Partially Coherent Compressive Phase Retrieval for Millimeter-Wave Massive MIMO Channel Estimation

Chen Hu, Xiaodong Wang, Linglong Dai, and Junjie Ma

***Abstract*—Channel estimation is challenging for millimeter-wave (mmWave) massive MIMO systems. Leveraging the spatial sparsity of mmWave channels, compressive sensing (CS) based channel estimation been extensively studied. However, the mmWave hardware imperfections may introduce random phase distortions to the received pilots, which makes the conventional CS methods fail to estimate the channel. The compressive phase retrieval (CPR) method can be employed to solve this challenging problem. In this paper, we exploit the partial coherence in hybrid mmWave systems, i.e., the pilots sent from different radio frequency (RF) chains share the same phase distortion in the same time frame, while the phase distortions are different across different time frames. Based on this property, we propose an on-grid *partially coherent* CPR (PC-CPR) algorithm for mmWave channel estimation in the presence of severe phase distortions. Unlike the existing *coherent* channel estimation schemes that require perfect phase information or the *noncoherent* channel estimation schemes that ignore the phases of measurements, the proposed on-grid PC-CPR algorithm exploits the partial coherence property to estimate the sparse angle-domain channel vector. Furthermore, to solve the resolution limitation of the on-grid PC-CPR algorithm, we propose an off-grid PC-CPR algorithm that directly estimates the parameters of channel paths. The proposed partially coherent channel estimation framework subsumes the existing coherent and noncoherent channel estimation methods as special cases. Simulation results show that under the presence of random phase distortions, the proposed PC-CPR algorithms outperform noncoherent channel estimation methods with higher reliability and lower pilot overhead by leveraging the partial coherence.**

***Index Terms*—Millimeter-Wave (mmWave), massive MIMO, hybrid precoding, sparse channel estimation, compressive phase retrieval, partial coherence.**

I. Introduction

MILLIMETER-WAVE (mmWave) communication is recognized as one of the key technologies to support the ever-increasing wireless data traffic, since it has the potential to increase both per-user and network capacity by several orders of magnitude [1]. In a mmWave massive multiple input multiple output (MIMO) system, the base station (BS) is usually equipped with an array with a large number of antennas. The large array is used to achieve high array gains to compensate for the severe path loss in mmWave band, and to form narrow beams to separate different users in space. To reduce the hardware cost and energy consumption in mmWave massive MIMO systems, hybrid precoding is regarded as a promising architecture [2], where hundreds of antennas are connected to a much smaller number of radio frequency (RF) chains via analog phase shifters [3]. To realize the potential gain of mmWave massive MIMO, estimating the channel state information is of great importance. In a hybrid precoding system, the BS needs to manipulate the analog precoders to broadcast pilots to different directions [3]. It takes multiple time frames to transmit enough pilots for a reliable channel estimation, which results in a high training overhead [4]. Thus, the channel estimation problem for mmWave massive MIMO is challenging.

A. Prior Works

To reduce the training overhead, several compressive sensing (CS) based channel estimation schemes have been proposed [5]–[9]. These schemes exploit the property that mmWave channels typically have a small number of dominant paths, thus the channel is sparse in the angle domain. They collect the complex-valued pilot measurements with reduced dimensionality, and estimate the sparse channel vector based on CS techniques. With the help of CS, the required number of pilot measurements can be much smaller than that of antennas. To acquire the complex-valued pilot measurements with both amplitudes and phases for channel estimation, the transmitter and receiver should be *coherent* during the entire training period. In this paper, *coherent* channel estimation means that the transceivers are synchronized in the process of channel estimation, particularly the phases at the transmitter and receiver are synchronized. However, due to the large carrier frequency offset (CFO) and random phase drift caused by the hardware imperfections, the received pilots are usually corrupted by random phase offsets in different time frames [10]. Moreover, it is not easy to compensate for the offset before the channel is estimated and the link is established. It has been reported that when the random phase offset is significant but not considered, the CS-based channel estimation schemes will fail to work [10], [11]. Thus, it is practically important to

Manuscript received September 18, 2019; revised January 9, 2020; accepted February 11, 2020. Date of publication February 24, 2020; date of current version March 13, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mats Bengtsson. This work was supported in part by the National Science and Technology Major Project of China under Grant 2018ZX03001004 and in part by the National Natural Science Foundation of China for Outstanding Young Scholars under Grant 61722109. *(Corresponding author: Linglong Dai.)*

Chen Hu and Linglong Dai are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: huc16@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn).

Xiaodong Wang is with the Electrical Engineering Department, Columbia University, New York, NY 10027 USA (e-mail: xw2008@columbia.edu).

Junjie Ma was with the Electrical Engineering Department, Columbia University, New York, NY 10027 USA. He is now with the Department of Electrical Engineering, Harvard University, Cambridge, MA 02138 USA (e-mail: junjiema@seas.harvard.edu).

Digital Object Identifier 10.1109/TSP.2020.2975914

study the problem of estimating the sparse mmWave channel in the presence of phase offsets.

Very recently, there have been several works on *noncoherent* channel estimation or beam alignment in the presence of phase offsets for mmWave massive MIMO systems [12]–[17]. *Noncoherent* channel estimation means that the channel estimation is realized without perfect synchronization, and thus phase offsets exist in the received pilots. Specifically, in [12], [13], the phase offsets and the mmWave channel are jointly estimated. These schemes are based on recovering high-dimensional sparse tensors, which have high computational complexity. In [14]–[17], all phase information in the received pilots is assumed unavailable. The mmWave channel is estimated only based on the amplitudes of received pilots. For example, [15] designs a set of special beam patterns for channel measurements, and the beam alignment problem is formulated as a coding-and-decoding problem. The directions and the magnitudes of the path gains can be efficiently recovered with a small number of samples. In [17], a two-stage method is proposed to recover the phase information at first and then estimate the sparse channel, but its performance degrades severely when the signal-to-noise ratio (SNR) is low.

Another straightforward approach is to estimate the channel using the compressive phase retrieval (CPR) technique by simply ignoring all phases of the received pilots. CPR is an emerging signal processing technique that has attracted increasing research interests in signal processing [18]–[22]. It studies the problem of how to recover a sparse signal if we have only the amplitudes of the complex measurements. However, CPR has not been considered for mmWave channel estimation.

B. Contributions

In this paper, we study the mmWave channel estimation problem under the presence of severe phase offsets.

Firstly, we exploit the partial coherence in hybrid mmWave systems and formulate the on-grid partially coherent channel estimation problem. In this paper, *on-grid* channel estimation means that we assume the channel power distributes only on a few discrete angle-domain grids so that the angle-domain channel vector can be estimated as a sparse vector. The *partial coherence* refers to the property that the pilots sent from different RF chains share the same phase shift in the same time frame but have different phase shifts across different time frames. We design a novel on-grid partially coherent CPR (PC-CPR) algorithm to solve the problem. Different from the coherent channel estimation that requires all the phase information and the noncoherent channel estimation that ignores all the phases of the measurements, the proposed on-grid PC-CPR algorithm makes use of the partial coherence property. To be specific, in the initialization stage, we propose new test statistics. We prove that the reliability of the initialization can be significantly improved by exploiting the partial coherence. In the iterative refinement stage, we estimate the phase offsets shared by each time frame rather than estimating the phases of all measured pilots.

Secondly, since the real path angles of the channel are continuously distributed, the on-grid assumption may result in a resolution loss that limits the channel estimation accuracy. To solve this problem, we further formulate the off-grid partially coherent channel estimation problem and propose an off-grid PC-CPR algorithm. Different from the on-grid channel estimation, the *off-grid* method estimates the continuous mmWave path parameters. The off-grid PC-CPR consists of an initialization stage and an iterative refinement stage. The initialization stage employs a successive path estimation and cancelation scheme, and the iterative refinement stage alternately updates the phase offset estimates and refines the channel parameter estimates using gradient descent.

A salient feature of the proposed partially coherent channel estimation framework is that it subsumes the conventional coherent and noncoherent channel estimation methods as special cases. To be specific, when there is only one RF chain, the partial coherence vanishes, so the performance of the proposed on-grid PC-CPR is reduced to that of the existing noncoherent CPR algorithm. When the number of RF chains becomes large, more pilot measurements share the same phase offset within the same time frame. Consequently, the partial coherence approaches full coherence, so the performances of the PC-CPR are close to those of existing coherent channel estimation algorithms. Simulation results[1] show that under the presence of random phase offset, the proposed PC-CPR algorithms outperform the existing noncoherent channel estimation methods by leveraging the partial coherence. Besides, the off-grid PC-CPR algorithm can achieve higher channel estimation accuracy than the on-grid PC-CPR algorithm at the expense of higher computational complexity.

C. Organization and Notations

The remainder of the paper is organized as follows. In Section II, we formulate the signal model and the partially coherent channel estimation problem in mmWave massive MIMO systems with hybrid precoding. In Section III, we propose the on-grid PC-CPR algorithm by leveraging the partially coherent phase structure. The off-grid PC-CPR algorithm is presented in Section IV. Simulation results are provided in Section V, and finally conclusions are drawn in Section VI.

Notation: In this paper, light symbols, boldface lower-case symbols and upper-case symbols denote scalars, vectors and matrices, respectively. For a vector $\mathbf{x}$, we use x_i for its i-th element. For a matrix $\mathbf{X}$, $\mathbf{X}(i,:)$ and $\mathbf{X}(:,j)$ denote its i-th row and j-th column respectively. Given an index set $\mathcal{S}$, $\mathbf{x}_{\mathcal{S}}$ is the sub-vector that consists of x_i's for all $i \in \mathcal{S}$, and $\mathbf{X}(\mathcal{S},:)$ is the sub-matrix that consists of rows whose indices are in $\mathcal{S}$. $(\cdot)^*$, $|\cdot|$ and $\angle(\cdot)$ denote the conjugate, the amplitude and the angle of a complex number, while $\|\cdot\|_0$ and $\|\cdot\|$ denote the ℓ_0-norm and the ℓ_2-norm of a vector, respectively. $(\cdot)^T$ and $(\cdot)^H$ are the transpose and the conjugate transpose of a matrix, respectively. $\otimes$ is the Kronecker product, while $\circ$ is the Hadamard product. $\mathbf{I}_K$ is the $K \times K$ identity matrix. $\mathrm{diag}(\mathbf{x})$ is the diagonal matrix with the vector $\mathbf{x}$ on its diagonal. Finally, $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$.

[1]Simulation codes are provided to reproduce the results presented in this paper: http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html.

II. System Model and Problem Formulation

A. MmWave Massive MIMO

Due to the high sampling rate and the large array size in mmWave massive MIMO, it is not practical to apply the fully-digital system where each antenna is driven by an independent RF chain. As an alternative to the costly fully-digital structure, the hybrid precoding architecture connects the antennas to a much smaller number of RF chains via analog phase shifters. In the hybrid precoding mmWave massive MIMO system with N BS antenna and N_{RF} RF chains, the BS transmits N_{S} data streams to K single-antenna users, the downlink signal model can be written as [2]

$$\mathbf{r} = \mathbf{H}^H \mathbf{P}_{\mathrm{A}} \mathbf{P}_{\mathrm{D}} \mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{r} \in \mathbb{C}^{K\times 1}$ is the received signal at the users, $\mathbf{H} \in \mathbb{C}^{N\times K}$ is the channel matrix, $\mathbf{P}_{\mathrm{A}} \in \mathbb{C}^{N\times N_{\mathrm{RF}}}$ is the analog precoder, $\mathbf{P}_{\mathrm{D}} \in \mathbb{C}^{N_{\mathrm{RF}}\times N_{\mathrm{S}}}$ is the digital precoder, $\mathbf{s} \in \mathbb{C}^{N_{\mathrm{S}}\times 1}$ is the transmitted data streams, and $\mathbf{n} \in \mathbb{C}^{K\times 1}$ is the noise. The channel matrix $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$ consists of the channel vectors of all K users. The channel model of a specific user is described as follows.

For a typical uniform planar array (UPA) with N_1 columns and N_2 rows of antenna elements, the total number of antennas $N = N_1 N_2$. We follow the widely adopted three-dimensional channel model as [1]

$$\mathbf{h} = \sqrt{\frac{1}{\kappa + L - 1}} \sum_{l=1}^{L} \bar{\alpha}_l \bar{\mathbf{a}}\left(\theta_l, \phi_l\right), \tag{2}$$

where L is the number of paths, κ is the ratio of line-of-sight (LoS) path power to non-line-of-sight (NLoS) path power. $\bar{\alpha}_l$, θ_l and ϕ_l are the normalized complex path gain, the physical azimuth angle, and the physical elevation angle of the l-th path, respectively. $\bar{\alpha}_1 \bar{\mathbf{a}}(\theta_1, \phi_1)$ is the LoS path, and $\bar{\alpha}_l \bar{\mathbf{a}}(\theta_l, \phi_l)$ for $2 \le l \le L$ denote the NLoS components. The normalized path gains $\bar{\alpha}_l$ for $2 \le l \le L$ are assumed to be circularly symmetric complex Gaussian, i.e., $\bar{\alpha}_l \sim \mathcal{CN}(0, 1)$, and the LoS component $\bar{\alpha}_1 \sim \mathcal{CN}(0, \kappa)$. The azimuth angles θ_l are uniformly distributed in $[0, 2\pi)$, and the elevation angles ϕ_l are uniformly distributed in $(-\pi/2, \pi/2)$. The array steering vector $\bar{\mathbf{a}}(\theta_l, \phi_l) \in \mathbb{C}^{N\times 1}$ is given by

$$\begin{aligned}\bar{\mathbf{a}}(\theta_l, \phi_l) = \frac{1}{\sqrt{N}} &\left[e^{j2\pi \frac{0 d \sin\theta_l \cos\phi_l}{\lambda}}, \ldots, e^{j2\pi \frac{(N_1-1) d \sin\theta_l \cos\phi_l}{\lambda}}\right]^T \\ &\otimes \left[e^{j2\pi \frac{0 d \sin\phi_l}{\lambda}}, e^{j2\pi \frac{1 d \sin\phi_l}{\lambda}}, \ldots, e^{j2\pi \frac{(N_2-1) d \sin\phi_l}{\lambda}}\right]^T,\end{aligned} \tag{3}$$

where d is the antenna spacing, λ is the carrier wavelength. In this paper, $d = \lambda/2$. By defining $\alpha_l = \bar{\alpha}_l \sqrt{\frac{1}{\kappa+L-1}}$, the spatial direction $(\vartheta_l, \varphi_l) = (\frac{2\pi d \sin\theta_l \cos\phi_l}{\lambda}, \frac{2\pi d \sin\phi_l}{\lambda})$ and $\mathbf{a}(\vartheta_l, \varphi_l) = \bar{\mathbf{a}}(\theta_l, \phi_l)$, we can rewrite (2) as

$$\mathbf{h} = \sum_{l=1}^{L} \alpha_l \mathbf{a}\left(\vartheta_l, \varphi_l\right). \tag{4}$$

Note that $\vartheta_l, \varphi_l \in [-\pi, \pi]$ according to their definitions, and $\mathbf{a}(\vartheta_l, \varphi_l) = \mathbf{a}(\vartheta_l + 2\pi, \varphi_l) = \mathbf{a}(\vartheta_l, \varphi_l + 2\pi)$ is a periodic function, so we equivalently assume $\vartheta_l, \varphi_l \in [0, 2\pi)$ for convenience.

Thanks to the form of the UPA steering vector, we can express the channel vector in the angle domain. Specifically, the angle-domain channel vector $\tilde{\mathbf{h}}$ is defined by [5]

$$\mathbf{h} = \mathbf{F}^H \tilde{\mathbf{h}} = \left(\mathbf{F}_{N_1\times N_1} \otimes \mathbf{F}_{N_2\times N_2}\right)^H \tilde{\mathbf{h}}, \tag{5}$$

where $\mathbf{F}_{N_1\times N_1}$ and $\mathbf{F}_{N_2\times N_2}$ are discrete Fourier transform (DFT) matrices of sizes $N_1 \times N_1$ and $N_2 \times N_2$, respectively. The $((i-1)N_2 + j)$-th element of $\tilde{\mathbf{h}}$, $\tilde{h}_{(i-1)N_2+j}$, is the channel component at the spatial direction of $(2\pi(i-1)/N_1, 2\pi(j-1)/N_2)$ $(1 \le i \le N_1,\ 1 \le j \le N_2)$. These discrete spatial directions are called angle-domain grids in this paper. Due to the fact that the mmWave channels are made up of a few paths, and only the angle-domain components that are close to the channel path directions have large values, the angular channel vector $\tilde{\mathbf{h}}$ is approximately sparse [1].

B. Traditional Coherent Pilot Transmission

To compensate for the severe path loss in the mmWave band, the precoding matrices in (1) must be carefully designed based on the knowledge about the channel $\mathbf{H}$. To acquire the channel information for beamforming, most existing channel estimation works [3]–[6] are based on traditional coherent pilot transmission, which means that the transceivers are perfectly synchronized during the pilot transmission phase.

In this paper, we consider the downlink channel estimation, where the BS broadcasts pilots to all users, and the users estimate their channels and feedback them to the BS. The reason why we choose the downlink channel estimation is that the channel estimation accuracy is sensitive to the power of the received pilots. Since the BS typically has a larger transmit power than the users, the power of the received pilots in the downlink channel estimation can be larger compared to that in the uplink channel estimation, so the downlink channel estimation can be more practical under the severe path loss.

For a specific user, the traditional coherent downlink pilot transmission in the t-th time slot can be modeled by

$$\begin{aligned}\bar{y} &= \mathbf{h}^H \mathbf{P}_{\mathrm{A}} \mathbf{P}_{\mathrm{D}} \mathbf{s} + \bar{n} \\ &= \mathbf{h}^H \mathbf{P}_{\mathrm{A}} \mathbf{x} + \bar{n},\end{aligned} \tag{6}$$

where $\bar{y} \in \mathbb{C}$ is the received pilot, $\mathbf{h} \in \mathbb{C}^{N\times 1}$ is the channel vector of the user, $\mathbf{P}_{\mathrm{A}}$, $\mathbf{P}_{\mathrm{D}}$, and $\mathbf{s}$ are the analog precoder, the digital precoder, and the transmitted pilot, respectively. $\bar{n} \sim \mathcal{CN}(0, \sigma_n^2)$ is the noise. $\mathbf{x} = \mathbf{P}_{\mathrm{D}}\mathbf{s}$ is the equivalent transmit pilots at the RF chains.

In a hybrid mmWave system, the analog precoder cannot be reconfigured for each time slot, e.g., phase-shifters need around 64 time slots to be updated for a new configuration [12]. During the b-th time frame with the analog precoder $\mathbf{P}_{\mathrm{A},b}$, we spend $N_{\mathrm{TS}} \ge N_{\mathrm{RF}}$ time slots to transmit orthogonal pilot sequences in different RF chains, which can be modeled by

$$\bar{\mathbf{y}}_b^H = \mathbf{h}^H \mathbf{P}_{\mathrm{A},b} \mathbf{X} + \bar{\mathbf{n}}_b^H, \tag{7}$$

where $\bar{\mathbf{y}}_b^H = [\bar{y}_{b,1}, \bar{y}_{b,2}, \ldots, \bar{y}_{b,N_{\rm TS}}]$ contains the received pilots in $N_{\rm TS}$ time slots, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_{\rm TS}}]$ is the orthogonal pilot matrix such that $\mathbf{X}\mathbf{X}^H = N_{\rm TS}P_{\rm T}\mathbf{I}_{N_{\rm RF}}$, $P_{\rm T}$ is the transmitted power of each RF chain, and $\bar{\mathbf{n}}^H = [\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_{N_{\rm TS}}]$.

Though we can receive $N_{\rm TS}$ pilots, the number of independent measurements is $N_{\rm RF}$ because $\mathbf{P}_{\rm A}$ has only $N_{\rm RF}$ columns. Then the $N_{\rm RF}$-dimensional effective received pilots are given by

$$\begin{aligned}\mathbf{y}_b &= \frac{1}{\sqrt{N_{\rm TS}P_{\rm T}}}\mathbf{X}\bar{\mathbf{y}}_b = \frac{1}{\sqrt{N_{\rm TS}P_{\rm T}}}\left(\mathbf{X}\mathbf{X}^H\mathbf{P}_{{\rm A},b}^H\mathbf{h} + \mathbf{X}\bar{\mathbf{n}}_b\right)\\ &= \sqrt{N_{\rm TS}P_{\rm T}}\mathbf{P}_{{\rm A},b}^H\mathbf{h} + \mathbf{n}_b, \end{aligned} \tag{8}$$

where $\mathbf{y}_b \in \mathbb{C}^{N_{\rm RF}\times 1}$ consists of $N_{\rm RF}$ independent effective received pilots, $\mathbf{n}_b = \frac{1}{\sqrt{N_{\rm TS}P_{\rm T}}}\mathbf{X}\bar{\mathbf{n}}_b$, and thus we have $\mathbf{n}_b \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I}_{N_{\rm RF}})$. We can equivalently write

$$\mathbf{y}_b = \mathbf{W}_b^H\mathbf{h} + \mathbf{n}_b = \mathbf{A}_b^H\tilde{\mathbf{h}} + \mathbf{n}_b, \tag{9}$$

where $\mathbf{W}_b = \sqrt{N_{\rm TS}P_{\rm T}}\mathbf{P}_{{\rm A},b}$, and $\mathbf{A}_b = \mathbf{F}\mathbf{W}_b$ according to (5). In order to receive more independent pilots, we repeat (7)–(9) for $1 \le b \le B$, where B is the number of time frames. Thus the number of independent pilots is $M = BN_{\rm RF}$. When there are no phase offsets, we have:

$$\mathbf{y} = \mathbf{W}^H\mathbf{h} + \mathbf{n} = \mathbf{A}^H\tilde{\mathbf{h}} + \mathbf{n}, \tag{10}$$

where $\mathbf{y} = [\mathbf{y}_1^H, \mathbf{y}_2^H, \ldots, \mathbf{y}_B^H]^H \in \mathbb{C}^M$, $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_B] \in \mathbb{C}^{N\times M}$, $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_B] \in \mathbb{C}^{N\times M}$, and $\mathbf{n} = [\mathbf{n}_1^H, \mathbf{n}_2^H, \ldots, \mathbf{n}_B^H]^H \in \mathbb{C}^{M\times 1}$. The measurement matrices $\mathbf{W}_b$ for $b = 1, 2, \ldots, B$ should be different from one another because the BS needs to collect uncorrelated measurements to estimate the channel [5].

C. Partially Coherent Pilot Transmission

The model (10) is the coherent pilot transmission. However, the coherence does not hold under severe hardware imperfections in mmWave communication systems [11], [16]. Before we proceed, we briefly explain the causes of the non-coherence. Due to the imperfect oscillators, there is a large CFO between the transmitter and the receiver, particularly in the high-frequency mmWave band [14]. It has been found that the received mmWave pilots are distorted by random phase offsets due to the CFO, which is not easy to compensate during the pilot training time. As a result, there are random phase offsets among the measured pilots.

The phase offset between two pilot measurements can be calculated by $\epsilon = 2\pi T\Delta f$ [26]. The CFO Δf is typically as large as several parts per millions (ppm) of the carrier frequency f. T is the time interval between two pilot measurements. On one hand, the phase offset between adjacent time slots within the same time frame can be neglected, because the sampling period is short (e.g., 1 nanosecond) due to the large bandwidth in mmWave communications. So we can still calculate the effective received pilots according to (8). On the other hand, the phase offset between adjacent time frames can not be neglected, because the time duration of a frame is much longer than the sampling period. [12] reported that the phase-shifters need around 64 samples to be updated for a new configuration, which means that T is on the order of tens to hundreds of nanoseconds. For example, assuming the carrier frequency $f = 73$ GHz, a CFO as small as 5 ppm can cause a phase shift as large as 0.25 rad between two adjacent measurements in 100 nanoseconds. [10] also reported that in a system operating at 28 GHz, the standard deviation of the random phase offset can be up to 0.27 rad within 0.5 microsecond. Such large phase offsets across frames can distort the complex value of the efficient received pilots, so they can no longer be neglected. The 802.11ad standard [24] also assumes phase coherence within a frame, but cannot maintain phase coherence across time frames. Therefore, the phase non-coherence across time frames is a serious issue that has to be dealt with in practical mmWave communication systems.

In traditional communication systems, such offsets can be compensated by carrier synchronization [25]. For data transmission in mmWave systems, given the estimated channel, by choosing fixed analog precoders, we can also insert synchronization reference signals into each data frame to compensate for the phase offset. However, it is challenging to synchronize during the pilot transmission period in mmWave systems. One reason is that we cannot optimize the analog precoders without knowing the channel. Without directional beamforming, we cannot ensure that the synchronization signal is heard to all users. Another reason is that the downlink pilots are broadcasted with different analog precoders in different time frames in (9). Both the changing precoders and the random phase offset can change the phases of the received synchronization signal, so it is non-trivial to remove the random phase offset during the pilot transmission. Consequently, we have to first estimate the spatial channel under the existence of the phase offsets, and only after that can we adopt fixed precoders to successfully transmit signals to the users.

On the other hand, in a hybrid mmWave massive MIMO system, there are $N_{\rm RF}$ RF chains that transmit pilot simultaneously. In the same time frame, the effective received pilots corresponding to different transmitter RF chains are polluted by the same phase offset because the offset is caused by the same CFO [12]. Therefore, the pilot model (9) should be modified to

$$\mathbf{y}_b = \mathbf{W}_b^H\mathbf{h}e^{j\omega_b} + \mathbf{n}_b = \mathbf{A}_b^H\tilde{\mathbf{h}}e^{j\omega_b} + \mathbf{n}_b, \quad b = 1, \ldots, B, \tag{11}$$

where $\omega_b \sim \mathcal{U}[0, 2\pi)$ represents the phase offset in time frame b. Moreover, (10) becomes

$$\begin{aligned}\mathbf{y} &= \begin{bmatrix} e^{j\omega_1}\mathbf{I}_{N_{\rm RF}} & & & \\ & e^{j\omega_2}\mathbf{I}_{N_{\rm RF}} & & \\ & & \ddots & \\ & & & e^{j\omega_B}\mathbf{I}_{N_{\rm RF}} \end{bmatrix}\mathbf{W}^H\mathbf{h} + \mathbf{n}\\ &= \begin{bmatrix} e^{j\omega_1}\mathbf{I}_{N_{\rm RF}} & & & \\ & e^{j\omega_2}\mathbf{I}_{N_{\rm RF}} & & \\ & & \ddots & \\ & & & e^{j\omega_B}\mathbf{I}_{N_{\rm RF}} \end{bmatrix}\mathbf{A}^H\tilde{\mathbf{h}} + \mathbf{n}. \end{aligned} \tag{12}$$

We call this model *partially coherent* pilot transmission in this paper, which means the phase coherence holds between the effective received pilots corresponding to different transmitter RF chains within the same time frame, but the coherence is lost across different time frames. Note that this model subsumes

the noncoherent pilot transmission and the coherent pilot transmission. When $N_{\mathrm{RF}} = 1$, each element of $\mathbf{y}$ is distorted by an independent phase offset, so the received pilots become fully noncoherent; while when $B = 1$, (12) becomes the coherent pilot transmission model (10) distorted by an overall phase offset.

D. Channel Estimation Problem Formulations

Due to the fact that the mmWave channels only contain a small number of dominant paths, the angular channel vector $\tilde{\mathbf{h}}$ in (5) is approximately sparse. By keeping only the k largest elements, $\tilde{\mathbf{h}}$ can be treated as a k-sparse vector to be estimated. Since k is usually unknown in mmWave channel estimation, we can set k to a multiple of the average number of paths. To solve the sparse channel estimation problem under the random phase offsets $\omega_1, \ldots, \omega_B$ based on the model in (12), one could simply ignore all phase information of the received pilots, and estimate the channel by using only the magnitudes $|\mathbf{y}|$. As a result, the sparse channel estimation with phase offsets can be formulated as the standard "*compressive phase retrieval*" (CPR) problem, which is given by [19]:

$$\min_{\tilde{\mathbf{h}} \in \mathbb{C}^N} \left\| |\mathbf{y}| - \left| \mathbf{A}^H \tilde{\mathbf{h}} \right| \right\|^2 \quad \text{s.t.} \left\| \tilde{\mathbf{h}} \right\|_0 \leq k. \tag{13}$$

Compared with the classical CS problem, the CPR problem (13) is much more difficult due to the lack of phase information in the measurements [27]. Several typical CPR algorithms [20]–[22] have been recently proposed to solve this problem. Moreover, by imposing the nonlinear absolute operator on the received pilots, the additive noise becomes more difficult to handle, which makes the ℓ_2 loss function suboptimal.

As an alternative, we consider the partial coherence in the received pilots, and formulate the sparse channel estimation problem by

$$\min_{\tilde{\mathbf{h}} \in \mathbb{C}^N, \omega_1, \ldots, \omega_B \in [0, 2\pi)} \sum_{b=1}^{B} \left\| \mathbf{y}_b - e^{j\omega_b} \mathbf{A}_b^H \tilde{\mathbf{h}} \right\|^2, \quad \text{s.t.} \left\| \tilde{\mathbf{h}} \right\|_0 \leq k. \tag{14}$$

Since the non-zero elements of $\tilde{\mathbf{h}}$ are the channel gains in the corresponding angle-domain grids defined in (5), this problem is an on-grid partially coherent channel estimation problem. Note that (13) can be subsumed as a special case of (14) with $N_{\mathrm{RF}} = 1$.

However, the real path directions can take any continuous values, and they do not always match well with the angle-domain grids. As a result, the on-grid channel estimation leads to unavoidable estimation error [28], [29]. To further improve the channel estimation accuracy by estimating the off-grid spatial directions, we formulate the off-grid partially coherent channel estimation problem by

$$\min_{\substack{\alpha_1, \ldots, \alpha_L \in \mathbb{C}, \\ \vartheta_1, \ldots, \vartheta_L, \varphi_1, \ldots, \varphi_L, \omega_1, \ldots, \omega_B \in [0, 2\pi)}} \sum_{b=1}^{B} \left\| \mathbf{y}_b - e^{j\omega_b} \mathbf{W}_b^H \mathbf{h} \right\|^2,$$

$$\text{s.t.} \, \mathbf{h} = \sum_{l=1}^{L} \alpha_l \mathbf{a}\left(\vartheta_l, \varphi_l\right). \tag{15}$$

Note that the number of paths L is usually unknown. As an alternative, we replace L by $\bar{L}$ such that $\bar{L} > L$ in most cases.

III. Proposed On-grid PC-CPR Channel Estimation Algorithm

In this section, we will propose an on-grid partially coherent compressive phase retrieval (PC-CPR) algorithm for the on-grid partially coherent channel estimation problem (14). To make it more clear, we will introduce the proposed on-grid PC-CPR algorithm by modifying the compressive reweighted amplitude flow (CRAF) algorithm [20] recently proposed for solving the standard CPR problem (13). However, CRAF does not leverage the phase offset structure in the signal model of the hybrid mmWave system, thus it requires a large number of measurements and has poor channel estimation performance. In contrast, the proposed on-grid PC-CPR can leverage the phase structure in hybrid mmWave systems, which has not been considered in existing CPR algorithms.

Both the proposed on-grid PC-CPR algorithm and CRAF consist of two stages: the initialization and the iterative refinement, which will be discussed in detail in Section III-A and Section III-B respectively. We will show how the proposed on-grid PC-CPR algorithm leverages the phase structure and estimates the mmWave channel with lower overhead and better accuracy compared with CRAF.

A. Initialization Methods

In general, CPR is a non-convex problem [30]. As a result, most CPR algorithms require careful initialization to avoid local minimizers [21]. In the family of all algorithms for CPR as well as the ordinary phase retrieval for recovering non-sparse vectors, "spectral initialization" is a popular initialization method. "Spectral initialization" refers to the category of methods that calculate the initial value of the vector by finding the principal direction of a matrix.

1) The Initialization Method in CRAF: In [20], a variant of spectral initialization was proposed. There are three steps in this initialization method, including determining the initial support, constructing the initialization matrix, and calculating the initial value by eigenvalue decomposition.

Consider the problem in (13). Firstly, given the preset sparsity level k, CRAF detects k indices as the initial guess of the support set. The initial support set $\hat{\mathcal{S}}$ is determined by indices of the k largest elements in the test statistics $\{\zeta_i\}_{i=1}^N$:

$$\zeta_i = \frac{1}{M} \left\| \left| \mathbf{A}\left(i, :\right) \right|^T \circ |\mathbf{y}| \right\|^2, \quad i = 1, \ldots, N, \tag{16}$$

$$\hat{\mathcal{S}} = \left\{ i_1, i_2, \ldots, i_k | \zeta_{i_1} > \cdots > \zeta_{i_k} > \zeta_{i_{k+1}} > \cdots > \zeta_{i_N} \right\}. \tag{17}$$

Secondly, the initialization matrix $\mathbf{D}_{\hat{\mathcal{S}}}$ is formed as

$$\mathbf{D}_{\hat{\mathcal{S}}} = \mathbf{A}\left(\hat{\mathcal{S}}, :\right) \operatorname{diag}\left(\left[\mathcal{T}\left(|y_1|; r\right), \ldots, \mathcal{T}\left(|y_M|; r\right)\right]\right) \times \left(\mathbf{A}\left(\hat{\mathcal{S}}, :\right)\right)^H, \tag{18}$$

where $\mathcal{T}$ is called the "pre-processing function," $r \triangleq \sqrt{\frac{1}{M}\sum_{m=1}^{M}|y_m|^2}$. In [20], the pre-processing function is defined as follows: first the indices of measurements $\{1,\ldots,M\}$ are divided into two sets:

$$\mathcal{I}^- \triangleq \left\{1 \le m \le M : |y_m|^2 \le r^2/2\right\}, \tag{19}$$

$$\mathcal{I}^+ \triangleq \left\{1 \le m \le M : |y_m|^2 > r^2/2\right\}. \tag{20}$$

Then, the empirical pre-processing function is given by

$$\mathcal{T}_{\mathrm{CRAF}}\left(|y_m|;r\right) = \begin{cases} \frac{\lambda^-}{|\mathcal{I}^-|}, & m \in \mathcal{I}^-, \\ \frac{\lambda^+}{|\mathcal{I}^+|}, & m \in \mathcal{I}^+, \end{cases} \tag{21}$$

where $\lambda^+ > 0$ and $\lambda^- < 0$ are preselected coefficients. There are also some theoretical works about the optimal design of the pre-processing function for the ordinary phase retrieval problem (not the compressive PR problem), which can be found in [31], [32].

Thirdly, the direction of the initial value $\tilde{\mathbf{h}}^{(0)}$ of the sparse channel vector $\tilde{\mathbf{h}}$ is estimated by the dominant eigenvector of $\mathbf{D}_{\hat{\mathcal{S}}}$:

$$\mathbf{d} = \arg\max_{\|\mathbf{z}\|=1} \mathbf{z}^H \mathbf{D}_{\hat{\mathcal{S}}} \mathbf{z}. \tag{22}$$

And the magnitude of $\tilde{\mathbf{h}}^{(0)}$ is estimated as

$$\begin{aligned} \hat{r} &= \arg\min_{r' \in \mathbb{R}} \left\| r' \left| \left(\mathbf{A}\left(\hat{\mathcal{S}},:\right)\right)^H \mathbf{d} \right| - |\mathbf{y}| \right\|^2 \\ &= \frac{\left| \left(\mathbf{A}\left(\hat{\mathcal{S}},:\right)\right)^H \mathbf{d} \right|^T |\mathbf{y}|}{\left\| \left(\mathbf{A}\left(\hat{\mathcal{S}},:\right)\right)^H \mathbf{d} \right\|^2}. \end{aligned} \tag{23}$$

Finally the initial estimate $\tilde{\mathbf{h}}^{(0)}$ is given by [20]

$$\begin{cases} \tilde{\mathbf{h}}^{(0)}_{\hat{\mathcal{S}}} = \hat{r}\mathbf{d}, \\ \tilde{\mathbf{h}}^{(0)}_{[1:N]\setminus\hat{\mathcal{S}}} = \mathbf{0}. \end{cases} \tag{24}$$

2) The Proposed Initialization Method in PC-CPR: Different from CRAF, the proposed on-grid PC-CPR algorithm utilizes the partial coherence property. The proposed initialization method also consists of three steps, and it differs from the CRAF initialization only in the initial support detection. That is, instead of (16)–(17), the proposed on-grid PC-CPR computes the following test statistics based on (11):

$$Z_i = \frac{1}{M}\sum_{b=1}^{B} \left|\mathbf{A}_b\left(i,:\right)\mathbf{y}_b\right|^2, \quad i = 1,\ldots,N \tag{25}$$

and then the initial support set is determined by

$$\hat{\mathcal{S}} = \left\{i_1, i_2, \ldots, i_k | Z_{i_1} > \cdots > Z_{i_k} > Z_{i_{k+1}} > \cdots > Z_{i_N}\right\}. \tag{26}$$

Note that compared with ζ_i in (16), our proposed statistic Z_i in (25) exploits the fact that the measurements in $\mathbf{y}_b$ are in the same time frame, thus these measurements have the same phase offset. We can further compare Z_i with ζ_i as follows. Assume that the entries of $\mathbf{A}$ are zero-mean independently and identically distributed, whose second-order moment and fourth-order moments are $\mathbb{E}\{|a|^2\}$ and $\mathbb{E}\{|a|^4\}$, respectively. Then we have

$$\begin{aligned} \mathbb{E}\left(\zeta_i\right) = &\left\|\tilde{\mathbf{h}}\right\|^2 \mathbb{E}^2\left\{|a|^2\right\} + \left|\tilde{h}_i\right|^2 \left[\mathbb{E}\left\{|a|^4\right\} - \mathbb{E}^2\left\{|a|^2\right\}\right] \\ &+ \mathbb{E}\left\{|a|^2\right\}\sigma_n^2, \end{aligned} \tag{27}$$

$$\begin{aligned} \mathbb{E}\left(Z_i\right) = &\left\|\tilde{\mathbf{h}}\right\|^2 \mathbb{E}^2\left\{|a|^2\right\} + \left|\tilde{h}_i\right|^2 \left[\mathbb{E}\left\{|a|^4\right\}\right. \\ &\left. + \left(N_{\mathrm{RF}} - 2\right)\mathbb{E}^2\left\{|a|^2\right\}\right] + \mathbb{E}\left\{|a|^2\right\}\sigma_n^2. \end{aligned} \tag{28}$$

The derivations of (27)–(28) can be found in Appendices A-B. Denote $\mathcal{S}$ as the support set of $\tilde{\mathbf{h}}$. Note that since $\mathbb{E}\{|a|^4\} \ge \mathbb{E}^2\{|a|^2\}$, $\mathbb{E}(\zeta_i)$ and $\mathbb{E}(Z_i)$ will have larger values for $i \in \mathcal{S}$ than for $i \notin \mathcal{S}$.

We have the following remarks about (27)–(28). Firstly, the second terms on the right-hand sides of (27)–(28) represent the gap between the statistics corresponding to $i \in \mathcal{S}$ and $i \notin \mathcal{S}$. The larger is the gap, the better is the support detection result. Secondly, since $N_{\mathrm{RF}} > 1$, it is clear that the gap for Z is larger than that for ζ. As a result, the proposed method achieves higher reliability for the initial support detection. Thirdly, [20] assumes that the entries of $\mathbf{A}$ are complex Gaussian random variables, so that $\mathbb{E}\{|a|^4\} = 2\mathbb{E}^2\{|a|^2\}$, and the gap of ζ equals to $|\tilde{h}_i|^2\mathbb{E}^2\{|a|^2\}$ which is strictly positive. But for other sensing matrices, the gap of ζ may not be large enough. For example, if all entries have unit modulus, $\mathbb{E}\{|a|^4\} = \mathbb{E}^2\{|a|^2\}$, then the gap of ζ is zero. In contrast, the gap for Z is strictly positive for all distributions of $\mathbf{A}$. This means by exploiting the phase offset structure, the proposed method offers robust performance under different sensing matrices.

After the initial support is determined, the proposed on-grid PC-CPR algorithm follows the same procedure as CRAF to construct the initialization matrix $\mathbf{D}_{\hat{\mathcal{S}}}$ according to (18) and (21), and then to obtain the initial value $\tilde{\mathbf{h}}^{(0)}$ according to (22)–(24).

In Algorithm 1, the proposed initialization method is described in Steps 1–3. However, the initialization results $\tilde{\mathbf{h}}^{(0)}$ is only a coarse estimate. We then iteratively refine the estimate of the sparse channel vector, which is discussed in the next subsection.

B. Iterative Refinement

1) The Iterative Refinement in CRAF: In CRAF [20], given the estimate of the sparse angle-domain channel vector in the $(t-1)$-th iteration, $\tilde{\mathbf{h}}^{(t-1)}$, in the t-th iteration, the phase estimate and the sparse channel estimate are performed. First, the phase of each received sample is estimated as

$$\begin{aligned} y_m^{(t)} &= \arg\min_{y \in \mathbb{C}} \left| y - \left(\mathbf{A}\left(:,m\right)\right)^H \tilde{\mathbf{h}}^{(t-1)} \right| \quad \text{s.t. } |y| = |y_m| \\ &= |y_m| \frac{\left(\mathbf{A}\left(:,m\right)\right)^H \tilde{\mathbf{h}}^{(t-1)}}{\left| \left(\mathbf{A}\left(:,m\right)\right)^H \tilde{\mathbf{h}}^{(t-1)} \right|}, \quad m = 1,2,\ldots,M. \end{aligned} \tag{29}$$

Algorithm 1: On-grid PC-CPR Channel Estimation Algorithm.

Input: Partially coherent noisy received pilots $\mathbf{y}$ in (12), equivalent sensing matrix $\mathbf{A}$, the number of coherent blocks B, the size of coherent block N_{RF}, the number of iterations T, the preset sparsity level k, the preset coefficients λ^+, λ^-, μ, τ_w, and β in (21), (31) and (32), and the termination threshold ε.

Output: Estimated sparse angle-domain channel vector $\tilde{\mathbf{h}}_{\mathrm{est}}$.

Stage 1 (initialization):

1: Determine the initial support $\hat{\mathcal{S}}$ by (16)–(17).
2: Form the matrix $\mathbf{D}_{\hat{\mathcal{S}}}$ according to (18) and (21).
3: Calculate the initial estimate $\tilde{\mathbf{h}}^{(0)}$ by the eigen-decomposition based method in (22)–(24).

Stage 2 (iterative refinement):

4: $\mathbf{y}^{(0)} = \mathbf{y}$, $t = 0$.
5: **while** $\|\mathbf{y}^{(t)} - \mathbf{A}\tilde{\mathbf{h}}^{(t)}\| > \varepsilon$ **do**
6: $t \leftarrow t + 1$.
7: $\mathbf{y}_b^{(t)} = \frac{\mathbf{y}_b \mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)}}{|\mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)}|}, b = 1, \ldots, B.$
8: $\mathbf{y}^{(t)} = [(\mathbf{y}_1^{(t)})^H, (\mathbf{y}_2^{(t)})^H, \ldots, (\mathbf{y}_B^{(t)})^H]^H$.
9: Calculate $\tilde{\mathbf{h}}^{(t)}$ according to (31)–(32).
10: **if** $t \geq T$ **then break**
11: **end while**
12: **return** $\tilde{\mathbf{h}}_{\mathrm{est}} = \tilde{\mathbf{h}}^{(t)}$

Given the above phase estimates, the sparse channel estimation then becomes a standard CS problem:

$$\tilde{\mathbf{h}}^{(t)} = \arg\min_{\tilde{\mathbf{h}} \in \mathbb{C}^N} \left\| \mathbf{y}^{(t)} - \mathbf{A}^H \tilde{\mathbf{h}} \right\| \quad \text{s.t.} \left\| \tilde{\mathbf{h}} \right\|_0 \leq k. \tag{30}$$

A hard thresholding method using reweighted gradients is proposed in [20], given by

$$\tilde{\mathbf{h}}^{(t)} = \mathcal{H}_k \left(\tilde{\mathbf{h}}^{(t-1)} - \frac{\mu}{M} \mathbf{A} \mathrm{diag} \left(\left[w_1^{(t)}, \cdots, w_M^{(t)} \right] \right) \times \left(\mathbf{A}^H \tilde{\mathbf{h}}^{(t-1)} - \mathbf{y}^{(t)} \right) \right), \tag{31}$$

where

$$w_m^{(t)} = \max \left\{ \tau_w, \frac{\left| (\mathbf{A}(:,m))^H \tilde{\mathbf{h}}^{(t-1)} \right|}{\left| (\mathbf{A}(:,m))^H \tilde{\mathbf{h}}^{(t-1)} \right| + \beta |y_m|} \right\}, \quad m = 1, 2, \ldots, M. \tag{32}$$

Here $\mathcal{H}_k : \mathbb{C}^N \to \mathbb{C}^N$ is the hard thresholding operator that keeps the k largest elements, and sets other elements to zeros. The step size μ, the threshold τ_w, the weight β are all preset parameters. The new estimate $\tilde{\mathbf{h}}^{(t)}$ is calculated by the gradient method based on the previous estimate $\tilde{\mathbf{h}}^{(t-1)}$. $\mathcal{H}_k$ ensures the sparsity level $\|\tilde{\mathbf{h}}^{(t)}\|_0 \leq k$, and the weight $w_m^{(t)}$ is designed to increase the efficacy of the method [20]. This method is able to find a better estimate $\tilde{\mathbf{h}}^{(t)}$ in the neighborhood of $\tilde{\mathbf{h}}^{(t-1)}$ that makes the objective function in (30) smaller.

2) The Proposed Iterative Refinement in PC-CPR: Similar to CRAF, each iteration of the proposed iterative refinement also consists of phase estimation and sparse channel estimation. The difference is in the phase estimation step, where we leverage the phase structure. Since every N_{RF} measurements share the same phase offset as in (11), we can estimate this common phase offset from $\mathbf{y}_b$ instead of estimating an independent phase offset for each scalar sample y_m. To be specific, we solve the phase estimation problem for block b as follows:

$$\begin{aligned} \omega_b^{(t)} &= \arg\min_{\omega_b \in [0, 2\pi)} \left\| \mathbf{y}_b - e^{j\omega_b} \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right\|^2 \\ &= \arg\min_{\omega_b \in [0, 2\pi)} \left\{ \mathbf{y}_b^H \mathbf{y}_b + \left(\mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right)^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right. \\ &\qquad \left. -2\mathrm{Re} \left(e^{j\omega_b} \mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right) \right\} \\ &= \arg\max_{\omega_b \in [0, 2\pi)} \mathrm{Re} \left(e^{j\omega_b} \mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right) \\ &= \angle \left(\mathbf{y}_b{}^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right)^*, \quad b = 1, \ldots, B. \end{aligned} \tag{33}$$

With the estimated phase offsets, we can compensate them by

$$e^{j\omega_b^{(t)}} = \frac{\left(\mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right)^*}{\left| \mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right|}, \quad b = 1, \ldots, B. \tag{34}$$

$$\mathbf{y}_b^{(t)} = \mathbf{y}_b e^{-j\omega_b^{(t)}} = \frac{\mathbf{y}_b \mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)}}{\left| \mathbf{y}_b^H \mathbf{A}_b \tilde{\mathbf{h}}^{(t-1)} \right|}, \quad b = 1, \ldots, B. \tag{35}$$

And we let

$$\mathbf{y}^{(t)} = \left[\left(\mathbf{y}_1^{(t)} \right)^H, \left(\mathbf{y}_2^{(t)} \right)^H, \ldots, \left(\mathbf{y}_B^{(t)} \right)^H \right]^H. \tag{36}$$

Then, we adopt the same hard thresholding method using reweighted gradients in (31)–(32) to update the estimate of the sparse channel vector. Note that the proposed on-grid PC-CPR subsumes the conventional noncoherent CPR as a special case with $N_{\mathrm{RF}} = 1$. In Algorithm 1, the proposed iterative refinement is in steps 4–11. The computational complexity of Algorithm 1 is mainly determined by the matrix-vector multiplications in Step 7 and Step 9, which is $\mathcal{O}(MNT) = \mathcal{O}(MN \log(1/\varepsilon))$.

IV. Extension to Off-Grid Partially Coherent Channel Estimation

A. Off-Grid PC-CPR

Formulated in (15), the off-grid PC-CPR estimates the channel parameters including the off-grid spatial directions and the path gains of $\bar{L}$ paths. Similar to the on-grid PC-CPR, the off-grid PC-CPR algorithm also alternately refines the phase offset estimates and the channel estimates in an iterative process. Rather than refining the estimate of the sparse angular channel vector $\tilde{\mathbf{h}}$ in the on-grid PC-CPR, the off-grid method refines the estimate of all the channel parameters, i.e., $\mathbf{p} \triangleq [\alpha_1, \ldots, \alpha_{\bar{L}}, \vartheta_1, \ldots, \vartheta_{\bar{L}}, \varphi_1, \ldots, \varphi_{\bar{L}}]^T$.

The main idea of the proposed off-grid PC-CPR method is to iteratively perform phase offsets estimate and channel

parameter estimate. In particular, given the channel estimate in the $(t-1)$-th iteration, we estimate and compensate for the phase shifts following the similar method as in (35)–(36), except that $\mathbf{A}_b^H\tilde{\mathbf{h}}^{(t-1)}$ is replaced by $\mathbf{W}_b^H\mathbf{h}^{(t-1)}$. Then, given the updated measurements $\mathbf{y}^{(t)}$ and the path parameter estimates in the previous iteration $\mathbf{p}^{(t-1)}$, we update the path parameters by solving

$$\min_{\substack{\alpha_1,\ldots,\alpha_{\bar{L}}\in\mathbb{C},\\ \vartheta_1,\ldots,\vartheta_{\bar{L}},\varphi_1,\ldots,\varphi_{\bar{L}}\in[0,2\pi)}} g\left(\mathbf{p};\mathbf{y}^{(t)}\right) \triangleq \left\|\mathbf{y}^{(t)}-\mathbf{W}^H\mathbf{h}\right\|^2,$$

$$\text{s.t.}\quad \mathbf{h}=\sum_{l=1}^{\bar{L}}\alpha_l\mathbf{a}\left(\vartheta_l,\varphi_l\right), \tag{37}$$

where the new estimate $\mathbf{p}^{(t)}$ are found in the neighborhood of $\mathbf{p}^{(t-1)}$. A simple yet effective approach is the gradient descent method by repeating the following refinement for $\tau=1,\ldots,T_1$:

$$\mathbf{p}^{(t,\tau)}=\mathbf{p}^{(t,\tau-1)}-\mu\boldsymbol{\nabla}_{\mathbf{p}^*}g\left(\mathbf{p}^{(t,\tau-1)};\mathbf{y}^{(t)}\right), \tag{38}$$

where T_1 is the number of inner iterations, $\mathbf{p}^{(t,0)}=\mathbf{p}^{(t-1)}$, $\mathbf{p}^{(t)}=\mathbf{p}^{(t,T_1)}$, and $\boldsymbol{\nabla}_{\mathbf{p}^*}$ is the conjugate gradient operator, whose expression can be found in Appendix C. The gradient descent refinement (38) should be repeated for a large number of iterations before it converges to the solution of (37). However, $\mathbf{y}^{(t)}$ contains estimated phase offsets, which is not accurate. Thus, it is not efficient to spend a large number of inner iterations to find the accurate solution of (37). As an alternative, we perform (38) only once rather than a number of inner iterations, which is written as

$$\mathbf{p}^{(t)}=\mathbf{p}^{(t-1)}-\mu\boldsymbol{\nabla}_{\mathbf{p}^*}g\left(\mathbf{p}^{(t-1)};\mathbf{y}^{(t)}\right). \tag{39}$$

After the gradient descent refinement of path parameters, the t-th iteration is ended with updating the channel vector $\mathbf{h}^{(t)}$ according to the channel model (4). The above phase-offset compensation and gradient descent refinement need to be repeated for T times before we get the accurate estimates of the channel. In Algorithm 2, the iterative process is summarized in Stage 3, Steps 11–19. The computational complexity of Algorithm 2 is much higher than that of Algorithm 2, because the gradient descent refinement need to calculate the partial derivatives to all the parameters of $\bar{L}$ paths. The computational complexity to calculate such a partial derivative is $\mathcal{O}(MN)$, and the computational complexity of Algorithm 2 is thus $\mathcal{O}(MN\bar{L}T)=\mathcal{O}(MN\bar{L}\log(1/\varepsilon))$. To compare Algorithm 2 with Algorithm 1, it is clear that there is a tradeoff between the channel estimation accuracy and the computational complexity. Algorithm 2 achieves higher accuracy with a higher cost of computation.

B. Initialization for Off-Grid PC-CPR

We next discuss how to detect $\bar{L}$ paths to get the initial values of $\mathbf{p}$. We solve this problem in two stages. Firstly, we find initial estimates of the phase shifts. Secondly, the channel parameters of all paths are initialized by a successive path estimation and cancelation method.

Algorithm 2: Off-grid PC-CPR Channel Estimation Algorithm.

Input: Partially coherent noisy received pilots $\mathbf{y}$, combining matrix $\mathbf{W}$, equivalent sensing matrix $\mathbf{A}$, the number of coherent blocks B, the size of coherent block N_{RF}, the preset initialization sparsity level k, the preset number of paths $\bar{L}$, the number of refinements in the path parameter initialization T_0, the gradient descent step size μ, the number of iterations in the third stage T, and termination threshold ε.

Output: Estimated channel vector $\hat{\mathbf{h}}$.

Stage 1 (phase initialization):

1: Run steps 1–3 of Algorithm 1 to get a k-sparse initial angle-domain channel estimate $\tilde{\mathbf{h}}^{(0)}$.
2: $\mathbf{y}_b^{(1)}=\frac{\mathbf{y}_b\mathbf{y}_b^H\mathbf{W}_b^H\mathbf{h}^{(0)}}{|\mathbf{y}_b^H\mathbf{W}_b^H\mathbf{h}^{(0)}|},\ b=1,\ldots,B.$
3: $\mathbf{y}^{(1)}=[(\mathbf{y}_1^{(1)})^H,(\mathbf{y}_2^{(1)})^H,\ldots,(\mathbf{y}_B^{(1)})^H]^H.$

Stage 2 (channel parameters initialization):

4: **for** $l=1:\bar{L}$ **do**
5: Detect the l-th path and determine $\alpha_l^{(0)},\vartheta_l^{(0)},\varphi_l^{(0)}$ according to (40)–(43).
6: **for** $t=1:T_0$ **do**
7: Calculate $\alpha_l^{(t)},\vartheta_l^{(t)},\varphi_l^{(t)}$ by the refinement method in (44).
8: **end for**
9: **end for**
10: Calculate the estimate of the channel vector $\mathbf{h}^{(T_0)}$ according to (4).

Stage 3 (iterative process):

11: $\mathbf{y}^{(T_0)}=\mathbf{y}^{(1)}$, $t=T_0$.
12: **while** $\|\mathbf{y}^{(t)}-\mathbf{W}^H\mathbf{h}^{(t)}\|>\varepsilon$ **do**
13: $t\leftarrow t+1$.
14: $\mathbf{y}_b^{(t)}=\frac{\mathbf{y}_b\mathbf{y}_b^H\mathbf{W}_b^H\mathbf{h}^{(t-1)}}{|\mathbf{y}_b^H\mathbf{W}_b^H\mathbf{h}^{(t-1)}|},\ b=1,\ldots,B.$
15: $\mathbf{y}^{(t)}=[(\mathbf{y}_1^{(i)})^H,(\mathbf{y}_2^{(t)})^H,\ldots,(\mathbf{y}_B^{(t)})^H]^H.$
16: Refine the parameters of the all paths according to (39), and update the current channel estimate $\mathbf{h}^{(t)}$ according to (4).
17: **if** $t\geq T_0+T$ **then break**
18: **end while**
19: **return** $\mathbf{h}_{\mathrm{est}}=\mathbf{h}^{(t)}$

The first stage is the initialization of the phase shifts. We use the on-grid PC-CPR algorithm which has been discussed in Section III-A, to find an initial estimate $\tilde{\mathbf{h}}^{(0)}$ of the angle-domain channel vector. Then, we calculate the initial estimates of the phase offsets $\omega_1^{(1)},\ldots,\omega_B^{(1)}$ by (33). We can get the phase compensated measurements $\mathbf{y}^{(1)}$ according to (35)–(36). The first stage is Steps 1–3 in Algorithm 2.

In the second stage, the path parameters of each path are successively estimated. Given the already estimated parameters of the 1st, …, $(l-1)$-th paths, i.e., $\alpha_1{}^{(T_0)},\vartheta_1{}^{(T_0)},\varphi_1{}^{(T_0)},\ldots,\alpha_{l-1}^{(T_0)},\vartheta_{l-1}^{(T_0)},\varphi_{l-1}^{(T_0)}$, we need to estimate $\alpha_l,\vartheta_l,\varphi_l$. We subtract the contribution of the $l-1$ already estimated paths from the

phase-compensated measurements $\mathbf{y}^{(1)}$:

$$\mathbf{y}_l = \mathbf{y}^{(1)} - \mathbf{W}^H \left(\sum_{l'<l} \alpha_{l'}^{(T_0)} \mathbf{a}\left(\vartheta_{l'}^{(T_0)}, \varphi_{l'}^{(T_0)}\right) \right). \tag{40}$$

Then, we estimate the l-th path by solving

$$\begin{aligned} \alpha_l, \vartheta_l, \varphi_l &= \arg \min_{\substack{\alpha\in\mathbb{C}, \\ \vartheta,\varphi\in[0,2\pi)}} f(\alpha, \vartheta, \varphi; \mathbf{y}_l) \\ &\triangleq \left\| \mathbf{y}_l - \mathbf{W}^H \alpha \mathbf{a}(\vartheta, \varphi) \right\|^2. \end{aligned} \tag{41}$$

Given ϑ_l, φ_l, the optimal value of α_l can be expressed as:

$$\alpha_l = \frac{\mathbf{a}^H(\vartheta_l, \varphi_l) \mathbf{W} \mathbf{y}_l}{\left\| \mathbf{W}^H \mathbf{a}(\vartheta_l, \varphi_l) \right\|^2}. \tag{42}$$

We exhaustively search through the angle-domain dense grids to get the initial estimates:

$$\alpha_l^{(0)}, \vartheta_l^{(0)}, \varphi_l^{(0)} = \arg \max_{\substack{\vartheta\in\{\frac{\pi(i-1)}{N_1} | 1\le i\le 2N_1\}, \\ \varphi\in\{\frac{\pi(i-1)}{N_1} | 1\le j\le 2N_2\}}} f(\alpha, \vartheta, \varphi; \mathbf{y}_l), \text{ s.t. } (42). \tag{43}$$

The reason why we choose the $2N_1 \times 2N_2$ dense grid rather than the standard $N_1 \times N_2$ grid defined in (5) is that we can get more accurate values of $\alpha_l^{(0)}, \vartheta_l^{(0)}, \varphi_l^{(0)}$ by searching through the denser one. After that, we perform the conjugate gradient descent refinement to the parameter estimates of the single path for a few iterations:

$$\begin{bmatrix} \alpha_l \\ \vartheta_l \\ \varphi_l \end{bmatrix}^{(t)} = \begin{bmatrix} \alpha_l \\ \vartheta_l \\ \varphi_l \end{bmatrix}^{(t-1)} - \mu \begin{bmatrix} \frac{\partial f(\alpha_l, \vartheta_l, \varphi_l; \mathbf{y}_l)}{\partial \alpha_l^*} \\ \frac{\partial f(\alpha_l, \vartheta_l, \varphi_l; \mathbf{y}_l)}{\partial \vartheta_l} \\ \frac{\partial f(\alpha_l, \vartheta_l, \varphi_l; \mathbf{y}_l)}{\partial \varphi_l} \end{bmatrix}^{(t-1)}, \tag{44}$$

where μ is the step size. The first element of the derivative term is the conjugate derivative because $\alpha_l \in \mathbb{C}$. The expressions of the first-order partial derivatives in (44) are provided in Appendix D. Since we do not need very accurate initial estimates, T_0 is set to be a small number, e.g., $T_0 = 5$ in our simulations. After estimating the parameters of the l-th path, we move forward to initialize the parameter estimates of the next path until we get the initial estimates of all the $\bar{L}$ paths. We can find this stage in Steps 4–10 of Algorithm 2.

V. Simulation Results

In this section, simulation results are provided to demonstrate the performance of our proposed PC-CPR channel estimation algorithms.

A. Simulation Setup

In our simulations, we consider the 32×16 UPA, so the size of the channel vector is $N = 512$. Each data point is drawn by running 1000 Monte-Carlo instances, and for each Monte-Carlo test, we randomly generate a channel vector. To be specific, the mmWave channel is composed of 1 LoS path and $(L-1)$ NLoS paths, where the random number of NLoS paths $(L-1)$ follows Poisson distribution with parameter $\lambda = 5$. Then, the normalized path gains are generated by $\bar{\alpha}_1 \sim \mathcal{CN}(0, \kappa)$ and $\bar{\alpha}_l \sim \mathcal{CN}(0, 1)$ for $2 \le l \le L$, where $\kappa = 10$. The azimuth angles $\{\theta_l\}_{l=1}^L$ are uniformly distributed in $[0, 2\pi)$, and the elevation angles $\{\phi_l\}_{l=1}^L$ are uniformly distributed in $(-\pi/2, \pi/2)$. The phase offsets $\{\omega_b\}_{b=1}^B$ are uniformly distributed in $[0, 2\pi)$. Each element of the combining matrix is generated by $W_{i,j} = 1/\sqrt{N} e^{j\gamma_{i,j}}$, where $\gamma_{i,j}$ is uniformly distributed in $[0, 2\pi)$. The angular dictionary $\mathbf{F}$ defined in (5) is of the size 512×512. Based on the signal model (12), we can define the SNR by $\mathrm{SNR} = 10\log_{10} \frac{N_{\mathrm{TS}} P_{\mathrm{T}}}{\sigma_n^2}$ dB, which is the ratio of the expected power of effective received pilots to the power of additive noise in the effective received pilots.

Since L is random, we choose the sparsity level $k = 20$ for problems (13)–(14), and $\bar{L} = 10$ for problem (15). In Algorithm 1, the maximum number of iterations $T = 20$. We set the algorithm parameters $\lambda^+ = 1$, $\lambda^- = -1$ in (21), and $\mu = 0.2$, $\tau_w = 0.1$, $\beta = 0.1$ in (31)–(32). In Algorithm 2, the number of refinements in the path parameter initialization is $T_0 = 5$, and the maximum number of iterations $T = 100$. The step size in (39) is $\mu = 10^{-6}$. The termination threshold is $\varepsilon = 10^{-8}$.

B. Performance Metrics

We evaluate several performance metrics of channel estimation, including the normalized mean squared error (NMSE), the average beamforming loss, and the channel estimation success rate. In this paper, due to the phase noncoherence problem, the definition of NMSE is slightly different from the conventional definition in literature. According to (12), for any $\omega \in [0, 2\pi)$,

$$\mathbf{y} = \begin{bmatrix} e^{j\omega_1}\mathbf{A}_1 \\ e^{j\omega_2}\mathbf{A}_2 \\ \vdots \\ e^{j\omega_B}\mathbf{A}_B \end{bmatrix} \tilde{\mathbf{h}} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_B \end{bmatrix} = \begin{bmatrix} e^{j(\omega_1-\omega)}\mathbf{A}_1 \\ e^{j(\omega_2-\omega)}\mathbf{A}_2 \\ \vdots \\ e^{j(\omega_B-\omega)}\mathbf{A}_B \end{bmatrix} \tilde{\mathbf{h}}' + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_B \end{bmatrix}, \tag{45}$$

where $\tilde{\mathbf{h}}' = \tilde{\mathbf{h}} e^{j\omega}$ is another angle-domain channel vector. This equation means that both $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{h}}'$ can result in the same measurement vector $\mathbf{y}$, thus we cannot distinguish them from our measurements with phase offsets. Hence, the channel vector can be estimated subject to an ambiguity phase rotation term $e^{j\omega}$. This phase ambiguity will not influence the beamforming SNR gain. In the data frames, this phase ambiguity and the phase offset in each data frame lead to a phase rotation on the data. To remove the phase rotation, we can insert a single reference signal into each data frame, so the phase rotation can thereby be estimated and compensated for. So the phase ambiguity will not bring problems in data transmission. To take into account the phase ambiguity in evaluating the estimation error, the squared error between the estimated channel $\hat{\mathbf{h}}$ and the true channel $\mathbf{h}$ is defined as

$$\begin{aligned} \mathcal{E}\left(\hat{\mathbf{h}}, \mathbf{h}\right) &\triangleq \min_{\omega\in[0,2\pi)} \left\| \hat{\mathbf{h}} e^{j\omega} - \mathbf{h} \right\|^2 \\ &= \left\| \hat{\mathbf{h}} \frac{\hat{\mathbf{h}}^H \mathbf{h}}{\left| \hat{\mathbf{h}}^H \mathbf{h} \right|} - \mathbf{h} \right\|^2. \end{aligned} \tag{46}$$

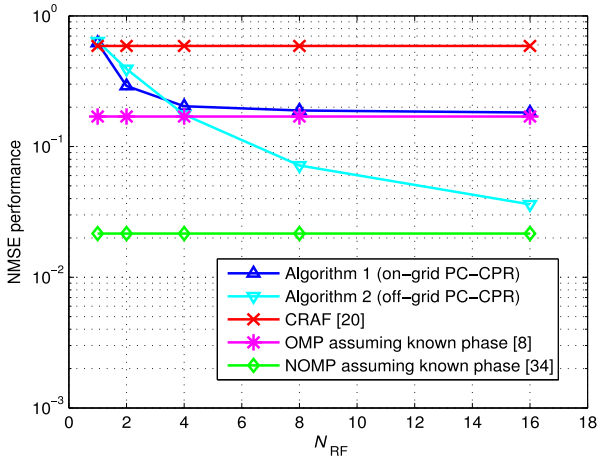

Fig. 1. NMSE for different number of RF chains, with $M = BN_{\mathrm{RF}} = 256$.

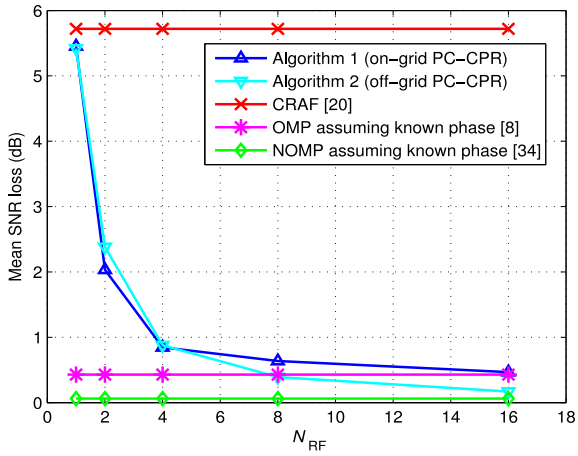

Fig. 2. Mean SNR losses for different number of RF chains, with $M = BN_{\mathrm{RF}} = 256$.

Then, we can define the NMSE by

$$\mathrm{NMSE} = \frac{\mathbb{E}\left\{\mathcal{E}\left(\hat{\mathbf{h}}, \mathbf{h}\right)\right\}}{\mathbb{E}\left\|\mathbf{h}\right\|^2}. \tag{47}$$

After the channel vector is estimated, the users feed the estimated channel back to the BS. Given the estimated channel vector $\mathbf{h}_{\mathrm{est}}$, the analog precoding vector is determined by maximizing the beamforming gain

$$\begin{aligned}\mathbf{w}\left(\mathbf{h}_{\mathrm{est}}\right) &= \arg\max_{\mathbf{w}\in\mathbb{C}^N} \left|\mathbf{w}^H\mathbf{h}_{\mathrm{est}}\right| \quad \text{s.t.}\, |w_1|, \ldots, |w_N| = \frac{1}{\sqrt{N}} \\ &= \frac{1}{\sqrt{N}} \exp \angle\left(\mathbf{h}_{\mathrm{est}}\right). \end{aligned} \tag{48}$$

The beamforming loss is calculated by subtracting the achieved beamforming gain based on the estimated channel from the maximal beamforming gain given the perfect channel knowledge, i.e., $\mathrm{loss} = 10\log|[\mathbf{w}(\mathbf{h})]^H\mathbf{h}| - 10\log|[\mathbf{w}(\mathbf{h}_{\mathrm{est}})]^H\mathbf{h}|$. We calculate the mean beamforming loss as a performance index. Similar to [14], if the beamforming loss is smaller than 3 dB, we call this channel estimation is "successful". The channel estimation success rate is also calculated to show the reliability of the channel estimation methods.

Based on the estimated channel, the BS can adopt the classic hybrid precoding algorithm in [33] to calculate the analog precoding matrix $\mathbf{P}_{\mathrm{A}}$ and the digital precoding matrix $\mathbf{P}_{\mathrm{D}}$ for a system with multiple users. Let $\mathbf{P} = \mathbf{P}_{\mathrm{A}}\mathbf{P}_{\mathrm{D}}$ whose columns $\{\mathbf{p}_u\}_{u=1}^K$ are the hybrid precoding vectors for the K users. The multi-user sum-rate is calculated by

$$R = \sum_{u=1}^{K} \log_2\left(1 + \frac{P_{\mathrm{T}}\left|\mathbf{h}_u^H\mathbf{p}_u\right|^2}{\sum_{u'\neq u} P_{\mathrm{T}}\left|\mathbf{h}_u^H\mathbf{p}_{u'}\right|^2 + \sigma_n^2}\right). \tag{49}$$

C. Results

In Figs. 1–2, the SNR is 10 dB. We fix the number of measurements $M = BN_{\mathrm{RF}} = 256$. We compare the NMSE and the mean SNR loss of different channel estimation algorithms for different number of RF chains N_{RF}. Note that we do not fix the number of time frames B. The reason is, given the same B, a larger N_{RF} leads to a larger M, then the channel estimation performances will be significantly better with more measurements. So we fix M rather than B to make the comparison more fair. Since N_{RF} is the size of a coherent block in the received pilots, Figs. 1–2 can show how the partial coherence influences the channel estimation performances. We can see from these figures that the proposed on-grid PC-CPR algorithm and off-grid PC-CPR algorithm perform better as N_{RF} increases. This is because the larger is N_{RF}, the more partial coherence can we leverage. We also compare the proposed algorithms with the conventional CPR algorithm CRAF [20] which requires no coherence, and the standard on-grid algorithm OMP [8] and off-grid algorithm NOMP [17], both of which require full phase coherence. When $N_{\mathrm{RF}} = 1$, the partial coherence vanishes, so the performances of the proposed algorithms are close to those of the noncoherent CPR algorithm (CRAF). As N_{RF} increases, the partial coherence approaches full coherence, so the performances of the on-grid/off-grid PC-CPR algorithms approach those of the coherent on-grid/off-grid channel estimation methods (OMP/NOMP), respectively. The results support that the proposed partially coherent channel estimation subsumes the conventional coherent and noncoherent channel estimation as special cases.

In Fig. 3, $N_{\mathrm{RF}} = 16$, and SNR $= 10$ dB. We run the proposed on-grid and off-grid PC-CPR algorithms with different initialization schemes, including the random initialization and the initialization in [20] without considering the partial coherence property. It is clear that the performance is poor when we adopt the random initialization. Moreover, the proposed initialization scheme is able to improve the channel estimation reliability by leveraging the partial coherence property. This figure can show the important role of the initializations in solving the nonconvex channel estimation problems in the existence of phase offsets.

In Figs. 4–5, SNR $= 10$ dB. we compare the channel estimation success rate and the NMSE performance against different number of measurements M. We compare the proposed on-grid PC-CPR algorithm, off-grid PC-CPR algorithm with CRAF [20]

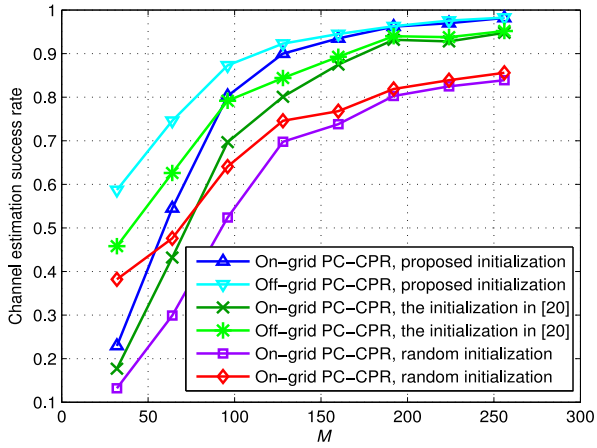

Fig. 3. Channel estimation success rate with different initialization schemes.

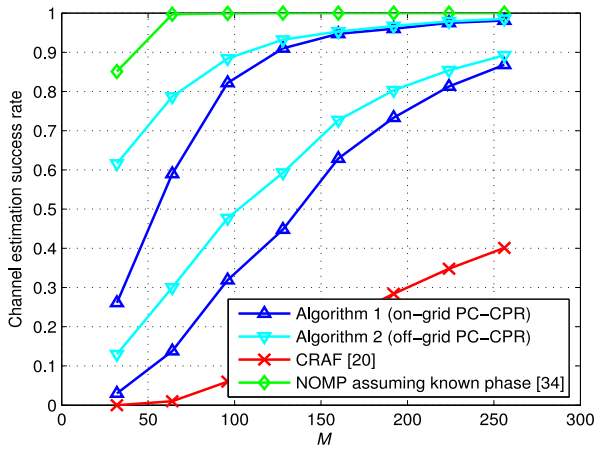

Fig. 4. Channel estimation success rate against the number of measurements for UPA.

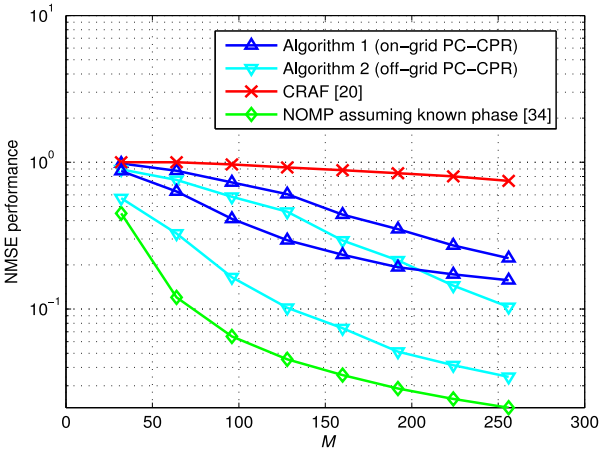

Fig. 5. NMSE performance against the number of measurements for UPA.

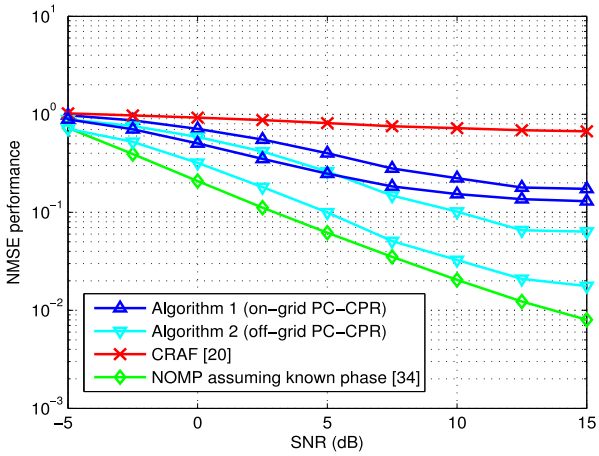

Fig. 6. NMSE performance against SNR for UPA.

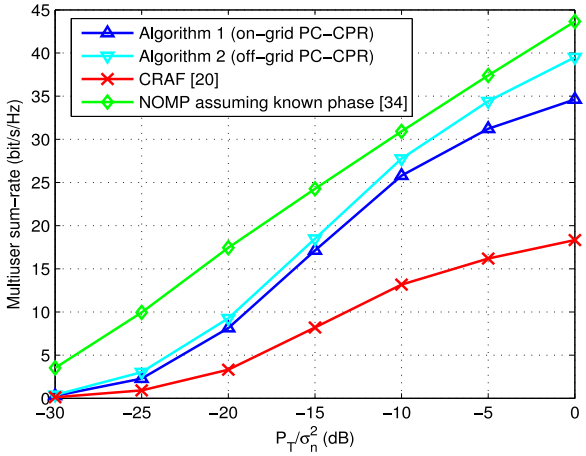

Fig. 7. Multiuser sum-rate comparison for UPA.

and NOMP [34]. Both the proposed on-grid PC-CPR channel estimation algorithm and the off-grid version can achieve higher success rate and lower NMSE than CRAF [20]. This is mainly because we can fully utilize the phase offset structure. Moreover, the success rate under the system with $N_{\mathrm{RF}} = 16$ is higher than that under $N_{\mathrm{RF}} = 4$.

Fig. 6 shows the NMSE performance of the above phase retrieval algorithms against the SNR. The number of measurements is $M = 256$. The NMSE performance gap between off-grid PC-CPR and on-grid PC-CPR becomes larger as SNR increases. This is because under high SNR, the NMSE of on-grid channel estimation is mainly restricted by the angle quantization caused by the on-grid assumption.

Fig. 7 evaluates the multiuser sum-rate. $M = 256$. The number of RF chains and the number of users are $N_{\mathrm{RF}} = K = 4$. After estimating the channels of the users, the BS can calculate the analog and digital precoding matrices to transmit data to the users. We can compare the achieved sum-rate of different channel estimation schemes against the signal SNR $P_{\mathrm{T}}/\sigma_n^2$. We can see that the proposed on-grid and off-grid PC-CPR algorithms can achieve higher sum-rate than CRAF [20].

Finally, we compare the proposed on-grid and off-grid PC-CPR algorithm with other noncoherent channel estimation algorithms, including the Fast Beam Alignment (FBA) algorithm in [14] and the noncoherent channel estimation (NoncoherentCE) algorithm in [17]. [14], [17] are proposed for uniform linear array (ULA). Since ULA can be considered as a reduced

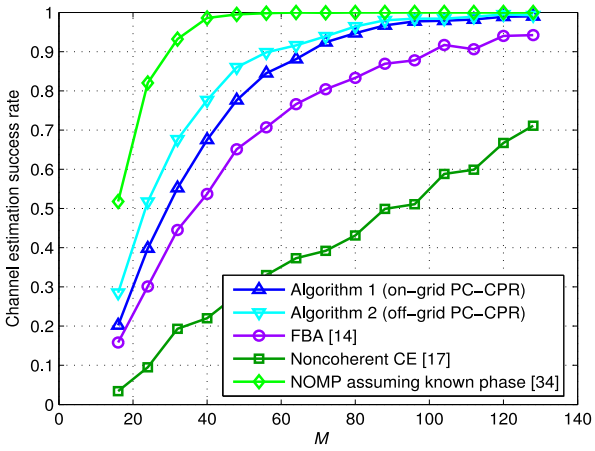

Fig. 8. Channel estimation success rate comparison for ULA.

UPA with only one row of antennas, our algorithm can also work under ULA. We simulate for a 128-antenna ULA-based mmWave massive MIMO system, where $N_{\mathrm{RF}} = 4$, and the SNR is 10 dB. From Fig. 8, we can observe that our proposed algorithms has higher success rate under the same number of measurements M. To achieve the same success rate, the required number of measurements of the proposed PC-CPR algorithms is lower than that of [14], [17].

VI. Conclusion

In this paper, we have proposed on-grid and off-grid PC-CPR algorithms to solve the partially coherent channel estimation problem in hybrid mmWave massive MIMO systems. Due to the unavoidable hardware imperfections in mmWave systems, the pilot measurements in different time frames are corrupted by random phase offsets; thus most CS techniques are not applicable. The proposed algorithms can exploit the property of hybrid mmWave system that the received pilots in different RF chains share the same phase offset in the same time frame, which is not considered in existing CPR algorithms. In the initialization stage of on-grid PC-CPR, we propose new test statistics to determine the initial support set. We show that the proposed statistics can significantly improve the initialization reliability by leveraging the phase structure. In the iterative refinement stage, the proposed on-grid PC-CPR differs from existing algorithms in the estimation of phase offsets. Furthermore, we have also proposed off-grid channel estimation to increase the estimation accuracy further. The proposed partially coherent channel estimation framework subsumes the conventional coherent and noncoherent channel estimation methods as special cases. Simulation results show that under the presence of random phase offsets where the coherent channel estimation schemes fail to work, the proposed on-grid and off-grid PC-CPR algorithms outperform noncoherent channel estimation methods with higher reliability and lower pilot overhead by leveraging the partial coherence. Besides, the off-grid PC-CPR can achieve better channel estimation accuracy than the on-grid PC-CPR.

Appendix A
Proof of (27)

Since the noise is independent to the entries of $\mathbf{A}$, by substituting (11) into (16), we have

$$\begin{aligned}
\mathbb{E}\left(\zeta_i\right) &= \frac{1}{M}\sum_{b=1}^{B}\mathbb{E}\left\|\left|\mathbf{A}_b\left(i,:\right)\right|^T \circ \left|\mathbf{A}_b^H\tilde{\mathbf{h}}e^{j\omega_b}+\mathbf{n}_b\right|\right\|^2 \\
&= \frac{1}{M}\sum_{b=1}^{B}\sum_{j=1}^{N_{\mathrm{RF}}}\mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\left|\mathbf{A}_b^H\left(:,j\right)\tilde{\mathbf{h}}\right|^2\right. \\
&\qquad \left. +\left|\mathbf{A}_b(i,j)\right|^2\left|\mathbf{n}_b(j)\right|^2\right\} \\
&= \frac{1}{M}\sum_{b=1}^{B}\sum_{j=1}^{N_{\mathrm{RF}}}\mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\mathbf{A}_b^*\left(n_1,j\right)\right. \\
&\qquad \left.\times\,\mathbf{A}_b\left(n_2,j\right)\tilde{h}_{n_1}\tilde{h}_{n_2}^*\right\} \\
&\quad +\mathbb{E}\left\{\left|a\right|^2\right\}\sigma_n^2 \\
&= \frac{1}{M}\sum_{b=1}^{B}\sum_{j=1}^{N_{\mathrm{RF}}}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\tilde{h}_{n_1}\tilde{h}_{n_2}^* \\
&\quad \times\mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\mathbf{A}_b^*\left(n_1,j\right)\mathbf{A}_b\left(n_2,j\right)\right\} \\
&\quad +\mathbb{E}\left\{\left|a\right|^2\right\}\sigma_n^2. \qquad (50)
\end{aligned}$$

On one hand, when $n_1 = n_2$, we have $n_1 = i$ or $n_1 \neq i$. If $n_1 = n_2 = i$,

$$\begin{aligned}
&\mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\mathbf{A}_b^*\left(n_1,j\right)\mathbf{A}_b\left(n_2,j\right)\right\} \\
&\quad = \mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^4\right\} = \mathbb{E}\left\{\left|a\right|^4\right\}. \qquad (51)
\end{aligned}$$

While if $n_1 = n_2 \neq i$,

$$\begin{aligned}
&\mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\mathbf{A}_b^*\left(n_1,j\right)\mathbf{A}_b\left(n_2,j\right)\right\} \\
&\quad = \mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\right\}\mathbb{E}\left\{\left|\mathbf{A}_b\left(n_1,j\right)\right|^2\right\} = \mathbb{E}^2\left\{\left|a\right|^2\right\}. \qquad (52)
\end{aligned}$$

On the other hand, when $n_1 \neq n_2$, we have $n_1 \neq i$ or $n_2 \neq i$. If $n_1 \neq i$,

$$\begin{aligned}
&\mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\mathbf{A}_b^*\left(n_1,j\right)\mathbf{A}_b\left(n_2,j\right)\right\} \\
&\quad = \mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\mathbf{A}_b\left(n_2,j\right)\right\}\mathbb{E}\left\{\mathbf{A}_b^*\left(n_1,j\right)\right\} = 0. \qquad (53)
\end{aligned}$$

While if $n_2 \neq i$, for the similar reason we also have $\mathbb{E}\{|\mathbf{A}_b(i,j)|^2\mathbf{A}_b^*(n_1,j)\mathbf{A}_b(n_2,j)\} = 0$.

Therefore, (50) becomes

$$\begin{aligned}
\mathbb{E}\left(\zeta_i\right) &= \frac{1}{M}\sum_{b=1}^{B}\sum_{j=1}^{N_{\mathrm{RF}}}\sum_{n_1=1}^{N}\left|\tilde{h}_{n_1}\right|^2 \\
&\quad \times\mathbb{E}\left\{\left|\mathbf{A}_b(i,j)\right|^2\mathbf{A}_b^*\left(n_1,j\right)\mathbf{A}_b\left(n_1,j\right)\right\} \\
&\quad +\mathbb{E}\left\{\left|a\right|^2\right\}\sigma_n^2
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{n_1=1,\\ n_1\neq i}}^{N} \left|\tilde{h}_{n_1}\right|^2 \mathbb{E}^2\left\{|a|^2\right\} + \left|\tilde{h}_i\right|^2 \mathbb{E}\left\{|a|^4\right\} \\
&+ \mathbb{E}\left\{|a|^2\right\}\sigma_n^2 \\
&= \left\|\tilde{\mathbf{h}}\right\|^2 \mathbb{E}^2\left\{|a|^2\right\} + \left|\tilde{h}_i\right|^2 \left[\mathbb{E}\left\{|a|^4\right\} - \mathbb{E}^2\left\{|a|^2\right\}\right] \\
&+ \mathbb{E}\left\{|a|^2\right\}\sigma_n^2. \qquad (54)
\end{aligned}
$$

APPENDIX B
PROOF OF (28)

By substituting (11) into (25), we have

$$
Z_i = \frac{1}{M}\sum_{b=1}^{B}\left|\mathbf{A}_b\left(i,:\right)\mathbf{A}_b^H\tilde{\mathbf{h}}e^{j\omega_b} + \mathbf{A}_b\left(i,:\right)\mathbf{n}_b\right|^2. \qquad (55)
$$

Denote $\mathbf{C}_b = \mathbf{A}_b\mathbf{A}_b^H$. Since the noise in independent of the entries of $\mathbf{A}$, we have

$$
\begin{aligned}
\mathbb{E}\left(Z_i\right) &= \frac{1}{M}\sum_{b=1}^{B}\mathbb{E}\left|\mathbf{C}_b\left(i,:\right)\tilde{\mathbf{h}}e^{j\omega_b}\right|^2 + \frac{1}{M}\sum_{b=1}^{B}\mathbb{E}\left|\mathbf{A}_b(i,:)\mathbf{n}_b\right|^2 \\
&= \frac{1}{M}\sum_{b=1}^{B}\mathbb{E}\left\{\tilde{\mathbf{h}}^H\mathbf{C}_b^H\left(i,:\right)\mathbf{C}_b\left(i,:\right)\tilde{\mathbf{h}}\right\} \\
&\quad + \frac{1}{M}\sum_{b=1}^{B}\mathbb{E}\left\|\mathbf{A}_b\left(i,:\right)\right\|^2\sigma_n^2 \\
&= \frac{1}{M}\sum_{b=1}^{B}\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}\tilde{h}_{n_1}^*\tilde{h}_{n_2}\mathbb{E}\left\{\mathbf{C}_b\left(n_1,i\right)\mathbf{C}_b\left(i,n_2\right)\right\} \\
&\quad + \mathbb{E}\left\{|a|^2\right\}\sigma_n^2. \qquad (56)
\end{aligned}
$$

On one hand, when $n_1 = n_2$, we have $n_1 = i$ or $n_1 \neq i$. If $n_1 = i$,

$$
\begin{aligned}
\mathbb{E}\left\{\mathbf{C}_b^2(i,i)\right\} &= \mathbb{E}\left\{\left[\mathbf{A}_b\left(i,:\right)\mathbf{A}_b^H\left(i,:\right)\right]^2\right\} \\
&= \mathbb{E}\left\{\left[\sum_{j=1}^{N_{\mathrm{RF}}}\left|\mathbf{A}_b(i,j)\right|^2\right]^2\right\} \\
&= \sum_{j=1}^{N_{\mathrm{RF}}}\mathbb{E}\left|\mathbf{A}_b(i,j)\right|^4 + \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{N_{\mathrm{RF}}}\mathbb{E}\left|\mathbf{A}_b\left(i,j_1\right)\right|^2\left|\mathbf{A}_b\left(i,j_2\right)\right|^2 \\
&= N_{\mathrm{RF}}\mathbb{E}\left\{|a|^4\right\} + N_{\mathrm{RF}}\left(N_{\mathrm{RF}}-1\right)\mathbb{E}^2\left\{|a|^2\right\}. \qquad (57)
\end{aligned}
$$

While if $n_1 \neq i$,

$$
\begin{aligned}
&\mathbb{E}\left\{\mathbf{C}_b\left(n_1,i\right)\mathbf{C}_b\left(i,n_1\right)\right\} \\
&= \mathbb{E}\left\{\mathbf{A}_b\left(n_1,:\right)\mathbb{E}\left\{\mathbf{A}_b^H\left(i,:\right)\mathbf{A}_b\left(i,:\right)\right\}\mathbf{A}_b^H\left(n_1,:\right)\right\} \\
&= \mathbb{E}\left\{\mathbf{A}_b\left(n_1,:\right)\left(\mathbf{I}_{N_{\mathrm{RF}}}\mathbb{E}\left\{|a|^2\right\}\right)\mathbf{A}_b^H\left(n_1,:\right)\right\} \\
&= N_{\mathrm{RF}}\mathbb{E}^2\left\{|a|^2\right\}. \qquad (58)
\end{aligned}
$$

On the other hand, when $n_1 \neq n_2$, we have $n_1 \neq i$ or $n_2 \neq i$. If $n_1 \neq i$,

$$
\begin{aligned}
&\mathbb{E}\left\{\mathbf{C}_b\left(n_1,i\right)\mathbf{C}_b\left(i,n_2\right)\right\} \\
&= \mathbb{E}\left\{\mathbf{A}_b\left(n_1,:\right)\mathbf{A}_b^H\left(i,:\right)\mathbf{A}_b\left(i,:\right)\mathbf{A}_b^H\left(n_2,:\right)\right\} \\
&\overset{(a)}{=} \mathbb{E}\left\{\mathbf{A}_b\left(n_1,:\right)\right\}\mathbb{E}\left\{\mathbf{A}_b^H\left(i,:\right)\mathbf{A}_b\left(i,:\right)\mathbf{A}_b^H\left(n_2,:\right)\right\} \\
&= 0, \qquad (59)
\end{aligned}
$$

where the equality (a) results from that $\mathbf{A}$ has i.i.d. entries. In the same way, if $n_2 \neq i$, we also have $\mathbb{E}\{\mathbf{C}_b(n_1,i)\mathbf{C}_b(i,n_2)\} = 0$.

Therefore, (56) becomes

$$
\begin{aligned}
\mathbb{E}\left(Z_i\right) &= \frac{1}{M}\sum_{b=1}^{B}\left|\tilde{h}_i\right|^2\mathbb{E}\left\{\mathbf{C}_b^2(i,i)\right\} \\
&\quad + \frac{1}{M}\sum_{b=1}^{B}\sum_{\substack{n=1\\ n\neq i}}^{N}\left|\tilde{h}_n\right|^2\mathbb{E}\left\{\mathbf{C}_b(n,i)\mathbf{C}_b(i,n)\right\} \\
&\quad + \frac{1}{M}\sum_{b=1}^{B}\sum_{\substack{n_1,n_2=1\\ n_1\neq n_2}}^{N}\tilde{h}_{n_1}^*\tilde{h}_{n_2}\mathbb{E}\left\{\mathbf{C}_b\left(n_1,i\right)\mathbf{C}_b\left(i,n_2\right)\right\} \\
&\quad + \mathbb{E}\left\{|a|^2\right\}\sigma_n^2 \\
&= \left|\tilde{h}_i\right|^2\left[\mathbb{E}\left\{|a|^4\right\} + \left(N_{\mathrm{RF}}-1\right)\mathbb{E}^2\left\{|a|^2\right\}\right] \\
&\quad + \sum_{\substack{n=1\\ n\neq i}}^{N}\left|\tilde{h}_n\right|^2\mathbb{E}^2\left\{|a|^2\right\} \\
&\quad + \mathbb{E}\left\{|a|^2\right\}\sigma_n^2 \\
&= \left\|\tilde{\mathbf{h}}\right\|^2\mathbb{E}^2\left\{|a|^2\right\} + \left|\tilde{h}_i\right|^2\left[\mathbb{E}\left\{|a|^4\right\}\right. \\
&\quad \left. + \left(N_{\mathrm{RF}}-2\right)\mathbb{E}^2\left\{|a|^2\right\}\right] \\
&\quad + \mathbb{E}\left\{|a|^2\right\}\sigma_n^2. \qquad (60)
\end{aligned}
$$

APPENDIX C
THE CONJUGATE GRADIENT OF $g(\mathbf{p};\mathbf{y})$ IN (37)

According to the definition of conjugate gradient [35],

$$
\boldsymbol{\nabla}_{\mathbf{p}^*}g = \left[\frac{\partial g}{\partial\alpha_1^*},\ldots,\frac{\partial g}{\partial\alpha_{\bar{L}}^*},\frac{\partial g}{\partial\vartheta_1},\ldots,\frac{\partial g}{\partial\vartheta_{\bar{L}}},\frac{\partial g}{\partial\varphi_1},\ldots,\frac{\partial g}{\partial\varphi_{\bar{L}}}\right]^T, \qquad (61)
$$

where

$$
\begin{aligned}
\frac{\partial g}{\partial\alpha_l^*} &= -\mathbf{a}^H\left(\vartheta_l,\varphi_l\right)\mathbf{W}\left[\mathbf{y} - \mathbf{W}^H\sum_{l'=1}^{\bar{L}}\alpha_{l'}\mathbf{a}\left(\vartheta_{l'},\varphi_{l'}\right)\right], \\
&\quad l = 1,\ldots,\bar{L}, \qquad (62)
\end{aligned}
$$

$$
\frac{\partial g}{\partial\vartheta_l} = 2\mathrm{Re}\left[\left(\mathbf{y} - \mathbf{W}^H\sum_{l'=1}^{\bar{L}}\alpha_{l'}\mathbf{a}\left(\vartheta_{l'},\varphi_{l'}\right)\right)^H\right.
$$

$$\cdot \frac{\partial}{\partial \vartheta_l}\left(\mathbf{y}-\mathbf{W}^H\sum_{l'=1}^{\bar{L}}\alpha_{l'}\mathbf{a}\left(\vartheta_{l'},\varphi_{l'}\right)\right)\Bigg]$$

$$=-2\mathrm{Re}\left[\alpha_l\left(\mathbf{y}-\mathbf{W}^H\sum_{l'=1}^{\bar{L}}\alpha_{l'}\mathbf{a}\left(\vartheta_{l'},\varphi_{l'}\right)\right)^H \times \mathbf{W}^H\frac{\partial\mathbf{a}\left(\vartheta_l,\varphi_l\right)}{\partial\vartheta_l}\right],\quad l=1,\ldots,\bar{L}, \tag{63}$$

$$\frac{\partial g}{\partial\varphi_l}=2\mathrm{Re}\left[\left(\mathbf{y}-\mathbf{W}^H\sum_{l'=1}^{\bar{L}}\alpha_{l'}\mathbf{a}\left(\vartheta_{l'},\varphi_{l'}\right)\right)^H \cdot \frac{\partial}{\partial\varphi_l}\left(\mathbf{y}-\mathbf{W}^H\sum_{l'=1}^{\bar{L}}\alpha_{l'}\mathbf{a}\left(\vartheta_{l'},\varphi_{l'}\right)\right)\Bigg]$$

$$=-2\mathrm{Re}\left[\alpha_l\left(\mathbf{y}-\mathbf{W}^H\sum_{l'=1}^{\bar{L}}\alpha_{l'}\mathbf{a}\left(\vartheta_{l'},\varphi_{l'}\right)\right)^H \times \mathbf{W}^H\frac{\partial\mathbf{a}\left(\vartheta_l,\varphi_l\right)}{\partial\varphi_l}\right],\quad l=1,\ldots,\bar{L}, \tag{64}$$

where the partial derivatives of the steering vector is given by

$$\frac{\partial\mathbf{a}\left(\vartheta,\varphi\right)}{\partial\vartheta}=\left[0je^{j0\vartheta},\ldots,\left(N_1-1\right)je^{j\left(N_1-1\right)\vartheta}\right]^T \otimes\left[e^{j0\varphi},\ldots,e^{j\left(N_2-1\right)\varphi}\right]^T, \tag{65}$$

$$\frac{\partial\mathbf{a}\left(\vartheta,\varphi\right)}{\partial\varphi}=\left[e^{j0\vartheta},\ldots,e^{j\left(N_1-1\right)\vartheta}\right]^T \otimes\left[0je^{j0\varphi},\ldots,\left(N_2-1\right)je^{j\left(N_2-1\right)\varphi}\right]^T. \tag{66}$$

APPENDIX D
THE CONJUGATE GRADIENT OF $f(\alpha,\vartheta,\varphi;\mathbf{y})$ IN (41)

Similar to Appendix C, we have

$$\frac{\partial f}{\partial\alpha^*}=-\mathbf{a}^H\left(\vartheta,\varphi\right)\mathbf{W}\left[\mathbf{y}-\mathbf{W}^H\alpha\mathbf{a}\left(\vartheta,\varphi\right)\right], \tag{67}$$

$$\frac{\partial f}{\partial\vartheta}=-2\mathrm{Re}\left[\alpha\big(\mathbf{y}-\mathbf{W}^H\alpha\mathbf{a}\left(\vartheta,\varphi\right)\big)^H\mathbf{W}^H\frac{\partial\mathbf{a}\left(\vartheta,\varphi\right)}{\partial\vartheta}\right], \tag{68}$$

$$\frac{\partial f}{\partial\varphi}=-2\mathrm{Re}\left[\alpha\big(\mathbf{y}-\mathbf{W}^H\alpha\mathbf{a}\left(\vartheta,\varphi\right)\big)^H\mathbf{W}^H\frac{\partial\mathbf{a}\left(\vartheta,\varphi\right)}{\partial\varphi}\right], \tag{69}$$

where $\frac{\partial\mathbf{a}(\vartheta,\varphi)}{\partial\vartheta}$ and $\frac{\partial\mathbf{a}(\vartheta,\varphi)}{\partial\varphi}$ follow (65) and (66), respectively.

REFERENCES

[1] S. Mumtaz, J. Rodriquez, and L. Dai, *MmWave massive MIMO: A paradigm for 5G*. New York, NY, USA: Academic, 2016.

[2] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr. 2018.

[3] 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.

[4] 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.

[5] A. Alkhateeb, G. Leus, and R. Heath, "Compressed sensing based multiuser millimeter wave systems: How many measurements are needed?" in *Proc. 40th IEEE Inter. Conf. on Acoust., Speech Signal Process.*, Brisbane, Australia, Apr. 2015, pp. 2909–2913.

[6] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-Wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

[7] A. Liu, V. K. N. Lau, and M. L. Honig, "Compressive RF training for massive MIMO with channel support side information," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3628–3641, Jul. 2019.

[8] J. Lee, G. T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[9] A. Liao, Z. Gao, H. Wang, S. Chen, M. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, Dec. 2019.

[10] N. J. Myers and R. W. Heath Jr., "A compressive channel estimation technique robust to synchronization impairments," in *Proc. IEEE Signal Process. Adv. Wireless Commun.*, Sapporo, 2017, pp. 1–5.

[11] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, "Millimeter wave communications: From point-to-point links to agile network connections," in *Proc. ACM HotNets*, ACM, 2016, pp. 169–175.

[12] J. R.-Fernandez, N. G.-Prelcic, and R. W. Heath, "Channel estimation for millimeter wave MIMO systems in the presence of CFO uncertainties," in *Proc. IEEE Int. Conf. Commun.*, Kansas City, MO, USA, 2018, pp. 1–6.

[13] N. J. Myers and R. W. Heath, "Message passing-based joint CFO and channel estimation in mmWave systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3064–3077, Jun. 2019.

[14] H. Hassanieh, O. Abari, M. Rodriguez, M. Abdelghany, D. Katabi, and P. Indyk, "Fast millimeter wave beam alignment," in *Proc. Conf. ACM Special Interest Group Data Commun.*, 2018, pp. 432–445.

[15] X. Li, J. Fang, H. Duan, Z. Chen, and H. Li, "Fast beam alignment for millimeter wave communications: A sparse encoding and phaseless decoding approach," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4402–4417, Sep. 2019.

[16] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave path tracking," in *Proc. 18th Int. Workshop Mobile Comput. Syst. Appl.*, 2017, pp. 13–18.

[17] M. E. Rasekh and U. Madhow, "Noncoherent compressive channel estimation for mmWave massive MIMO," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, 2018, pp. 889–894.

[18] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.

[19] G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen, "Sparse phase retrieval via truncated amplitude flow," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 479–491, Jan. 2018.

[20] L. Zhang, G. Wang, G. B. Giannakis, and J. Chen, "Compressive phase retrieval via reweighted amplitude flow," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5029–5040, Oct. 2018.

[21] R. Pedarsani, D. Yin, K. Lee, and K. Ramchandran, "Phasecode: Fast and efficient compressive phase retrieval based on sparse-graph codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3663–3691, Jun. 2017.

[22] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 928–938, Feb. 2014.

[23] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.
[24] *Wireless LAN Medium Access Control and Physical Layer Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Std 802.11ad-2012, Dec. 28, 2012, pp. 1–628.
[25] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.
[26] N. J. Myers, A. Mezghani, and R. W. Heath, "Swift-link: A compressive beam alignment algorithm for practical mmWave radios," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1104–1119, Feb. 2019.
[27] R. Chandra, Z. Zhong, J. Hontz, V. McCulloch, C. Studer, and T. Goldstein, "PhasePack: A phase retrieval library," in *Proc. 51st Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, 2017, pp. 1617–1621.
[28] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.
[29] Y. Tsai, L. Zheng, and X. Wang, "Millimeter-wave beamformed full-dimensional MIMO channel estimation based on atomic norm minimization," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6150–6163, Dec. 2018.
[30] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.
[31] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," *Found. Comput. Math.*, vol. 19, pp. 703–773, 2019.
[32] W. Luo, W. Alghamdi, and Y. M. Lu, "Optimal spectral initialization for signal recovery with applications to phase retrieval," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2347–2356, May 2019.
[33] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
[34] B. Mamandipoor, D. Ramasamy, and U. Madhow, "Newtonized orthogonal matching pursuit: Frequency estimation over the continuum," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5066–5081, Oct. 2016.
[35] Ken Kreutz-Delgado, "The complex gradient operator and the CR-calculus," Jun. 2009, *arXiv:0906.4835*.

Chen Hu (Student Member, IEEE) received the B.E. degree in electronic engineering in 2016 from Tsinghua University, Beijing, China, where he is currently working toward the Ph.D. degree in electronic engineering. His research interests include sparse signal processing and mmWave massive MIMO, with the emphasis on channel estimation. He was the recipient of the Freshman Scholarship of Tsinghua University in 2012, the Excellent Thesis Award of Tsinghua University in 2016, and the IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer Award in 2018.

Xiaodong Wang (Fellow, IEEE) received the Ph.D degree in electrical engineering from Princeton University, Princeton, NJ, USA. He is currently a Professor of electrical engineering with Columbia University, New York, NY, USA. He has authored or coauthored several research papers on the topics of his research interests, which include general areas of computing, signal processing, and communications. Among his publications is a book entitled *Wireless Communication Systems: Advanced Techniques for Signal Reception* (Prentice-Hall, 2003). His current research interests include wireless communications, statistical signal processing, and genomic signal processing. Dr. Wang was the recipient of the 1999 NSF CAREER Award, the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2011 IEEE Communication Society Award for Outstanding Paper on New Communication Topics. He was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON INFORMATION THEORY. He is listed as an ISI highly-cited author.

Linglong Dai (Senior Member, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2003, the M.S. degree (with the highest Hons.) from the China Academy of Telecommunications Technology, Beijing, China, in 2006, and the Ph.D. degree (with the highest Hons.) from Tsinghua University, Beijing, China, in 2011. From 2011 to 2013, he was a Postdoctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, where he was an Assistant Professor from 2013 to 2016 and has been an Associate Professor since 2016. He has made a large amount of simulation code publicly available. He has coauthored the book *MmWave Massive MIMO: A Paradigm for 5G* (Academic Press, 2016). He also holds 16 granted patents. His current research interests include massive MIMO, millimeter-wave/THz communications, NOMA, reconfigurable intelligent surface (RIS), and machine learning for wireless communications, with emphasis on reproducible research. Dr. Dai was the recipient of five IEEE Best Paper Awards at the IEEE ICC 2013, the IEEE ICC 2014, the IEEE ICC 2017, the IEEE VTC 2017-Fall, and the IEEE ICC 2018. He was also the recipient of the Tsinghua University Outstanding Ph.D. Graduate Award in 2011, the Beijing Excellent Doctoral Dissertation Award in 2012, the China National Excellent Doctoral Dissertation Nomination Award in 2013, the URSI Young Scientist Award in 2014, the IEEE TRANSACTIONS ON BROADCASTING Best Paper Award in 2015, the *Electronics Letters* Best Paper Award in 2016, the National Natural Science Foundation of China for Outstanding Young Scholars in 2017, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2017, the IEEE ComSoc Asia-Pacific Outstanding Paper Award in 2018, and the *China Communications* Best Paper Award in 2019. He is an Area Editor of the IEEE COMMUNICATIONS LETTERS, and an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He has authored or coauthored more than 60 IEEE journal papers and more than 40 IEEE conference papers.

Junjie Ma received the B.E. degree from Xidian University, Xi'an, China, in 2010, and the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2015. He was a Postdoctoral Researcher with the City University of Hong Kong, from 2015 to 2016, and with Columbia University, New York, NY, USA, from 2016 to 2019. Since July 2019, he has been a Postdoctoral Fellow with the Department of Electrical Engineering, Harvard University, Cambridge, MA, USA. His current research interests include message passing algorithms and their applications for high-dimensional signal processing.