

Channel Estimation for RIS Assisted Millimeter Wave Systems via OMP with Optimization

You You, Yufei Xue, *Student Member, IEEE*, Li Zhang, *Senior Member, IEEE*,
Xiaohu You, *Fellow, IEEE*, and Chuan Zhang, *Senior Member, IEEE*

Abstract—Reconfigurable intelligence surface (RIS) can be deployed to assist the communications in millimeter wave (mmWave) systems. Employing the sparsity of the mmWave channel, the compressive sensing (CS) techniques can be leveraged to reduce the pilot overhead of the channel estimation (CE). However, conventional CS-based algorithms are based on the discrete grids and the difference between the real continuous angle and its nearest grid point is called off-grid error. Off-grid errors seriously deteriorate the CE performance. In this paper, we propose the orthogonal matching pursuit with discrete-continuous optimization (DC-OMP) method for RIS assisted mmWave systems to mitigate the impact of the off-grid errors. Simulation results show that compared with existing works, the proposed DC-OMP can efficiently mitigate the impact of the off-grid errors without adding much complexity.

Index Terms—Reconfigurable intelligence surface (RIS), channel estimation (CE), compressive sensing (CS), orthogonal matching pursuit (OMP), optimization.

I. INTRODUCTION

THE reconfigurable intelligence surface (RIS) is regarded as a promising technology to overcome the high propagation loss of millimeter wave (mmWave) communications [1]. To unleash the potential of RIS assisted mmWave systems, accurate channel state information (CSI) is required. However, the pilot overhead in the channel estimation (CE) is too high due to the large number of passive elements without signal processing capabilities at the RIS. Although conventional on-grid compressive sensing (CS) technologies can be leveraged to reduce the pilot overhead in the CE. The difference between the real continuous angle and its nearest grid point (named as the off-grid error) seriously deteriorates the CS based CE performance.

CS-based RIS-assisted mmWave CE techniques can be classified into on-grid CE [2, 3], off-grid CE [4, 5] and

gridless CE [6]. A cascaded channel model was proposed in [2] by exploiting the inherent sparse structure of the cascaded channel and conventional orthogonal matching pursuit (OMP) is used for CE. In [3], double-structured OMP (DS-OMP) was proposed by further analyzing the double-structure sparsity of cascaded channel. However, both OMP and DS-OMP ignore the off-grid errors. To mitigate the impact of the off-grid errors, discrete Fourier transform (DFT)-OMP [5] was proposed as a two-stage CE scheme and part of the angle parameters are optimized in the first stage. Gridless CE methods such as the atomic norm minimization (ANM) method [6] are proposed to radically avoid the off-grid errors. Specifically, ANM estimates angle parameters by searching infinite atomic set, which achieves more accurate CSI compared with on-grid and off-grid methods. However, both DFT-OMP and ANM are two-stage CE methods. Only part of the angles are estimated in the first stage. As a result, the error propagation significantly deteriorates the CE performance. Moreover, ANM [6] recovers angle parameters by searching infinite atomic set, which costs unacceptable complexity.

In this paper, we propose OMP with discrete-continuous optimization (DC-OMP) for RIS-assisted mmWave systems to mitigate the impact of the off-grid errors with relatively low complexity. Contributions of this paper are summarized as follows:

- 1). Impact of the off-grid errors on the RIS assisted mmWave CE is analyzed. Then DC-OMP is proposed to mitigate the CE performance deterioration due to the off-grid errors in the RIS assisted mmWave systems. First, OMP with a coarse grid is employed as an initial estimation of all the angle parameters. Then the optimization of angle parameters is formulated into a nonlinear constrained problem which can be solved by classical convex optimization methods.
- 2). Simulation results demonstrate that the proposed method outperforms other existing CS-based off-grid CE methods considering off-grid errors in terms of the pilot overhead, CE accuracy and complexity. Compared with gridless CE methods, DC-OMP achieves slightly worse normalized mean square error (NMSE) performance but with far less pilot overhead and complexity.

Notation: Lower-case and upper-case boldface letters \mathbf{a} and \mathbf{A} denote a vector and a matrix, respectively; \mathbf{a}^T denotes the conjugate of vector \mathbf{a} ; \mathbf{A}^* , \mathbf{A}^T and \mathbf{A}^H denote the complex conjugate, transpose and conjugate transpose of matrix \mathbf{A} , respectively; $\text{diag}(\mathbf{x})$ denotes the diagonal matrix with the

Copyright © 2023 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. This work was supported in part by National Key R&D Program of China under Grant 2020YFB2205503, in part by CPSF under Grants 2022M722433, in part by NSFC under Grants 62122020 and 61871115, in part by the Jiangsu Provincial NSF under Grant BK20211512, in part by the Major Key Project of PCL under Grant PCL2021A01-2, in part by Jiangsu Excellent Postdoctoral Program, and in part by the Fundamental Research Funds for the Central Universities. (You You and Yufei Xue contributed equally to this work.) (Corresponding author: Chuan Zhang.)

Y. You, Y. Xue, X. You, and C. Zhang are with the LEADS, the National Mobile Communications Research Laboratory, and the Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 211189, China, and also with the Purple Mountain Laboratories, Nanjing 211100, China. (email: chzhang@seu.edu.cn).

L. Zhang is with School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, U.K..

vector \mathbf{x} on its diagonal; $\text{vec}(\mathbf{A})$ and $\text{vec}^{-1}(\mathbf{a})$ denote the vectorized \mathbf{A} and anti-vectorized \mathbf{a} ; $\text{reshape}(\mathbf{A}, p, q)$ makes \mathbf{A} a $p \times q$ matrix; $\langle \mathbf{a}, \mathbf{b} \rangle$, $\mathbf{a} \otimes \mathbf{b}$ and $\mathbf{a} \bullet \mathbf{b}$ denotes the inner product, Kronecker product and Khatri-Rao product of \mathbf{a} and \mathbf{b} . The symbols $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the 2-norm and Frobenius-norm, respectively.

II. SYSTEM MODEL

A. Channel Model

We consider a RIS assisted mmWave system with an M -elements ($M = M_x \times M_y$) uniform planar array (UPA) RIS, an N -antennas uniform linear array (ULA) base station (BS) and a single antenna user. Specifically, the BS-to-RIS channel $\mathbf{G} \in \mathbb{C}^{M \times N}$ can be expressed as [2]

$$\mathbf{G} = \sqrt{\frac{MN}{L_G}} \sum_{l_1=1}^{L_G} \delta_{l_1} \mathbf{a}_r(\vartheta'_{l_1}, \gamma'_{l_1}) \mathbf{a}_t^H(\phi'_{l_1}), \quad (1)$$

where L_G denotes the number of paths between the RIS and BS, δ_{l_1} represents the complex gain for the l_1 -th path, $\gamma'_{l_1} = \cos(\gamma_{l_1})$, $\vartheta'_{l_1} = \sin(\gamma_{l_1}) \cos(\vartheta_{l_1})$ and $\phi'_{l_1} = \cos(\phi_{l_1})$. $\vartheta_{l_1}(\gamma_{l_1})$ and $\vartheta'_{l_1}(\gamma'_{l_1})$ denotes the azimuth (elevation) physical and spatial angle of arrival (AoA) at the RIS respectively. ϕ_{l_1} and ϕ'_{l_1} are the physical and spatial angle of departure (AoD) at the BS respectively, $\mathbf{a}_r(\cdot)$ and $\mathbf{a}_t(\cdot)$ represent the array steering vectors associated to the RIS and BS, respectively. For the N -antennas ULA, $\mathbf{a}_t(\phi'_{l_1})$ can be represented as

$$\mathbf{a}_t(\phi'_{l_1}) = \frac{1}{\sqrt{N}} \left[e^{j2\pi\phi'_{l_1} \frac{d}{\lambda} \mathbf{n}} \right]^T, \quad (2)$$

where $\mathbf{n} = [0 \ 1 \ \dots \ N-1]$, λ is the carrier wavelength, d is the antennas spacing usually satisfying $d = \lambda/2$. For the M -elements UPA, $\mathbf{a}_r(\vartheta'_{l_1}, \gamma'_{l_1})$ can be written as

$$\mathbf{a}_r(\vartheta'_{l_1}, \gamma'_{l_1}) = \frac{1}{\sqrt{M}} \left[e^{j2\pi\gamma'_{l_1} \frac{d}{\lambda} \mathbf{m}_x} \otimes e^{j2\pi\vartheta'_{l_1} \frac{d}{\lambda} \mathbf{m}_y} \right]^T, \quad (3)$$

where $\mathbf{m}_x = [0 \ 1 \ \dots \ M_x-1]$, $\mathbf{m}_y = [0 \ 1 \ \dots \ M_y-1]$. Similarly, the RIS-to-user channel $\mathbf{h} \in \mathbb{C}^M$ can be expressed as

$$\mathbf{h} = \sqrt{\frac{M}{L_h}} \sum_{l_2=1}^{L_h} \xi_{l_2} \mathbf{a}_r(\vartheta'_{l_2}, \gamma'_{l_2}), \quad (4)$$

where L_h denotes the number of paths between the RIS and user, $\gamma'_{l_2} = \cos(\gamma_{l_2})$, $\vartheta'_{l_2} = \sin(\gamma_{l_2}) \cos(\vartheta_{l_2})$, $\vartheta_{l_2}(\gamma_{l_2})$ and $\vartheta'_{l_2}(\gamma'_{l_2})$ denotes the azimuth (elevation) physical and spatial AoD at the RIS, respectively.

To explore the channel sparsity in angular domain, virtual channel representation [7] is used to express \mathbf{G} without considering the off-grid errors as

$$\mathbf{G} = (\mathbf{D}_x \otimes \mathbf{D}_y) \mathbf{A} \mathbf{D}_h^H = \mathbf{D}_g \mathbf{A} \mathbf{D}_h^H, \quad (5)$$

where $\mathbf{D}_x \in \mathbb{C}^{M_x \times M_{G,x}}$ ($\mathbf{D}_y \in \mathbb{C}^{M_y \times M_{G,y}}$) is the dictionary of the spatial elevation (azimuth) angles at the RIS and each of its column has a form of $\frac{1}{\sqrt{M_x}} [e^{j2\pi\gamma'_i \frac{d}{\lambda} \mathbf{m}_x}]^T \left(\frac{1}{\sqrt{M_y}} [e^{j2\pi\vartheta'_j \frac{d}{\lambda} \mathbf{m}_y}]^T \right)$, with γ'_i ($i = 1, 2, \dots, M_{G,x}$) and ϑ'_j ($j = 1, 2, \dots, M_{G,y}$) chosen from discrete grids \mathbf{T}_x and \mathbf{T}_y , where $M_{G,x}$ and $M_{G,y}$ are grid points

number (grid size) of \mathbf{T}_x and \mathbf{T}_y , respectively. Specifically, \mathbf{T}_x and \mathbf{T}_y are defined as

$$\begin{aligned} \mathbf{T}_x &= [-1 \ -1 + \frac{2}{M_{G,x}} \ \dots \ 1 - \frac{2}{M_{G,x}}], \\ \mathbf{T}_y &= [-1 \ -1 + \frac{2}{M_{G,y}} \ \dots \ 1 - \frac{2}{M_{G,y}}]. \end{aligned} \quad (6)$$

$\mathbf{D}_h \in \mathbb{C}^{N_G \times N}$ is defined in the similar way as \mathbf{D}_x and \mathbf{D}_y . Each column in \mathbf{D}_h has a form of $\mathbf{a}_t(\phi'_k)$ with ϕ'_k ($k = 1, 2, \dots, N_G$) chosen from discrete grid \mathbf{T}_h , where N_G is the grid points number (grid size) of \mathbf{T}_h . And \mathbf{T}_h is defined as

$$\mathbf{T}_h = [-1 \ -1 + \frac{2}{N_G} \ \dots \ 1 - \frac{2}{N_G}]. \quad (7)$$

$\mathbf{D}_g = \mathbf{D}_x \otimes \mathbf{D}_y \in \mathbb{C}^{M \times M_G}$, and $\mathbf{A} \in \mathbb{C}^{M_G \times N_G}$ is a sparse matrix with L_G nonzero entries, where $M_G = M_{G,x} \times M_{G,y}$.

Considering the limited scattering characteristics [2], \mathbf{h} can be written as

$$\mathbf{h} = \mathbf{D}_g \mathbf{\Sigma}, \quad (8)$$

where $\mathbf{\Sigma} \in \mathbb{C}^{M_G}$ is a sparse vector with L_h nonzero entries.

However, off-grid errors can not be ignored when the virtual channel representation is used. Accordingly, considering the continuous angles, a more accurate virtual channel representation of \mathbf{G} and \mathbf{h} can be presented as $\hat{\mathbf{G}}$ and $\hat{\mathbf{h}}$ as

$$\hat{\mathbf{G}} = \mathbf{G} + \mathbf{E}_G, \quad \hat{\mathbf{h}} = \mathbf{h} + \mathbf{E}_h, \quad (9)$$

where \mathbf{E}_G and \mathbf{E}_h are the off-grid errors matrices in the BS-to-RIS channel and RIS-to-user channel.

B. Problem Formulation

The received signal y_t at the t -th time slot can be written as

$$\begin{aligned} y_t &= \hat{\mathbf{h}}^H \text{diag}(\mathbf{v}_t^H) \hat{\mathbf{G}} \mathbf{w}_t s_t + n_t \\ &= \mathbf{v}_t^H \text{diag}(\mathbf{h}^H) \mathbf{G} \mathbf{w}_t s_t + n_t + e_t, \end{aligned} \quad (10)$$

where \mathbf{v}_t and \mathbf{w}_t are the phase shift vector and the precoding vector, respectively. \mathbf{v}_t can be optimized by maximizing the transmitting rate and received signal power [8]. In this letter, without loss of generality, we assume that the phase shift vector is random and time-varying in the CE. n_t denotes the noise in the t -th time slots and e_t is induced by the off-grid errors (\mathbf{E}_G and \mathbf{E}_h). $\text{diag}(\mathbf{h}^H) \mathbf{G}$ is defined as the cascaded channel \mathbf{H} as

$$\begin{aligned} \mathbf{H} &= \text{diag}(\mathbf{h}^H) \mathbf{G} = \mathbf{h}^* \bullet \mathbf{G} \\ &\stackrel{(a)}{=} (\mathbf{D}_g \mathbf{\Sigma})^* \bullet (\mathbf{D}_g \mathbf{A} \mathbf{D}_h^H) \\ &\stackrel{(b)}{=} (\mathbf{D}_g^* \bullet \mathbf{D}_g) (\mathbf{\Sigma}^* \otimes (\mathbf{A} \mathbf{D}_h^H)) \\ &\stackrel{(c)}{=} (\mathbf{D}_g^* \bullet \mathbf{D}_g) (\mathbf{\Sigma}^* \otimes \mathbf{\Lambda}) (1 \otimes \mathbf{D}_h^H) \\ &\stackrel{(d)}{=} \mathbf{U} (\mathbf{\Sigma}^* \otimes \mathbf{\Lambda}) \mathbf{D}_h^H \\ &\stackrel{(e)}{=} \mathbf{U}_s \mathbf{H}_{CS} \mathbf{D}_h^H, \end{aligned} \quad (11)$$

where (a) employs (5) and (8), (b) follows the property of Khatri-Rao product, i.e., $(\mathbf{A}\mathbf{B}) \bullet (\mathbf{C}\mathbf{D}) = (\mathbf{A} \bullet \mathbf{C})(\mathbf{B} \otimes \mathbf{D})$, (c) employs a principle of Kronecker product, i.e., $\mathbf{A}\mathbf{B} \otimes \mathbf{C}\mathbf{D} = (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D})$. In (d), $\mathbf{U} = \mathbf{D}_g^* \bullet \mathbf{D}_g \in \mathbb{C}^{M \times M_G^2}$ only

contains M_G distinct columns which are the first M_G columns of \mathbf{U} [2]. In (e), $\mathbf{H}_{CS} = \mathbf{\Sigma}^* \otimes \mathbf{A}$, $\mathbf{U}_s = \mathbf{U}(:, 1 : M_G)$ denotes the matrix constructed by the first M_G columns of \mathbf{U} . In this work, we focus on the cascaded channel estimation. The direct channel between the BS and UE can be easily estimated by turning off the RIS and employing various mature conventional channel estimation methods, and thus it is ignored in this work.

Assuming the pilot signal $s_t = 1$ ($t = 1, 2, \dots, T$), the received signal at the t -th time slot can be rewritten as

$$\begin{aligned} y_t &= \mathbf{v}_t^H \mathbf{H} \mathbf{w}_t + n_t + e_t \\ &\stackrel{(a)}{=} (\mathbf{w}_t^T \otimes \mathbf{v}_t^H) \text{vec}(\mathbf{H}) + n_t + e_t \\ &\stackrel{(b)}{=} (\mathbf{w}_t^T \otimes \mathbf{v}_t^H) (\mathbf{D}_h^* \otimes \mathbf{U}_s) \text{vec}(\mathbf{H}_{CS}) + n_t + e_t. \end{aligned} \quad (12)$$

In (a) and (b), $\text{vec}(\mathbf{M}\mathbf{X}\mathbf{N}) = (\mathbf{N}^T \otimes \mathbf{M}) \text{vec}(\mathbf{X})$ is employed for y_t and \mathbf{H} , respectively. After T pilots transmission, measurement matrix $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_T]^T$ can be represented as

$$\begin{aligned} \mathbf{y} &\stackrel{(a)}{=} \mathbf{W}_v (\mathbf{D}_h^* \otimes \mathbf{U}_s) \text{vec}(\mathbf{H}_{CS}) + \mathbf{N} + \mathbf{E} \\ &\stackrel{(b)}{=} \mathbf{\Psi} \mathbf{x} + \mathbf{N} + \mathbf{E}, \end{aligned} \quad (13)$$

where in (a) $\mathbf{N} = [n_1 \ n_2 \ \dots \ n_T]$, $\mathbf{E} = [e_1 \ e_2 \ \dots \ e_T]$ and

$$\mathbf{W}_v = \begin{pmatrix} \mathbf{w}_1^T \otimes \mathbf{v}_1^H \\ \vdots \\ \mathbf{w}_T^T \otimes \mathbf{v}_T^H \end{pmatrix}. \quad (14)$$

In (b), $\mathbf{\Psi} = \mathbf{W}_v (\mathbf{D}_h^* \otimes \mathbf{U}_s)$, $\mathbf{x} = \text{vec}(\mathbf{H}_{CS})$. The CE problem has been converted into a sparse signal recovery problem and conventional CS-based algorithms can be employed to estimate the \mathbf{H}_{CS} with high probability [2]. And \mathbf{H} can be further recovered according to (11).

However, as (13), the CE is not only affected by the noise \mathbf{N} , but also by the off-grid errors \mathbf{E} . Intuitively, the off-grid errors can be mitigated by simply increasing the grid size. However, the complexity is increasing significantly with the expansion of grid size in the CS-based algorithms. In addition, the increasing coherence between columns in $\mathbf{\Psi}$ may lead to the even worse CE performance.

III. PROPOSED METHODS

A. Impact of the off-grid errors on the CE

In order to show the impact of the off-grid errors on the RIS assisted mmWave CE, the CE performance of conventional OMP with different levels of off-grid errors are shown in Fig. 1. Specifically, the same grid size is assumed as $M_{G,x} = M_{G,y} = 10$ and $N_G = 64$ for spatial angles γ' , ϑ' and ϕ' , so that the grid intervals are $2/M_{G,x}$, $2/M_{G,y}$ and $2/N_G$ as shown in (6) and (7), respectively. For OMP OFF0, OMP OFF1, OMP OFF2, OMP OFF3 and OMP OFF4, the actual continuous spatial angles of γ' , ϑ' and ϕ' are generated at the points which are 0, 1/16, 1/8, 1/4 and 1/2 times of the corresponding grid interval ($2/M_{G,x}$, $2/M_{G,y}$ and $2/N_G$) away from the nearest grid points. It can be seen that OMP OFF0 achieves the best CE performance without the error floor at high SNRs. CE performances become much worse with larger off-grid errors and there are error floors for OMP OFF1, OMP OFF2, OMP OFF3 and OMP OFF4 at high SNRs.

Therefore, for on-grid based techniques, the impact of the off-grid errors can not be ignored and the mitigation methods are needed to improve the CE performance.

B. Description of the proposed DC-OMP algorithm

To mitigate the impact of off-grid errors efficiently, DC-OMP is proposed as a discrete-to-continuous optimization method and summarized in Algorithm 1.

Algorithm 1 DC-OMP

Input: \mathbf{y} , $\mathbf{\Psi}$ and sparsity S
Output: estimated channel $\tilde{\mathbf{H}}$

- 1: Initialization: residual $\mathbf{u}_0 = \mathbf{y}$ and index set $\Upsilon = \emptyset$
- 2: **for** $s = 1, 2, \dots, S$ **do**
- 3: $v_s = \arg \max_{j=1, \dots, N} |\langle \mathbf{u}_{s-1}, \mathbf{k}_j \rangle|$
- 4: Calculate $\{\phi'_s, \vartheta'_s, \gamma'_s\}$ by Algorithm 2.
- 5: $\{\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s\} = \arg \max_{\phi'_s, \vartheta'_s, \gamma'_s} f(\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s)$
- 6: $\tilde{\mathbf{k}}_{v_s} = \tilde{\mathbf{k}}_{v_s}(\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s)$
- 7: $\mathbf{\Omega}_s = [\mathbf{\Omega}_{s-1}, \tilde{\mathbf{k}}_{v_s}]$
- 8: $\mathbf{\Upsilon}_s = [\mathbf{\Upsilon}_{s-1}, v_s]$
- 9: $\tilde{\mathbf{D}}_h(:, A, s) = \mathbf{a}'_t(\tilde{\phi}'_s)$
- 10: $\tilde{\mathbf{U}}_s(:, B, s) = \text{diag}(\mathbf{D}_g(:, 1)^T) \mathbf{a}'_r(\tilde{\vartheta}'_s, \tilde{\gamma}'_s)$
- 11: $\mathbf{x}_s = \arg \min \|\mathbf{y} - \mathbf{\Omega}_s \mathbf{x}\|_2$
- 12: $\mathbf{u}_s = \mathbf{y} - \mathbf{\Omega}_s \mathbf{x}_s$
- 13: **end for**
- 14: **for** $s = 1, 2, \dots, S$ **do**
- 15: $\mathbf{X}(1, \Upsilon_s(s), s) = \mathbf{x}_s(s)$
- 16: $\tilde{\mathbf{H}}_{CS}(:, :, s) = \text{reshape}(\mathbf{X}(:, :, s), M_G, N_G)$
- 17: **if** $s = 1$ **then**
- 18: $\tilde{\mathbf{H}} = \tilde{\mathbf{U}}_s(:, :, s) \tilde{\mathbf{H}}_{CS}(:, :, s) \tilde{\mathbf{D}}_h^H(:, :, s)$
- 19: **end if**
- 20: $\tilde{\mathbf{H}} = \tilde{\mathbf{H}} + \tilde{\mathbf{U}}_s(:, :, s) \tilde{\mathbf{H}}_{CS}(:, :, s) \tilde{\mathbf{D}}_h^H(:, :, s)$
- 21: **end for**
- 22: **return** $\tilde{\mathbf{H}}$

In step 3-5, the initial estimated angles based on the discrete grids are optimized based on the continuous dictionary. Specifically, in step 3, as in conventional OMP, the most correlated column in the sensing matrix with the received signal can be found. The most relevant v_s -th column \mathbf{k}_{v_s} in $\mathbf{\Psi}$ can be formulated as

$$\mathbf{k}_{v_s}(\phi'_s, \vartheta'_s, \gamma'_s) = \mathbf{W}_v \{ \mathbf{a}'_t(\phi'_s) \otimes [\text{diag}(\mathbf{D}_h^*(:, 1)) \mathbf{a}'_r(\vartheta'_s, \gamma'_s)] \}. \quad (15)$$

In step 4, the initial estimated spatial angle group $\{\phi'_s, \vartheta'_s, \gamma'_s\}$ corresponding to \mathbf{k}_{v_s} can be found by Algorithm 2 according to the relationship between the angle group and the column index. In Algorithm 2, A , C and D are defined as the index of spatial angles ϕ'_s , ϑ'_s and γ'_s in grids \mathbf{T}_h , \mathbf{T}_x and \mathbf{T}_y , respectively.

In step 5, the initial estimated angle group $\{\phi'_s, \vartheta'_s, \gamma'_s\}$ from Algorithm 2 is further optimized to mitigate the impact of the off-grid errors. It is known that the largest coherence corresponding to the v_s -th column is only corresponding to the

Algorithm 2 Angle group $\{\phi'_s, \vartheta'_s, \gamma'_s\}$ estimation

Input: $v_s, \mathbf{T}_h, \mathbf{T}_x$ and \mathbf{T}_y
Output: $\{\phi'_s, \vartheta'_s, \gamma'_s\}$

- 1: $B = \text{mod}(v_s, M_G)$
- 2: **if** $B \neq 0$ **then**
- 3: $A = \frac{v_s - B}{M_G} + 1$; $D = \text{mod}(B, M_{G,y})$
- 4: **if** $D \neq 0$ **then**
- 5: $C = \frac{B - D}{M_{G,y}} + 1$
- 6: **else**
- 7: $C = \frac{B}{M_{G,y}}$; $D = M_{G,y}$
- 8: **end if**
- 9: **else**
- 10: $A = \frac{v_s}{M_G}$; $C = M_{G,x}$; $D = M_{G,y}$
- 11: **end if**
- 12: $\gamma'_s = \mathbf{T}_y(D)$; $\vartheta'_s = \mathbf{T}_x(C)$; $\phi'_s = \mathbf{T}_h(A)$
- 13: **return** $\{\phi'_s, \vartheta'_s, \gamma'_s\}$;

discrete grid points. In fact, if continuous angles are considered, there may exist a larger coherence which corresponds to the continuous angles near the discrete grid points (not exactly on the grids). Thus, to mitigate the off-grid errors, the optimized angle group $\{\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s\}$ can be obtained by solving a nonlinear constrained continuous optimization problem. The objective function can be defined as the coherence as

$$f(\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s) = |\langle \mathbf{k}_{v_s}(\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s), \mathbf{u}_{s-1} \rangle|. \quad (16)$$

Then, the nonlinear constrained continuous optimization problem can be formulated as

$$\begin{aligned} & \max_{\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s} f(\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s) \\ & \text{s.t.} \begin{cases} |\phi'_s - \tilde{\phi}'_s| \leq \frac{1}{2N_G} \\ |\vartheta'_s - \tilde{\vartheta}'_s| \leq \frac{1}{2M_{G,x}} \\ |\gamma'_s - \tilde{\gamma}'_s| \leq \frac{1}{2M_{G,y}}. \end{cases} \end{aligned} \quad (17)$$

Then, some classical convex optimization methods such as SQP and the Newton method can be employed to find the optimized angle group $\{\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s\}$ efficiently. Steps 6 to 21 are explained as follows:

- In step 6: the initial chosen column $\mathbf{k}_{v_s}(\phi'_s, \vartheta'_s, \gamma'_s)$ is updated to $\mathbf{k}_{v_s}(\tilde{\phi}'_s, \tilde{\vartheta}'_s, \tilde{\gamma}'_s)$. In this way, the sensing matrix corresponding to the discrete grids is updated to a sensing matrix with continuous angles.
- In step 7-8: the index set and chosen column set are updated by $\Omega_s = [\Omega_{s-1}, \tilde{\mathbf{k}}_{v_s}]$ and $\Upsilon_s = [\Upsilon_{s-1}, v_s]$.
- In step 9-10: the three-dimension (3D) dictionaries are updated as $\tilde{\mathbf{D}}_h(:, A, s) = \mathbf{a}'_t(\phi'_s)$ and $\tilde{\mathbf{U}}_s(:, B, s) = \text{diag}(\mathbf{D}_g(:, 1)^T) \mathbf{a}'_t(\vartheta'_s, \gamma'_s)$.
- In step 11-12: the cascaded channel gains and the updated residuals are calculated, respectively.
- In step 14-21: the sparse cascaded channel $\tilde{\mathbf{H}}_{CS}$ can be constructed as $\tilde{\mathbf{H}}_{CS} = \text{vec}^{-1}(\mathbf{x}_s)$ after step 13. And $\tilde{\mathbf{H}}_{CS}$ is further constructed to \mathbf{H} in angular domain through some matrix transformations. $\tilde{\mathbf{U}}_s$, $\tilde{\mathbf{H}}_{CS}$ and $\tilde{\mathbf{D}}_h$ are all 3D matrix, where $\tilde{\mathbf{U}}_s(x, y, z)$ denotes the x -th row, y -th column and z -th layer element in $\tilde{\mathbf{U}}_s$, with

$\tilde{\mathbf{H}}_{CS}$ and $\tilde{\mathbf{D}}_h$ following the same rule. Each layer of $\tilde{\mathbf{H}}_{CS}$ contains a non-zero point and its value is the cascaded channel gain and its column index and row index indicates the corresponding the column vector in $\tilde{\mathbf{U}}_s$ and the row vector in $\tilde{\mathbf{D}}_h$. In the s -th iteration of step 14-21, one cascaded channel are calculated as $\tilde{\mathbf{U}}_s(:, :, s) \tilde{\mathbf{H}}_{CS}(:, :, s) \tilde{\mathbf{D}}_h^H(:, :, s)$. For S -sparse matrix $\tilde{\mathbf{H}}_{CS}$, \mathbf{H} in the angular domain can be fully recovered after S times iteration.

IV. SIMULATION RESULTS

In this section, the simulation results are presented to evaluate the performance of our proposed methods. All simulation results are averaged over 1000 channel realizations. The number of elements at the RIS and number of antennas at the BS are set as $M = 64$ and $N = 64$ respectively with grid size $M_G = 100$ and $N_G = 64$. The channel gains δ_{l_1} and ξ_{l_2} follow $\mathcal{CN}(0, 1)$ and the paths number $L_G = L_h = 3$. All physical angles including ϑ , γ and ϕ are randomly generated from $(-\frac{\pi}{2}, \frac{\pi}{2}]$. NMSE is defined as

$$\text{NMSE} = \mathbb{E} \left[\|\tilde{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2 \right], \quad (18)$$

and SNR is defined as $1/\sigma^2$ with the noise $n_t \sim \mathcal{CN}(0, \sigma^2)$. In this section, the proposed DC-OMP is compared with the existing works including the on-grid method: OMP [2], discrete optimization method: two-grid OMP [9], two-stage methods: DFT-OMP [5] and gridless method ANM [6].

A. Performance Analysis

In Fig. 2, NMSE performances with varying SNR and 220 pilots are shown. It can be found that the proposed DC-OMP can significantly improve the CE performance compared with OMP. Two-grid OMP shows superiority to OMP but the optimized estimated angles are still chosen from the discrete grids. So the proposed DC-OMP with continuous optimization outperforms. DFT-OMP performs better than two-grid OMP but worse than DC-OMP. Because DFT-OMP is a two-stage CE method which experiences some performance loss due to the error propagation. Then the combination of other angle parameters and channel gains are estimated in the second stage by the OMP estimator. As a result, DFT-OMP is worse than DC-OMP which optimizes all angle parameters together. ANM completely avoids the off-grid errors and is able to achieve remarkable CE performance at high SNR ranges by searching the angles in an infinite set. But the unaffordable complexity makes it impractical for RIS assisted mmWave CE. Moreover, ANM is also a two-stage CE method which suffers from the error propagation. It is worth noticing that DC-OMP has slightly worse performance in low SNR ranges because the stronger noise has a negative impact on the selection of the most correlated column, thus leading to lower tolerance to noises compared with DFT-OMP and ANM.

In Fig. 3, NMSE performances are shown when SNR = 10 dB with varying T . It can be found that the NMSE of DC-OMP, OMP, two-grid OMP and ANM decrease progressively with the increasing T . DC-OMP can greatly reduce NMSE

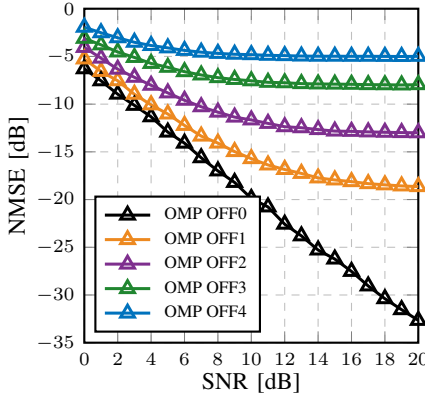


Fig. 1. NMSE performance v.s. SNR [dB] with the different off-grid errors

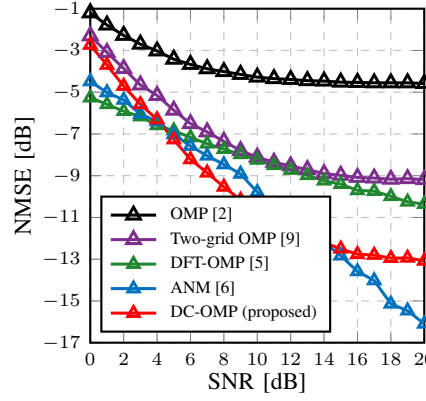


Fig. 2. NMSE performance v.s. SNR [dB]

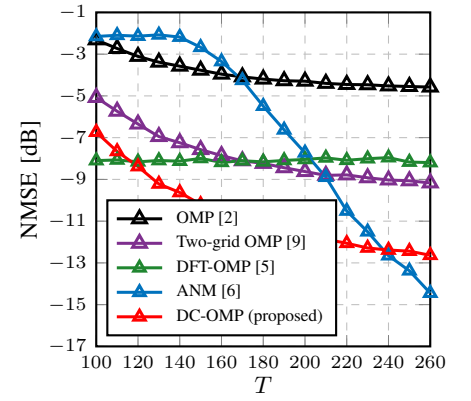


Fig. 3. NMSE performance v.s. pilots T

compared with OMP and save more than 50% pilots compared with two-grid OMP when $\text{NMSE} = -8\text{dB}$. As a two-stage CE method, DFT-OMP requires fewer overall pilots (pilots used in two stages) to reach the error floor and achieves better NMSE CE performance in low- T regime. However, as Fig.3 shown, the error floor of DFT-OMP appears early at the low- T regime with much worse achievable NMSE CE performance because of the error propagation between the two stages. As a result, compared with DFT-OMP, DC-OMP can achieve more remarkable NMSE CE performance with more pilots. It can also be found that the CE performance of ANM extremely relies on the accuracy of the AoD estimated in the first stage. As a result, if an insufficient number of pilots are used in the first stage, ANM may fail to estimate the AoD at BS. Thus, more pilots should be considered in the first stage of ANM to guarantee a very accurate estimation. Fig. 3 shows that, to achieve a desirable NMSE performance, the proposed DC-OMP can significantly reduce the pilot overhead compared with the benchmarks except for ANM.

B. Complexity Analysis

For complexity analysis, considering that DC-OMP and OMP have the same order of complexity, we use runtime calculated by MATLAB R2021a to provide a rough comparison between them ($\text{SNR} = 10\text{dB}$ and $T = 220$). The runtime of OMP is assumed as 1, the corresponding results of other benchmarks can be found in Table I. It can be found that the complexity of DC-OMP is slightly higher than the benchmarks (except for ANM, ANM has a lot higher computational complexity) but with the remarkable performance improvement.

V. CONCLUSION

In this paper, we proposed a cascaded channel estimation method DC-OMP for RIS assisted mmWave systems to mitigate the impact of the off-grid errors. Simulation results demonstrate that the proposed DC-OMP can successfully mitigate the off-grid problem. In comparison with existing methods, it is able to achieve significantly improved CE performance without introducing much computational complexity.

TABLE I
COMPLEXITY OF THE PROPOSED METHOD AND BENCHMARKS

Algorithm	Complexity	Runtime
OMP [2]	$\mathcal{O}(TN_G M_G)$	1.0
Two-grid OMP [9]	$\mathcal{O}(TNM + TN_G M_G)$	8.5
DFT-OMP [5]	$\mathcal{O}(N^2 + 8DL_G)^a$	1.0
ANM [6]	$\mathcal{O}((N + T_{\text{stage-1}})^{3.5})^b$	45.0
DC-OMP (proposed)	$\mathcal{O}(TN_G M_G)$	11.7

^a D ($D \gg M$) is the overgrid size used in the second stage;

^b $T_{\text{stage-1}}$ ($T_{\text{stage-1}} \gg L_G$) is the pilots number used in the first stage.

REFERENCES

- [1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Nov. 2020.
- [2] P. Wang, J. Fang, H. Duan *et al.*, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Processing Lett.*, vol. 27, pp. 905–909, May 2020.
- [3] X. Wei, D. Shen, and L. Dai, "Channel estimation for RIS assisted wireless communications part II: An improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, Jan. 2021.
- [4] M. Jian and Y. Zhao, "A modified off-grid SBL channel estimation and transmission strategy for RIS-assisted wireless communication systems," in *Proc. Int. Wireless Commun. and Mobile Comput. (IWCMC)*, Limassol, Cyprus, 2020, pp. 1848–1853.
- [5] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave systems," *IEEE Trans. Signal Processing*, vol. 70, pp. 1478–1492, 2022.
- [6] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmwave MIMO systems via atomic norm minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5786–5797, 2021.
- [7] A. Sayeed, "Deconstructing multi antenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.
- [8] J. An, C. Xu, L. Gan, and L. Hanzo, "Low-complexity channel estimation and passive beamforming for RIS-assisted MIMO systems relying on discrete phase shifts," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1245–1260, 2022.
- [9] J. Lee, G.-T. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, 2014, pp. 3326–3331.