *Proc. IEEE Veh. Technol. Conf. Spring* (Nanjing, China), May 15–18, 2016, pp. 1–5.

[36] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[37] *Physical Channels and Modulation*, 3GPP TS 36.211 (Release 14), 2017.

Ziyuan Sha received the B.S. degree in 2017 from Tsinghua University, Beijing, China, where he is currently working toward the Ph.D degree. His main research interest is millimeter-wave communications.

Zhaocheng Wang (M'09–SM'11) received the B.S., M.S., and Ph.D. degrees from Tsinghua University, Beijing, China, in 1991, 1993, and 1996, respectively. From 1996 to 1997, he was a Postdoctoral fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was with OKI Techno Centre Pte. Ltd., Singapore, where he was first a Research Engineer and later became a Senior Engineer. From 1999 to 2009, he was with Sony Deutschland GmbH, where he was first a Senior Engineer and later became a Principal Engineer. He is currently a Professor of Electronic Engineering with Tsinghua University and serves as the Director of the Broadband Communication Key Laboratory, Tsinghua National Laboratory for Information Science and Technology. He has authored or coauthored more than 120 journal papers. He holds 34 granted U.S./EU patents. He has co-authored two books, one of which, *Millimeter Wave Communication Systems*, was selected by IEEE Series on Digital and Mobile Communication (Wiley-IEEE Press). His research interests include wireless communications, visible light communications, millimeter-wave communications, and digital broadcasting. He is a fellow of the Institution of Engineering and Technology. He was the Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2015 and IEEE COMMUNICATIONS LETTERS from 2013 to 2016, and was also the Technical Program Committee Co-Chair of various international conferences.

Sheng Chen (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from City University, London, U.K., in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments at the Universities of Sheffield, Edinburgh, and Portsmouth, all in U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he is currently a Professor in intelligent systems and signal processing. He has authored or coauthored more than 600 research papers. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods, and optimization. He is a fellow of the United Kingdom Royal Academy of Engineering, a fellow of IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an ISI highly cited researcher in engineering in 2004. He has 11 400 Web of Science citations and more than 24 000 Google Scholar citations.