Article:
You, Y., Chen, F., Zhang, L. orcid.org/0000-0002-4535-3200 et al. (2 more authors) (2024) Adaptive Channel Estimation for RIS-Assisted Systems in Time-Varying mmWave Channels. IEEE Transactions on Communications. ISSN 0090-6778

https://doi.org/10.1109/tcomm.2024.3425592

Adaptive Channel Estimation for RIS-Assisted Systems in Time-Varying mmWave Channels

You You, Fengyu Chen, Li Zhang, *Senior Member, IEEE*, Yongming Huang, *Senior Member, IEEE*, and Chuan Zhang, *Senior Member, IEEE*

***Abstract*—To improve channel estimation (CE) for reconfigurable intelligent surface (RIS)-assisted systems in time-varying mmWave channels, this paper proposes a two-stage adaptive CE scheme. This is the first attempt to develop a CE scheme without assumptions on specific timescales for channel variations. In the first stage, the adaptive scheme incorporates the estimation of partial channel state information and a channel status check process. The introduced check process can monitor the changing status of the channels and provide information for the second stage. In the second stage, based on the results from the check process, the adaptive scheme adaptively selects from two proposed candidate CE algorithms: Two-Phase orthogonal matching pursuit (TP-OMP) and Structured-Shift OMP (SS-OMP). Simulation results show that both TP-OMP and SS-OMP can reduce pilot overhead by around 33%, and respectively lower the computational complexity of existing works by about 55% and 65%. Additionally, the check process obtains an accuracy rate of approximately 92% so that the proposed CE scheme can maintain stable CE performance in time-varying channels.**

***Index Terms*—Reconfigurable intelligent surface, channel estimation, compressive sensing, orthogonal matching pursuit.**

I. Introduction

COMPRISED of numerous reflecting elements, reconfigurable intelligent surface (RIS) can dynamically alter the characteristics of impinging signals, thereby creating a controllable wireless transmission environment [1], [2]. To cater to different scenarios, various types of RISs have been designed, including fully passive RISs with minimal power consumption, active RISs with amplifiers to mitigate double path loss attenuation [3], and hybrid RISs that strike a balance between resource cost and functionality [4]. Regardless of the RIS type, accurate acquisition of channel state information (CSI) is crucial to fully exploit the potential of RISs. However, channel estimation (CE) poses significant challenges in RIS-assisted millimeter wave (mmWave) systems due to the large number of RIS elements and their limited signal processing abilities. Additionally, the need for frequent and time-consuming CE across multiple channel coherence blocks introduces substantial computational complexity and transmission delays. Consequently, developing efficient CE schemes for RIS-assisted systems operating in time-varying mmWave channels becomes imperative.

This work was supported in part by National Key R&D Program of China under Grant 2020YFB2205503, in part by NSFC under Grants 62122020, 62331009, and 62001108, in part by the Jiangsu Provincial NSF under Grants BK20211512 and BM2023016, in part by the Major Key Project of PCL under Grant PCL2023AS1-4, and in part by the Fundamental Research Funds for the Central Universities. *(You You and Fengyu Chen contributed equally to this paper.) (Corresponding author: Chuan Zhang.)*

You You, Fengyu Chen, Yongming Huang, and Chuan 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 Purple Mountain Laboratories, Nanjing 211189, China. (email: chzhang@seu.edu.cn)

Li Zhang is with the School of Electronic and Electrical Engineering, University of Leeds, LS2 9JT Leeds, U.K.(email: l.x.zhang@leeds.ac.uk)

A. Prior Work

Previous studies have proposed various methods to address the challenge of CE in a single block [5], [6]. The least squares (LS) algorithms have been applied either by sequentially activating individual RIS elements to estimate the channel [7] or by activating all RIS elements simultaneously [8]. To further alleviate the unbearable pilot overhead, element-grouping methods have been proposed [9]. By employing structured patterns for the pilots and RIS configurations, the received signals can be modeled as parallel factor tensors. The Khatri-Rao factorization and bilinear alternating LS (BALS) algorithms are then adopted in [10].

Leveraging the sparsity of mmWave channels in the angular domain, CE for RIS-assisted communications has been formulated as an ℓ_1 norm regularized optimization problem with fixed-rank constraints [11], enabling the utilization of alternating minimization and manifold optimization techniques. In addition, CE can be cast as sparse signal recovery problems [11], [12]. Various Bayesian-based compressive sensing (CS) techniques, such as sparse Bayesian learning (SBL) [13] and fast Bayesian matching pursuit (FBMP) [14] can be employed. However, the high computational complexity and parameter dependence limit their applicability. With a much lower complexity, greedy methods like orthogonal matching pursuit (OMP) [12] are widely exploited in the CE for RIS-assisted systems. Building upon the distinctive structures in RIS-assisted channels, improvements have been made to conventional OMP to enhance CE performance [15]–[20]. For example, a double-structured OMP (DS-OMP) was proposed to exploit the double-structured sparsity of cascaded channels among multiple users [15]. The authors in [16] discovered a common offset parameter shared by each user, leading to the development of a multi-user (MU) triple-structured CS-based CE algorithm. In [17], [18], the CSI of a typical user was obtained by exploiting the correlation between the channels of multiple users. For the remaining users, the re-parameterized user-RIS channels were estimated instead of the entire cascaded channels. To mitigate the error propagation, an ambiguous RIS-BS channel was first constructed in [19], followed by estimations of each ambiguous user equipment (UE)-RIS channel to obtain the full CSI. Nonetheless, the

abovementioned approaches still encounter challenges related to high pilot overhead or computational complexity, hindering further improvements in CE efficiency within a single channel coherence block.

Additionally, diverse CE schemes have been proposed to tackle the issues of high computational complexity and transmission delay in time-varying RIS-assisted channels. These schemes leverage the characteristics of RIS-assisted mmWave channels under certain assumptions [21]–[25]. In [21], a two-timescale CE framework was introduced, assuming quasi-static RIS-BS channels and more frequently changing UE-RIS channels. In [22], a three-stage angular domain CE (TAD-CE) scheme was proposed to divide the RIS-assisted channels into static RIS-BS angles, quasi-static UE-RIS angles, and time-varying channel gains. The TAD-CE scheme then estimates these three components in diverse frequencies with reduced pilots [22]. To alleviate the dependence on frequent CE when designing the reflection patterns of RIS, the quality of service (QoS) of UE was used as a metric to determine the frequency of UE re-localization and RIS configuration in [23]. Different CE schemes for semi-passive RIS [24] and anchor-assisted [25] CE approaches have also been proposed, albeit with higher energy consumption. While the aforementioned multi-block CE strategies can decrease pilot overhead in the long term, they all rely on exact prior information about some attributes of the time-varying channels. For instance, all of these works assume the duration (or relative duration) of the unchanged RIS-BS channel to be known.

B. Motivations

Retrospecting the existing literature, it is evident that formidable obstacles persist in current research on CE for RIS-assisted mmWave systems. Firstly, the existing CS-based methods have not fully leveraged the inherent structural characteristics of the cascaded channel matrix. Additionally, most current approaches heavily rely on ideal assumptions regarding the availability of channel variation information among blocks, which restricts their practical applicability. When these assumptions are not met, the existing CE schemes fail to achieve the expected performance or may even encounter failures. Hence, further research is necessary to develop more practical CE schemes accompanied by suitable algorithms.

C. Main Contributions

In this paper, we propose a CE scheme that can adaptively choose specific CE algorithms in different blocks by leveraging various channel properties. To the best of the authors' knowledge, this is the first attempt to tackle the CE in RIS-assisted mmWave systems without any prior information on the time-varying channels. The main contributions of this work are summarized as follows:

- The structural characteristics of the cascaded uplink UE-RIS-BS channel are revealed based on the sparsity and virtual angular domain (VAD) representations of the RIS-assisted mmWave channels. These analyzed structural features provide the theoretical foundation for the design of an adaptive CE scheme and two CE algorithms.
- A two-stage multi-block adaptive CE scheme is proposed. The proposed scheme does not rely on any knowledge of the channel variation but dynamically selects an appropriate CE algorithm from two proposed ones for each block. This selection process is based on a novel channel status check process.
- In the first stage, a check process is proposed to detect the channel variations across the multiple blocks. The angles of arrival (AoAs) at the BS are first estimated. By comparing the obtained partial CSI and that from the previous blocks, the check process can detect whether the RIS-BS channel has changed. Importantly, this process does not incur extra computational complexity or pilots. Simulation results demonstrate that the proposed check process achieves an accuracy rate of about 92%.
- In the second stage, two candidate CE algorithms are proposed. The Two-Phase OMP (TP-OMP) leverages the structure induced by the common paths of the UE-RIS channel within a single block. It is designed to handle CE regardless of whether the RIS-BS channel has changed. Additionally, we introduce a low-complexity algorithm named Structured-Shift OMP (SS-OMP) for CE when the RIS-BS channel remains unchanged from previous blocks. Simulation results show that compared to existing methods, both TP-OMP and SS-OMP can reduce pilots by approximately 30% with respectively 55% and 65% lower computational complexity.
- From a long-term perspective, the proposed CE scheme can be adaptive to the time-varying conditions of the channel. With a well-designed check process and the switching mechanism of the candidate CE algorithms, simulation results demonstrate the stable performance of the adaptive CE scheme across multi-block CE.

The remainder of this paper is organized as follows. In Section II, we introduce the system model. In Section III, we sequentially analyze three structures of the cascaded channel matrix. In Section IV, detailed descriptions of the proposed two-stage adaptive CE scheme are presented. Extensions of the proposed methods to mitigate the quantization errors are provided in Section V. Simulation results are depicted and analyzed in Section VI. In Section VII, we conclude this work.

Notation: Upper-case and lower-case boldface letters denote matrices and vectors, respectively. $(\cdot)^{\mathsf{T}}$, $(\cdot)^*$, $(\cdot)^{\mathsf{H}}$, and $(\cdot)^\dagger$ represent the transpose, conjugate, conjugate transpose and pseudo-inverse. $(\mathbf{A})_{:,n}$ and $(\mathbf{A})_{m,:}$ respectively represent the n-th column and m-th row of $\mathbf{A}$. $\|\cdot\|_F$ and $|\cdot|$ denote the Frobenius norm and absolute value. The Kronecker product is represented by $\otimes$. $\mathcal{CN}(\mu, \sigma^2)$ denotes the circular symmetric complex Gaussian distribution with its mean being μ and variance being σ^2. $\mathtt{diag}(\mathbf{a})$ denotes the diagonal matrix with elements of $\mathbf{a}$ on its diagonal. $\mathtt{mod}(a, b)$ returns the remainder after the division a/b. $\mathtt{unvec}_{\{x,y\}}(\mathbf{a})$ returns the $x \times y$ matrix whose elements are taken columnwisely from $\mathbf{a}$.

II. System Model

A. Channel Model

We consider the uplink CE in a RIS-assisted mmWave system. A RIS with $M = M_x \times M_y$ passive elements

is equipped to enhance the communication between an N-antenna BS and a single-antenna UE. The RIS and BS are assumed to have uniform planar array (UPA) elements and uniform linear array (ULA) antenna, respectively. The direct UE-BS channel is neglected in this work. The channels are assumed to be constant within a single block.

The classical Saleh-Valenzuela model is adopted to characterize the channels. The RIS-BS channel $\mathbf{G} \in \mathbb{C}^{N\times M}$ and the UE-RIS channel $\mathbf{f} \in \mathbb{C}^{M\times 1}$ can be respectively formulated as

$$\begin{aligned}\mathbf{G} &= \sqrt{\frac{MN}{L_G}} \sum_{l_g=1}^{L_G} \alpha_{l_g} \mathbf{b}_N(\theta_{l_g}^{G_r}) \mathbf{a}_M(\upsilon_{l_g}^{G_t}, \rho_{l_g}^{G_t})^{\mathsf{T}}, \\ \mathbf{f} &= \sqrt{\frac{M}{L_f}} \sum_{l_f=1}^{L_f} \beta_{l_f} \mathbf{a}_M(\upsilon_{l_f}^{r}, \rho_{l_f}^{r}),\end{aligned} \tag{1}$$

where L_G, L_f denote the path number of $\mathbf{G}$ and $\mathbf{f}$, respectively. $\alpha_{l_g}, \theta_{l_g}^{G_r}, \upsilon_{l_g}^{G_t}, \rho_{l_g}^{G_t}$ represent the complex gain, the spatial AoA at the BS, the spatial azimuth angle of departure (AoD) at the RIS, and the spatial elevation AoD at the RIS for the l_g-th path. $\beta_{l_f}, \upsilon_{l_f}^{r}(\rho_{l_f}^{r})$ represent the complex gain, the spatial azimuth (elevation) AoA at the RIS for the l_f-th path. $\mathbf{b}_N$ and $\mathbf{a}_M$ are the normalized array steering vectors for BS and RIS. $\mathbf{b}_N$ and $\mathbf{a}_M$ can be represented as

$$\begin{aligned}\mathbf{b}_N(\theta) &= \frac{1}{\sqrt{N}}[e^{j2\pi\theta\mathbf{n}}]^{\mathsf{T}}, \\ \mathbf{a}_M(\upsilon,\rho) &= \mathbf{b}_{M_x}(\upsilon) \otimes \mathbf{b}_{M_y}(\rho),\end{aligned} \tag{2}$$

where $\mathbf{n} = [0, \cdots, N-1]$. With the carrier wavelength being λ, and the antenna spacing being d ($d = \lambda/2$ is assumed in this paper), the relationship between the spatial angles and their corresponding physical angles are

$$\begin{aligned}\theta &= \frac{d\,\cos(\theta_{phy})}{\lambda}, \upsilon = \frac{d\,\cos(\upsilon_{phy})}{\lambda}, \\ \rho &= \frac{d\,\cos(\rho_{phy})\sin(\upsilon_{phy})}{\lambda},\end{aligned} \tag{3}$$

where θ_{phy}, υ_{phy} and ρ_{phy} are respectively the corresponding physical angles at the BS, the physical azimuth angle at the RIS, and the physical elevation angle at the RIS.

Considering the mmWave communication scenario, the commonly used linear cascaded channel model is adopted in this paper, i.e., the cascaded UE-RIS-BS channel is defined as $\mathbf{H} \triangleq \mathbf{G}\mathtt{diag}(\mathbf{f})$.[1] By using the VAD representations, $\mathbf{H} \in \mathbb{C}^{N\times M}$ can be decomposed as

$$\mathbf{H} = \mathbf{U}_N \bar{\mathbf{H}} \mathbf{U}_M^{\mathsf{T}}, \tag{4}$$

where $\mathbf{U}_N \in \mathbb{C}^{N\times N_G}$ and $\mathbf{U}_M \in \mathbb{C}^{M\times M_G}$ are the dictionary matrices at the BS and the RIS.

$$\begin{aligned}\mathbf{U}_N &= [\mathbf{b}_N(\theta_1), \cdots, \mathbf{b}_N(\theta_{N_G})], \\ \mathbf{U}_M &= \mathbf{U}_{M_x} \otimes \mathbf{U}_{M_y} \\ &= [\mathbf{b}_{M_x}(\upsilon_1), \cdots, \mathbf{b}_{M_x}(\upsilon_{M_{G_x}})] \\ &\quad \otimes [\mathbf{b}_{M_y}(\rho_1), \cdots, \mathbf{b}_{M_y}(\rho_{M_{G_y}})],\end{aligned} \tag{5}$$

[1]There are scenarios that the linear cascaded channel assumption may not hold when taking the mutual coupling effect into account [26], [27]. But the linear cascaded channel model remains applicable when the reverberation is weak at higher frequencies (such as mmWave in this paper) due to stronger absorption effects.

where N_G, M_{G_x}, M_{G_y} denote the number of the predetermined grids, which usually satisfy $N_G \geq N, M_{G_x} \geq M_x, M_{G_y} \geq M_y$. $\theta_n = -\frac{1}{2} + \frac{(n-1)}{N_G}, n \in [1, N_G]$. $\upsilon_{m_x} = -\frac{1}{2} + \frac{(m_x-1)}{M_{G_x}}$, $\rho_{m_y} = -\frac{1}{2} + \frac{(m_y-1)}{M_{G_y}}$, $m_x \in [1, M_{G_x}]$, $m_y \in [1, M_{G_y}]$, $M_G = M_{G_x} M_{G_y}$. $\bar{\mathbf{H}}$ is the sparse cascaded channel matrix in the angular domain, which contains a few non-zero elements due to the limited paths of mmWave channels. Note that the off-grid errors can not be avoided when uniformly distributed grids are used to characterize the continuous angles. Detailed discussions of this off-grid error will be provided in SectinoV.

B. Problem Formulation

The received signal at the BS in the t-th timeslot can be represented as:

$$\mathbf{y}_t = \mathbf{G}\mathtt{diag}(\boldsymbol{\omega}_t)\mathbf{f}x_t + \mathbf{n}_t = \mathbf{H}\boldsymbol{\omega}_t x_t + \mathbf{n}_t, \tag{6}$$

where x_t is the pilot symbol sent by UE, $\boldsymbol{\omega}_t \in \mathbb{C}^M$ is the reflecting coeffcients of RIS at the t-th timeslot. $\mathbf{n}_t \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise (AWGN) at the BS. For simplicity, we assume that $x_t = 1$. Over T timeslots, the received signal $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_T] \in \mathbb{C}^{N\times T}$ is

$$\mathbf{Y} = \mathbf{H}\mathbf{W} + \mathbf{N} \overset{(a)}{=} \mathbf{U}_N \bar{\mathbf{H}} \mathbf{U}_M^{\mathsf{T}} \mathbf{W} + \mathbf{N}, \tag{7}$$

where $\mathbf{W} = [\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_T]$, $\mathbf{N} = [\mathbf{n}_1, \cdots, \mathbf{n}_T]$. (a) is obtained by applying (4). Let $\tilde{\mathbf{Y}} = (\mathbf{U}_N^{\mathsf{H}}\mathbf{Y})^{\mathsf{H}}$, $\tilde{\mathbf{W}} = (\mathbf{U}_M^{\mathsf{T}}\mathbf{W})^{\mathsf{H}}$, $\tilde{\mathbf{H}} = \bar{\mathbf{H}}^{\mathsf{H}}$ and $\tilde{\mathbf{N}} = (\mathbf{U}_N^{\mathsf{H}}\mathbf{N})^{\mathsf{H}}$, (7) can then be rewritten as a sparse signal recovery model

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{W}}\tilde{\mathbf{H}} + \tilde{\mathbf{N}}, \tag{8}$$

where $\tilde{\mathbf{Y}} \in \mathbb{C}^{T\times N_G}$ is the measurement matrix, $\tilde{\mathbf{W}} \in \mathbb{C}^{T\times M_G}$ is the sensing matrix, and $\tilde{\mathbf{H}} \in \mathbb{C}^{M_G\times N_G}$ is the sparse signal.

III. Structure Analysis of the Sparse Channel

In this section, structures of $\tilde{\mathbf{H}}$ (referred to as the sparse channel in the following content) are analyzed and summarized. Firstly, we unveiled the relationship between the indices of the non-zero elements in $\tilde{\mathbf{H}}$ and the angles of the channels. Then, three structural features of $\tilde{\mathbf{H}}$, namely column-structure, row-structure A, and row-structure B are analyzed successively for further utilization. In addition, a showcase is presented to illustrate these structures.

As proven in [28], $\mathbf{H}$ can be represented as

$$\begin{aligned}\mathbf{H} =& \sqrt{\frac{MN}{L_G}} \sum_{l_g=1}^{L_G} \alpha_{l_g} \mathbf{b}_N(\theta_{l_g}^{G_r}) \mathbf{a}_M^{\mathsf{T}}(\upsilon_{l_g}^{G_t}, \rho_{l_g}^{G_t}) \\ & \mathtt{diag}\left(\sqrt{\frac{M}{L_f}} \sum_{l_f=1}^{L_f} \beta_{l_f} \mathbf{a}_M(\upsilon_{l_f}^{r}, \rho_{l_f}^{r})\right) \\ =& \sqrt{\frac{MN}{L_G L_f}} \sum_{l_g=1}^{L_G} \sum_{l_f=1}^{L_f} \alpha_{l_g}\beta_{l_f} \mathbf{b}_N(\theta_{l_g}^{G_r}) \\ & \times \mathbf{a}_M^{\mathsf{T}}(\upsilon_{l_g}^{G_t} + \upsilon_{l_f}^{r}, \rho_{l_g}^{G_t} + \rho_{l_f}^{r}).\end{aligned} \tag{9}$$

Applying (4) and (9), $\tilde{\mathbf{H}}$ in (8) can then be rewritten as

$$\begin{aligned}\tilde{\mathbf{H}} =&\bar{\mathbf{H}}^{\mathsf{H}} = (\mathbf{U}_N^{\mathsf{H}}\mathbf{H}\mathbf{U}_M^*)^{\mathsf{H}} = \sqrt{\frac{MN}{L_G L_f}}\sum_{l_g=1}^{L_G}\sum_{l_f=1}^{L_f}\alpha_{l_g}^*\beta_{l_f}^* \\ &\times \tilde{\mathbf{a}}_M^*(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^r,\rho_{l_g}^{G_t}+\rho_{l_f}^r)\tilde{\mathbf{b}}_N^{\mathsf{H}}(\theta_{l_g}^{G_r}),\end{aligned} \tag{10}$$

where $\tilde{\mathbf{b}}_N(\theta_{l_g}^{G_r}) \triangleq \mathbf{U}_N^{\mathsf{H}}\mathbf{b}_N(\theta_{l_g}^{G_r})$, $\tilde{\mathbf{a}}_M(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^r,\rho_{l_g}^{G_t}+\rho_{l_f}^r) \triangleq \mathbf{U}_M^{\mathsf{H}}\mathbf{a}_M(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^r,\rho_{l_g}^{G_t}+\rho_{l_f}^r)$. To further clarify the structure of the $\tilde{\mathbf{H}}$, Proposition 1 is introduced. Here, $(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^r)$ and $(\rho_{l_g}^{G_t}+\rho_{l_f}^r)$ are respectively referred to as the azimuth and elevation cascaded AoA-AoD of the (l_g,l_f)-th subpath (the cascaded subpath consists of the l_g-th path of $\mathbf{G}$ and the l_f-th path of $\mathbf{f}$).

Proposition 1: Both $\tilde{\mathbf{b}}_N(\theta_{l_g}^{G_r}) \in \mathbb{C}^{N_G}$ and $\tilde{\mathbf{a}}_M(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^r,\rho_{l_g}^{G_t}+\rho_{l_f}^r) \in \mathbb{C}^{M_G}$ have only one non-zero element, which are respectively determined by the AoA at the BS and the cascaded AoA-AoD (both azimuth and elevation) at the RIS.

Proof: Please refer to Appendix A.

In total, there are $L_G L_f$ subpaths of the cascaded UE-RIS-BS channel, contributing to $L_G L_f$ non-zero elements in $\tilde{\mathbf{H}}$. Based on Proposition 1, it can be found that, each non-zero element in $\tilde{\mathbf{H}}$, denoted as h^{l_g,l_f}, corresponds to one specific subpath consisting of the l_g-th path of $\mathbf{G}$ and the l_f-th path of $\mathbf{f}$. Column and row indices of this non-zero elements in $\tilde{\mathbf{H}}$ depend on $\theta_{l_g}^{G_r}$ and $\{\upsilon_{l_g}^{G_t},\upsilon_{l_f}^r,\rho_{l_g}^{G_t},\rho_{l_f}^r\}$, respectively.

When $\{\theta_{l_g}^{G_r},\upsilon_{l_g}^{G_t},\rho_{l_g}^{G_t},\upsilon_{l_f}^r,\rho_{l_f}^r\}$ respectively equal to the $\{n^{l_g},m_x^{l_g},m_y^{l_g},m_x^{l_f},m_y^{l_f}\}$-th pre-determined grid point, Proposition 2 unveils the relationship between the angles of the channels and the indices of the non-zero elements in $\tilde{\mathbf{H}}$.

Proposition 2: The row and column indices of the non-zero element h^{l_g,l_f} in $\tilde{\mathbf{H}}$ are $(p^{l_g,l_f}-1)M_{G_x}+q^{l_g,l_f}$ and n^{l_g} respectively, where p^{l_g,l_f} and q^{l_g,l_f} are defined in (11).

Proof: Please refer to Appendix B.

Overall, there are no more than L_G non-zero columns in $\tilde{\mathbf{H}}$ due to the L_G BS AoAs, and each non-zero column contains no more than L_f non-zero elements due to the L_f cascaded AoA-AoDs at the RIS.

A. Column-structure

It can be seen from (10) that the indices of the non-zero columns in $\tilde{\mathbf{H}}$ only depend on the spatial AoAs $\theta_{l_g}^{G_r}$ at the BS, and are independent of the angles at the RIS. Considering that the RIS-BS channel varies more slowly than the UE-RIS channel [21], $\mathbf{G}_k$ may last for $\tau(\tau \geq 1)$ blocks, i.e., $\mathbf{G}_k = \cdots = \mathbf{G}_{k+\tau-1} = \mathbf{G}$. Consequently, $\tilde{\mathbf{H}}_k,\cdots,\tilde{\mathbf{H}}_{k+\tau-1}$ share the indices of non-zero columns

$$\mathbf{\Omega}_k^c = \mathbf{\Omega}_{k+1}^c = \cdots, \mathbf{\Omega}_{k+\tau-1}^c = \mathbf{\Omega}^c, \tag{12}$$

where $\mathbf{\Omega}^c$ represents the supports of non-zero columns in $\tilde{\mathbf{H}}$.

B. Row-structure A: common spacings within a single block

For the (l_g,l_f) subpath, the azimuth cascaded AoA-AoD at the RIS can be rewritten as

$$\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^r = \upsilon_{l_g}^{G_t}+\upsilon_{l'_f}^r-(\upsilon_{l'_f}^r-\upsilon_{l_f}^r) = \upsilon_{l_g}^{G_t}+\upsilon_{l'_f}^r-\Delta\upsilon_{l'_f,l_f}^r, \tag{13}$$

where $\Delta\upsilon_{l'_f,l_f}^r \triangleq \upsilon_{l'_f}^r-\upsilon_{l_f}^r$. $l'_f,l_f \in [1,L_f]$, $l'_f \neq l_f$. Similarly, the elevation cascaded AoA-AoD can be decomposed as

$$\rho_{l_g}^{G_t}+\rho_{l_f}^t = \rho^{G_t}+\rho_{l'_f}^t-\Delta\rho_{l'_f,l_f}^r, \tag{14}$$

where $\Delta\rho_{l'_f,l_f}^r \triangleq \rho_{l'_f}^r-\rho_{l_f}^r$. Obviously, the angular differences $\Delta\upsilon_{l'_f,l_f}^r$ and $\Delta\rho_{l'_f l_f}^r$ solely depend on the l'_f-th and l_f-th paths of $\mathbf{f}$, and are irrelevant to $\mathbf{G}$. The two non-zero elements $(h^{l_g,l'_f},h^{l_g,l_f})$ resulting from the (l_g,l'_f) and (l_g,l_f) subpaths lie on the same non-zero column in $\tilde{\mathbf{H}}$ due to the shared $\theta_{l_g}^{G_r}$. In addition, the corresponding row indices $(p^{l_g,l'_f}-1)M_{G_y}+q^{l_g,l'_f}$ and $(p^{l_g,l_f}-1)M_{G_y}+q^{l_g,l_f}$ satisfy[2]

$$\begin{aligned}\mathtt{mod}(p^{l_g,l'_f}-p^{l_g,l_f},M_{G_x}) &= \mathtt{mod}(m_x^{l'_f}-m_x^{l_f},M_{G_x}) \triangleq \Delta p_{l'_f,l_f}^r \\ \mathtt{mod}(q^{l_g,l'_f}-q^{l_g,l_f},M_{G_y}) &= \mathtt{mod}(m_y^{l'_f}-m_y^{l_f},M_{G_y}) \triangleq \Delta q_{l'_f,l_f}^r,\end{aligned} \tag{15}$$

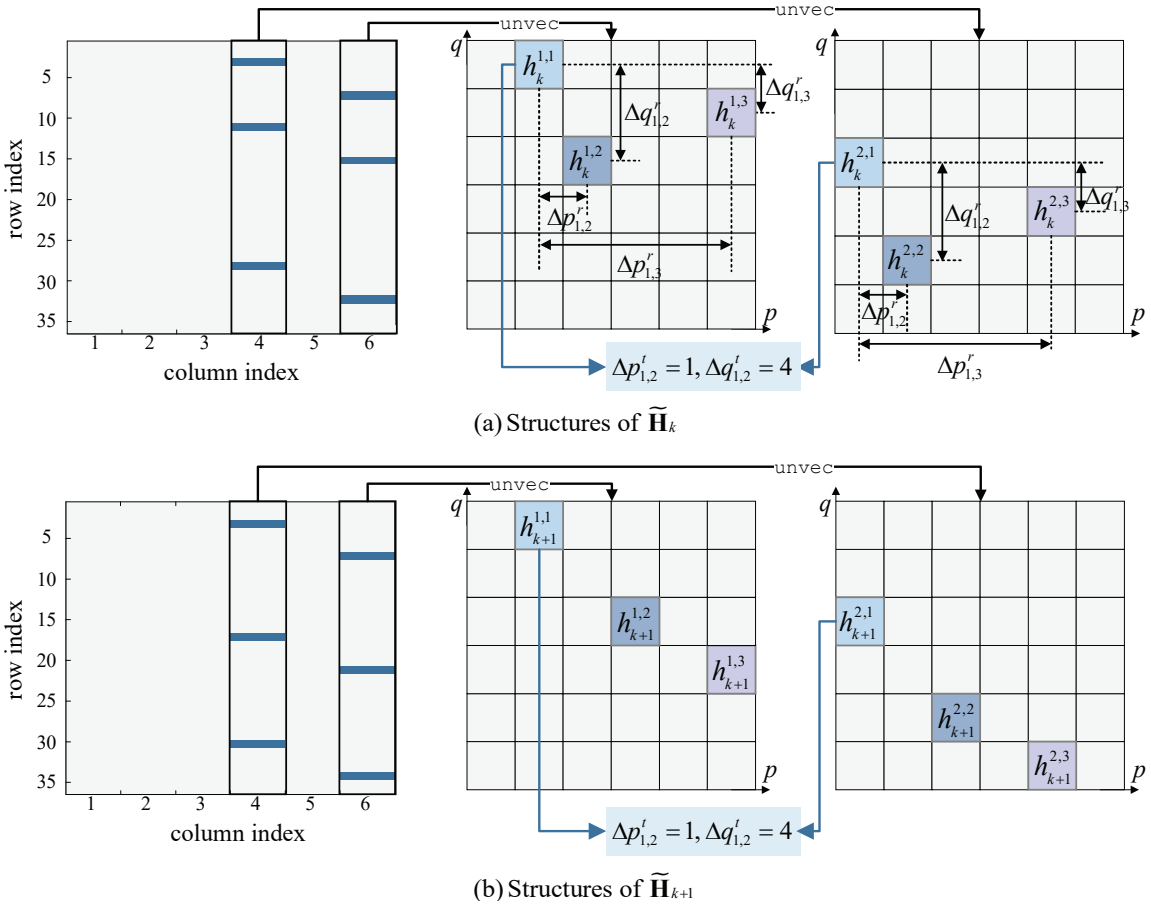
where $\Delta p_{l'_f,l_f}^r$ and $\Delta q_{l'_f,l_f}^r$ are defined as the azimuth and elevation row spacings between h^{l_g,l'_f} and h^{l_g,l_f}.

From (13)-(15), we observe that within a single block, there exist certain commonalities among the non-zero elements located in different non-zero columns. Specifically, $\{\Delta p_{l'_f,l_f}^r,\Delta q_{l'_f,l_f}^r\}$ are the same for $\{h^{l_g,l_f},h^{l_g,l'_f}\}$, $\forall l_g \in [1,L_G]$. In other words, the spacings between the non-zero elements resulting from the (l_g,l'_f) and (l_g,l_f) subpaths are the same for $\forall l_g$. This commonality can be exploited to simplify CE. For instance, once $\{h^{1,1},\cdots,h^{1,L_f}\}$ are obtained. By choosing a reference element h^{1,l'_f}, the spacings $\Delta p_{l'_f,l_f}^r,\Delta q_{l'_f,l_f}^r$ can be computed to serve as prior information to simplify the estimation of the remaining non-zero elements.

[2]Eq. (15) is obtained based on (11), detailed derivation process is omitted here for brevity.

$$p^{l_g,l_f} = \begin{cases} m_x^{l_g}+m_x^{l_f}+\frac{M_{G_x}-2}{2} & \text{if } (m_x^{l_g}+m_x^{l_f}) \in [2,\frac{M_{G_x}}{2}+2) \\ m_x^{l_g}+m_x^{l_f}-\frac{M_{G_x}+2}{2} & \text{if } (m_x^{l_g}+m_x^{l_f}) \in [\frac{M_{G_x}}{2}+2,\frac{3M_{G_x}}{2}+1] \\ m_x^{l_g}+m_x^{l_f}-\frac{3M_{G_x}+2}{2} & \text{if } (m_x^{l_g}+m_x^{l_f}) \in (\frac{3M_{G_x}}{2}+1,2M_{G_x}], \end{cases} \tag{11a}$$

$$q^{l_g,l_f} = \begin{cases} m_y^{l_g}+m_y^{l_f}+\frac{M_{G_y}-2}{2} & \text{if } (m_y^{l_g}+m_y^{l_f}) \in [2,\frac{M_{G_y}}{2}+2) \\ m_y^{l_g}+m_y^{l_f}-\frac{M_{G_y}+2}{2} & \text{if } (m_y^{l_g}+m_y^{l_f}) \in [\frac{M_{G_y}}{2}+2,\frac{3M_{G_y}}{2}+1] \\ m_y^{l_g}+m_y^{l_f}-\frac{3M_{G_y}+2}{2} & \text{if } (m_y^{l_g}+m_y^{l_f}) \in (\frac{3M_{G_y}}{2}+1,2M_{G_y}]. \end{cases} \tag{11b}$$



(a) Structures of $\widetilde{\mathbf{H}}_k$

(b) Structures of $\widetilde{\mathbf{H}}_{k+1}$

Fig. 1. Structure of two cascaded sparse channels where $\mathbf{G}_k = \mathbf{G}_{k+1}$, $M = 36(6\times 6), N = 6, L_G = 2, L_f = 3$. Colored rectangles represent the non-zero elements, while gray rectangles represent the zero elements.

C. Row-structure B: common spacings among blocks

For the subpath consists of the l_g-th path of $\mathbf{G}$ and the l_f-th path of $\mathbf{f}$, the cascaded AoA-AoD at the RIS satisfy

$$\begin{aligned}\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r} &= \upsilon_{l'_g}^{G_t}+\upsilon_{l_f}^{r}-(\upsilon_{l'_g}^{G_t}-\upsilon_{l_g}^{G_t}) = \upsilon_{l'_g}^{G_t}+\upsilon_{l_f}^{r}-\Delta\upsilon_{l'_g,l_g}^{G_t},\\ \rho_{l_g}^{G_t}+\rho_{l_f}^{r} &= \rho_{l'_g}^{G_t}+\rho_{l_f}^{r}-(\rho_{l'_g}^{G_t}-\rho_{l_g}^{G_t}) = \rho_{l'_g}^{G_t}+\rho_{l_f}^{r}-\Delta\rho_{l'_g,l_g}^{G_t},\end{aligned} \tag{16}$$

where $l'_g, l_g \in [1, L_G], l'_g \neq l_g$. $\Delta\upsilon_{l'_g,l_g}^{G_t} \triangleq \upsilon_{l'_g}^{G_t} - \upsilon_{l_g}^{G_t}$, $\Delta\rho_{l'_g,l_g}^{G_t} \triangleq \rho_{l'_g}^{G_t} - \rho_{l_g}^{G_t}$ are referred to as the azimuth and elevation angular differences between the l'_g-th and the l_g-th of $\mathbf{G}$. Apparently, when $\mathbf{G}$ is constant, $\Delta\upsilon_{l'_g,l_g}^{G_t}, \Delta\rho_{l'_g,l_g}^{G_t}$ remain unchanged. Besides, the indices of the non-zero elements resulting from the (l'_g, l_f) and (l_g, l_f) subpaths are related to each other as follows

$$\begin{aligned}\mathtt{mod}(p^{l'_g,l_f}-p^{l'_g,l_f}, M_{G_x}) &= \mathtt{mod}(m_x^{l'_g}-m_x^{l_g}, M_{G_x}) \triangleq \Delta p_{l'_g,l_g}^{t},\\ \mathtt{mod}(q^{l'_g,l_f}-q^{l_g,l_f}, M_{G_y}) &= \mathtt{mod}(m_y^{l'_g}-m_y^{l_g}, M_{G_y}) \triangleq \Delta q_{l'_g,l_g}^{t},\end{aligned} \tag{17}$$

where $\Delta p_{l'_g,l_g}^{t}, \Delta q_{l'_g,l_g}^{t}$ are the azimuth and elevation row spacings between $h^{l'_g,l_g}$ and h^{l_g,l_f} in $\tilde{\mathbf{H}}$.

Similarly, $\{\Delta\upsilon_{l'_g,l_g}^{G_t}, \Delta\rho_{l'_g,l_g}^{G_t}\}$ and $\{\Delta p_{l'_g,l_g}^{t}, \Delta q_{l'_g,l_g}^{t}\}$ are unrelated to $\mathbf{f}$. Consequently, for the estimation of $\tilde{\mathbf{H}}_k$ across different blocks with the same $\mathbf{G}_k$, these common angular differences (or spacings) can be utilized to reduce the complexity of the CE process. For example, with the CSI acquired in previous blocks, $\{\Delta\upsilon_{l'_g,l_g}^{G_t}, \Delta\rho_{l'_g,l_g}^{G_t}\}$ or $\{\Delta p_{l'_g,l_g}^{t}, \Delta q_{l'_g,l_g}^{t}\}$ can be calculated and leveraged to improve the estimation of CSI in the current block when $\mathbf{G}$ remains unchanged.

D. Illustration of a showcase

A showcase has been presented in Fig. 1, where $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{H}}$ share the same $\mathbf{G}$. Therefore, for $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{H}}_{k+1}$, the indices of the non-zero columns are the same, i.e., $\mathbf{\Omega}_k^c = \mathbf{\Omega}_{k+1}^c = \{4, 6\}$.

For the subpaths that share the common UE-RIS paths but different RIS-BS paths, subfigures in Fig. 1(a) reveal the derived row-structure A. Specifically, choosing $h_k^{1,1}$ and $h_k^{2,1}$ as reference elements in each non-zero column (i.e., $l'_f = 1$ in (15)), $\{\Delta p_{1,2}^r = 4, \Delta q_{1,2}^r = 5\}$ holds for both $\{h^{1,1}, h^{1,2}\}$ and $\{h^{2,1}, h^{2,2}\}$. Similarly, $\{\Delta p_{1,3}^r = 2, \Delta q_{1,3}^r = 5\}$ applies for both $\{h_k^{1,1}, h_k^{1,3}\}$ and $\{h_k^{2,1}, h_k^{2,3}\}$.

As discussed in Section III-C, while $\mathbf{f}_k, \mathbf{f}_{k+1}$ are different for $\mathbf{H}_k$ and $\mathbf{H}_{k+1}$, the common RIS-BS channel $\mathbf{G}$ enables $\{h_k^{1,1}, h_k^{2,1}\}$ and $\{h_{k+1}^{1,1}, h_{k+1}^{2,1}\}$ to share the same value of spacings, i.e., $\{\Delta p_{1,2}^t = 1, \Delta q_{1,2}^t = 4\}$ applied to both $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{H}}_{k+1}$(depicted in Fig. 1(b)).

IV. Proposed Adaptive CE Scheme

RIS-associated channels are typically considered to exhibit the two-timescale property [21]. This property facilitates im-

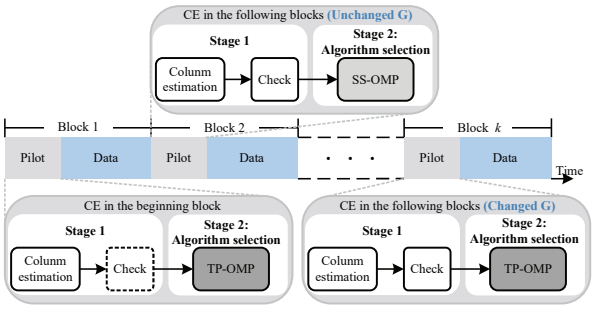


Fig. 2. The proposed adaptive CE scheme.

proved and simplified estimation of time-varying channels over consecutive blocks by leveraging CSI of the quasi-static RIS-BS channel. Existing approaches for RIS-assisted time-varying CE usually assume a known and fixed variation mode of the RIS-BS channel, i.e., the duration of unchanged $\mathbf{G}$. Thus, frequencies of estimating $\mathbf{G}$ and $\mathbf{f}$ vary accordingly [21], [24], [25]. However, in practical scenarios, the variation of $\mathbf{G}$ may deviate from such assumption, leading to a considerable degradation in CE performance. To address this limitation, an adaptive CE scheme is proposed in this paper. The proposed adaptive scheme can dynamically select appropriate CE algorithms for each block without relying on the aforementioned ideal assumption.

As depicted in Fig. 2, the proposed adaptive scheme consists of two stages within each block. In the initial block, the first stage omits the check process due to the absence of prior CSI, and the second stage exploits TP-OMP. In the subsequent k-th ($k > 1$) block, estimation of the indices of the non-zero columns in $\tilde{\mathbf{H}}_k$ is initially performed, followed by a check process to detect the changing status of $\mathbf{G}_k$. During the second stage, based on the result of the check process, the adaptive scheme selects a more appropriate algorithm to acquire full CSI. Specifically, SS-OMP is chosen if $\mathbf{G}_k$ remains unchanged; otherwise, TP-OMP is employed. The complete adaptive CE scheme is summarized in Algorithm 1. Further detailed explanations of the two stages are provided in Section IV-A and Section IV-B respectively.

A. Stage 1: Estimation of the Non-zero Columns and Check

1) Column Support Estimation (Step 2 in Alg. 1): Based on (8), each column of $\tilde{\mathbf{Y}}_k$ can be represented as

$$\tilde{\mathbf{Y}}_k(:,n) = \tilde{\mathbf{W}}_k\tilde{\mathbf{H}}_k(:,n) + \tilde{\mathbf{N}}_k(:,n),\ n\in[1,N_G]. \tag{18}$$

It can be found that the columns of the received signal $\tilde{\mathbf{Y}}_k$ with higher energy correspond to the non-zero columns in $\tilde{\mathbf{H}}_k$. Thus, the indices of the non-zero columns in $\tilde{\mathbf{H}}_k$ can be obtained by simply calculating the energy of the columns in $\tilde{\mathbf{Y}}_k$.

$$\mathbf{\Omega}_k^c = \Gamma(\|\tilde{\mathbf{Y}}_k(:,n)\|_F^2, L_G),\ n=1,\cdots,N_G, \tag{19}$$

where $\Gamma(\mathbf{A},l)$ denotes the operator to choose indices of the l largest elements in $\mathbf{A}$. After obtaining $\mathbf{\Omega}_k^c$, CE can be simplified as

$$\tilde{\mathbf{Y}}_k(:,\mathbf{\Omega}_k^c) = \tilde{\mathbf{W}}\tilde{\mathbf{H}}_k(:,\mathbf{\Omega}_k^c) + \tilde{\mathbf{N}}_k(:,\mathbf{\Omega}_k^c), \tag{20}$$

Algorithm 1 Adaptive Scheme for Multi-block CE

Input: $\tilde{\mathbf{Y}}_k$, $\tilde{\mathbf{W}}_k$, L_G, L_f, M, N, M_{G_x}, M_{G_y}.
Output: $\hat{\mathbf{H}}_k$.

% Initialization
$\hat{\mathbf{H}}_k = \mathbf{0}_{M\times N}$
% Stage 1: Column estimation and the check process
obtain $\mathbf{\Omega}_k^c$ via (19).
if $k=1$ **then**
  $\mathbf{\Omega}^c = \mathbf{\Omega}_1^c$
  Obtain $\hat{\tilde{\mathbf{H}}}_1(:,\mathbf{\Omega}^c)$ by **TP-OMP**.
else
  Obtain χ via (22).
% Stage 2: Algorithm selection
  if $\chi=0$ **then**
    Update $\mathbf{\Omega}^c = \mathbf{\Omega}_k^c$.
    Obtain $\hat{\tilde{\mathbf{H}}}_k(:,\mathbf{\Omega}^c)$ by **TP-OMP**.
  else
    Obtain $\hat{\tilde{\mathbf{H}}}_k(:,\mathbf{\Omega}^c)$ by **SS-OMP**.
  end if
end if
Construct $\hat{\mathbf{H}}_k$ via (21).

where $\tilde{\mathbf{Y}}_k(:,\mathbf{\Omega}_k^c)\in\mathbb{C}^{T\times L_G}$, $\tilde{\mathbf{H}}_k(:,\mathbf{\Omega}_k^c)\in\mathbb{C}^{M_G\times L_G}$. Compared to (8), the dimension of the sparse signal to be estimated in (20) reduces by N/L_G. And final estimation of the cascaded channel $\mathbf{H}_k$ can be obtained via

$$\hat{\mathbf{H}}_k = \mathbf{U}_N(:,\mathbf{\Omega}_k^c)\hat{\tilde{\mathbf{H}}}_k^{\mathsf{H}}(:,\mathbf{\Omega}_k^c)\mathbf{U}_M^{\mathsf{T}}. \tag{21}$$

2) The Check Process (Step 7 in Alg. 1): Exploiting the two-timescale of the RIS-assisted channels, existing works assume that $\mathbf{G}_k$ remain the same during a certain number of blocks, i.e., τ in (12) is known and fixed. Consequently, the estimation of $\mathbf{G}$ is only required in the $(n\tau+1)$-th $(n\geq 0)$ block. However, the exact value of τ is unknown and may vary in practical scenarios. Therefore, instead of assuming a fixed and known τ, our proposed scheme incorporates a check process to detect whether $\mathbf{G}_k$ remains the same as in the previous blocks. Specifically, by comparing the AoAs at the BS in the k-th block with those in the preceding blocks, the proposed check process can effectively identify if $\mathbf{G}_k$ has changed. Notably, the estimation of the non-zero column indices (which correspond to the AoAs at the BS) can be obtained with low computational complexity, as presented earlier. The check process can then effectively detect the changing status of $\mathbf{G}_k$ after column estimation without introducing extra computational complexity or pilots. Details of the check process are explained as follows.

In the initial block, the check process is omitted. $\mathbf{\Omega}^c$ is then initialized as $\mathbf{\Omega}_1^c$. For the k-th ($k>1$) block, the check process compares $\mathbf{\Omega}_k^c$ with the stored $\mathbf{\Omega}^c$. The result of the check process (denoted as χ) can be computed by

$$\chi = \begin{cases} 1, & \text{if } \mathtt{diff}(\mathbf{\Omega}_k^c,\mathbf{\Omega}^c) = \varnothing \\ 0, & \text{if } \mathtt{diff}(\mathbf{\Omega}_k^c,\mathbf{\Omega}^c) \neq \varnothing, \end{cases} \tag{22}$$

Algorithm 2 TP-OMP for cascaded CE in the k-th block

Input: $\tilde{\mathbf{Y}}_k$, $\tilde{\mathbf{W}}_k$, $\mathbf{\Omega}^c$, L_G, L_f, M_{G_x}, M_{G_y}.
Output: $\hat{\tilde{\mathbf{H}}}_k(:,\mathbf{\Omega}^c)$, $\Delta\mathbf{p}^t$, $\Delta\mathbf{q}^t$.

% Initialization
1: $\hat{\tilde{\mathbf{H}}}_k(:,\mathbf{\Omega}^c)=\mathbf{0}_{M_G\times L_G}$, $\Delta\mathbf{p}^t=\varnothing$, $\Delta\mathbf{q}^t=\varnothing$
2: **for** $l=1$ to L_G **do**
3: $\tilde{\mathbf{y}}_l=\tilde{\mathbf{Y}}_k(:,\mathbf{\Omega}^c(l)),\tilde{\mathbf{r}}=\tilde{\mathbf{y}}_l$
4: **if** $l=1$ **then**
% Phase 1: estimate a reference non-zero column and obtain the spacings
5: **for** $j=1$ to L_f **do**
6: $\mathbf{C}=\texttt{unvec}_{\{M_{G_x},M_{G_y}\}}(|\tilde{\mathbf{W}}_k^{\mathsf{H}}\tilde{\mathbf{r}}|)$
7: $[\mathbf{q}_1(j),\mathbf{p}_1(j)]=\texttt{max}(\mathbf{C})$
8: $\mathbf{s}_1(j)=(\mathbf{p}_1(j)-1)M_x+\mathbf{q}_1(j)$
9: $\tilde{\mathbf{h}}_1=\tilde{\mathbf{W}}_k(:,\mathbf{s}_1)^\dagger\tilde{\mathbf{y}}_i$,
10: $\tilde{\mathbf{r}}=\tilde{\mathbf{y}}_1-\tilde{\mathbf{W}}_k(:,\mathbf{s}_1)\tilde{\mathbf{h}}_1$
11: **end for**
// Calculate the spacings in (15).
12: $\Delta\mathbf{p}^r=\mathbf{p}_1(1)-\mathbf{p}_1(2:L_f)$
$\Delta\mathbf{q}^r=\mathbf{q}_1(1)-\mathbf{q}_1(2:L_f)$
13: **else**
% Phase 2: Estimate the remaining non-zero columns
14: $\mathbf{C}=\texttt{unvec}_{\{M_{G_x},M_{G_y}\}}(|\tilde{\mathbf{W}}_k^{\mathsf{H}}\tilde{\mathbf{r}}|)$
15: $\tilde{\mathbf{C}}=\mathbf{C}+\sum_{j=1}^{L_f-1}\texttt{circ}[\mathbf{C},\Delta\mathbf{p}^r(j),\mathbf{q}^r(j)]$
16: $[\mathbf{p}_l(1),\mathbf{q}_l(1)]=\texttt{max}(\tilde{\mathbf{C}})$
17: **for** $i=1$ to L_f-1 **do**
18: $\mathbf{p}_l(i+1)=\texttt{circ}[\mathbf{p}_l(1),\Delta\mathbf{p}^r(i)]_{M_{G_x}}$
$\mathbf{q}_l(i+1)=\texttt{circ}[\mathbf{q}_l(1),\Delta\mathbf{q}^r(i)]_{M_{G_y}}$
19: **end for**
20: $\mathbf{s}_l=(\mathbf{p}_l-1)M_{G_x}+\mathbf{q}_l$
21: $\tilde{\mathbf{h}}_l=\tilde{\mathbf{W}}_k(:,\mathbf{s}_l)^\dagger\tilde{\mathbf{y}}_l$
22: **end if**
23: $\tilde{\mathbf{H}}_k(\mathbf{s}_l,\mathbf{\Omega}^c(l))=\tilde{\mathbf{h}}_l$
// Calculate the spacings in (17) for SS-OMP.
24: $\Delta\mathbf{p}^t=\Delta\mathbf{p}^t\cup\{\mathbf{p}_1(1)-\mathbf{p}_l(1)\}$
$\Delta\mathbf{q}^t=\Delta\mathbf{q}^t\cup\{\mathbf{q}_1(1)-\mathbf{q}_l(1)\}$
25: **end for**

where $\texttt{diff}(\mathbf{a},\mathbf{b})$ returns the vector that contains different entries of $\mathbf{a}$ and $\mathbf{b}$. In particular, $\chi=1$ indicates that $\mathbf{G}_k$ remains unchanged, while $\chi=0$ means that $\mathbf{G}_k$ has changed.

B. Stage 2: Selection From the Candidate Algorithms

With the obtained χ, stage 2 of the proposed adaptive CE scheme chooses a specific CE algorithm from the two proposed candidate algorithms: **T**wo-**P**hase OMP (TP-OMP) and **S**tructured-**S**hift OMP (SS-OMP). In particular, if $\mathbf{G}_k=\mathbf{G}_{k-1}$ ($\chi=1$), SS-OMP can be employed to acquire the CSI with lower computational complexity. Otherwise, TP-OMP is chosen to guarantee the CE performance.

1) The Proposed TP-OMP (Steps 5 and 10 in Alg. 1): TP-OMP is proposed by exploiting the row-structure A analyzed in Section III-B. The main procedure of TP-OMP is summarized in **Algorithm 2**. After obtaining $\mathbf{\Omega}^c$, the estimation of each non-zero column in $\tilde{\mathbf{H}}_k$ can be transformed into a sparse vector recovery model

$$\tilde{\mathbf{y}}_{k,l}=\tilde{\mathbf{W}}_k\tilde{\mathbf{h}}_{k,l}+\tilde{\mathbf{n}}_{k,l},\quad l=1,\cdots,L_G, \tag{23}$$

where $\tilde{\mathbf{y}}_{k,l}=\tilde{\mathbf{Y}}_k(:,\mathbf{\Omega}^c(l))\in\mathbb{C}^T$, $\tilde{\mathbf{h}}_{k,l}=\tilde{\mathbf{H}}_k(:,\mathbf{\Omega}^c(l))\in\mathbb{C}^{M_G}$, $\tilde{\mathbf{n}}_{k,l}=\tilde{\mathbf{N}}_k(:,\mathbf{\Omega}^c(l))$.

As revealed in the derivation of row-structure A, the L_GL_f non-zero elements in $\tilde{\mathbf{H}}$ have some common features. Specifically, for each non-zero column, there are L_f non-zero elements, and the intervals between these L_f elements are determined by the angular differences among the L_f UE-RIS channel paths. Within a single block, the intervals between the L_f elements remain consistent across all L_G non-zero columns. Thus, the CE process can be simplified. In the proposed TP-OMP, CE is divided into two phases.

Phase 1 (including steps 4-12 of Alg. 2): In phase 1 of TP-OMP, one of the non-zero columns, referred to as the reference non-zero column $\tilde{\mathbf{h}}_1$, is firstly recovered via conventional OMP[3]. Using the obtained $\tilde{\mathbf{h}}_1$, the azimuth and elevation row spacings between the L_f non-zero elements in $\tilde{\mathbf{h}}_1$ can be calculated, which provide the prior information for the estimation of the remaining (L_G-1) non-zero columns. Specifically, in step 12, by taking the firstly estimated non-zero element $h^{1,1}$ in $\tilde{\mathbf{h}}_1$ as the reference non-zero element ($l'_f=1$), the specific spacing vectors $\Delta\mathbf{p}^r$ and $\Delta\mathbf{q}^r$ can be easily obtained. It should be noted that both $\Delta\mathbf{p}^r$ and $\Delta\mathbf{q}^r$ contain (L_f-1) entries since there are L_f non-zero elements in a single non-zero column.

Phase 2 (including steps 13-24 of Alg. 2): In phase 2 of TP-OMP, with the obtained $\Delta\mathbf{p}^r$ and $\Delta\mathbf{q}^r$, recovery of remaining non-zero columns $\tilde{\mathbf{h}}_l(l\in[2,L_G])$ can be simplified. For each of the rest L_G-1 non-zero columns, indices of the reference non-zero element $h^{l,1}$ is estimated in step 14-16. Different

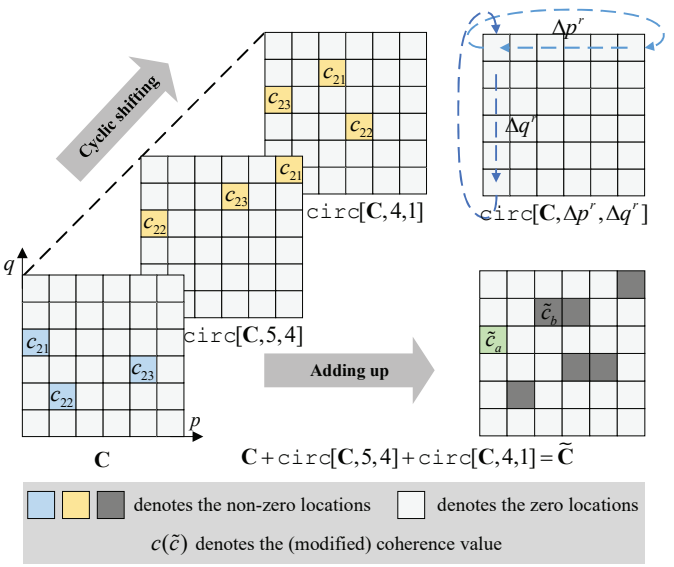


Fig. 3. The example of the correlation modification for the case of finding the position of $h_k^{2,1}$ in Fig. 1, where $\Delta\mathbf{p}^r=\{\Delta p^r_{1,2},\Delta p^r_{1,3}\}=\{5,2\}$, $\Delta\mathbf{q}^r=\{\Delta q^r_{1,2},\Delta q^r_{1,3}=\{4,5\}$.

[3]In the following descriptions, the subscript k is omitted. A more detailed description of the OMP algorithm can be found in [29].

from OMP, in step 15, the initially calculated correlation ($\mathbf{C}$) between the sensing matrix and residual is modified into $\tilde{\mathbf{C}}$ to find $h^{l,1}$. An example is presented in Fig. 3 to illustrate this modification process.

As shown in Fig. 3, the elements in $\mathbf{C}$ are first cyclically shifted according to $\Delta\mathbf{p}^r$ and $\Delta\mathbf{q}^r$. Subsequently, by adding up all the cyclically shifted elements, step 16 is able to find the specific index more accurately (Please refer to Remark 1 for a detailed explanation). In the following steps 17-19, as (15) implicates, the acquisition of $\Delta\mathbf{p}^r, \Delta\mathbf{q}^r$ allows the TP-OMP to directly compute the indices of the remaining elements in $\tilde{\mathbf{h}}_l$ one by one without repeating correlation calculation. Step 21 gets the estimation of $\tilde{\mathbf{h}}_l$ via LS. Finally, the spacings in (17) can be directly computed in step 24, which are included in the two sets $\Delta\mathbf{p}^t$ and $\Delta\mathbf{q}^t$, with each of them having $L_G{-}1$ entries.

Remark 1: The modification of coherence is able to mitigate the impact of noise. Taking the first iteration as an example, where the residual $\tilde{\mathbf{r}}$ is initialized as $\tilde{\mathbf{y}}_l$. For the conventional OMP, each entry of $\mathbf{c} = |\tilde{\mathbf{W}}^{\mathsf{H}}\tilde{\mathbf{r}}|$ can be represented as

$$\mathbf{c}(m) = \left| \sum_{t=1}^{T} \tilde{w}_{t,m}^{*} \left(\sum_{x=1}^{M_G} \tilde{w}_{t,x} \tilde{\mathbf{h}}_l(x) + \tilde{\mathbf{n}}_l(x) \right) \right|, \tag{24}$$

where $\tilde{w}_{t,m}$ denotes the entry in the t-th row and m-th column of $\tilde{\mathbf{W}}$. According to the CS theory, indices of the entries in $\mathbf{C}$ with larger correlation values correspond to the indices of the non-zero elements in $\tilde{\mathbf{h}}_l$. However, the existence of noise makes this approach less effective, especially when SNR is low. In contrast to the traditional OMP, after cyclic shifting and adding up, each element in $\tilde{\mathbf{C}}$ is a combination of multiple entries in $\mathbf{C}$, denoted as

$$\tilde{c}(p,q) \overset{(a)}{=} \sum_{\mathbb{I}(\{p,q\})} \mathbf{C}(m), \tag{25}$$

where $m = (p-1)M_{G_x} + q$. In (a), $\mathbb{I}(\{p,q\})$ represents the indices of the corresponsing positions after $\{p,q\}$ being cyclically shifted according to the spacings $\{\Delta\mathbf{p}^r, \Delta\mathbf{q}^r\}$.

An exemplary modification process is shown in Fig. 3, and the spacings are computed via the showcase in Fig. 1. For the recovery of $\tilde{\mathbf{h}}_2$, we have to first find the indices $h^{2,1}$. With known $\Delta\mathbf{p}^r = \{\Delta p_{1,2}^r, \Delta p_{1,3}^r\}$ and $\Delta\mathbf{q}^r = \{\Delta q_{1,2}^r, \Delta q_{1,3}^r\}$ (obtained in step 12 of TP-OMP), each of the entries in $\mathbf{C}$ successively conducts the cyclic shifting accordingly, resulting in $\texttt{circ}[\mathbf{C}, 5, 4]$ and $\texttt{circ}[\mathbf{C}, 4, 1]$. The top right corner of Fig. 3 shows the direction of this cyclic shift. Then, after an addition operation, the modified coherence matrix $\tilde{\mathbf{C}}$ is obtained. Each of the entries in $\tilde{\mathbf{C}}$ contains L_f initial coherence (3 in this example). It can be found that the value of $\tilde{c}_a$ in Fig. 3 is the sum of all L_f non-zero locations, i.e., $\tilde{c}_a = c_{21} + c_{22} + c_{23}$[4]. In contrast, other entries contain the coherence of either one or no non-zero location, i.e., $\tilde{c}_b = c_{21}$. Thus, the impact of noise on $\tilde{\mathbf{C}}$ is relatively lower since the noise is randomly distributed while the non-zero elements are distributed with certain patterns. Therefore, the modification of the coherence can mitigate the noise's influence.

[4]The coherence values of the zero locations are omitted here for simplicity.

Algorithm 3 SS-OMP for cascaded CE in the k-th block.

Input: $\tilde{\mathbf{Y}}_k$, $\tilde{\mathbf{W}}_k$, $\mathbf{\Omega}^c$, $\Delta\mathbf{p}^t$, $\Delta\mathbf{q}^t$, L_G, L_f, M_{G_x}, M_{G_y}.
Output: $\hat{\tilde{\mathbf{H}}}_k(:, \mathbf{\Omega}^c)$.

% initialization
1: $\hat{\tilde{\mathbf{H}}}_k(:, \mathbf{\Omega}^c) = \mathbf{0}_{M_G \times L_G}$
% Estimate the reference non-zero column via OMP
2: $\tilde{\mathbf{y}}_1 = \tilde{\mathbf{Y}}_k(:, \mathbf{\Omega}^c(1)), \tilde{\mathbf{r}} = \tilde{\mathbf{y}}_1$
3: **for** $j = 1$ to L_f **do**
4: $\quad \mathbf{C} = \texttt{unvec}_{\{M_{G_x}, M_{G_y}\}}(|\tilde{\mathbf{W}}_k^{\mathsf{H}}\tilde{\mathbf{r}}|)$
5: $\quad [\mathbf{q}_1(j), \mathbf{p}_1(j)] = \texttt{max}(\mathbf{C})$
6: $\quad \mathbf{s}_1(j) = (\mathbf{p}_1(j) - 1)M_{G_x} + \mathbf{q}_1(j)$
7: $\quad \tilde{\mathbf{h}}_1 = \tilde{\mathbf{W}}_k(:, \mathbf{s}_1)^{\dagger}\tilde{\mathbf{y}}_1$
8: $\quad \tilde{\mathbf{r}} = \tilde{\mathbf{y}}_1 - \tilde{\mathbf{W}}_k(, : \mathbf{s}_1)\tilde{\mathbf{h}}_1$
9: **end for**
10: $\hat{\tilde{\mathbf{H}}}_k(\mathbf{s}_1, \mathbf{\Omega}^c(1)) = \tilde{\mathbf{h}}_1$
% Estimate the remaining non-zero columns exploiting row-structure B
11: **for** $l = 2$ to L_G **do**
12: $\quad \tilde{\mathbf{y}}_l = \tilde{\mathbf{Y}}_k(:, \mathbf{\Omega}^c(l))$
// Obtain the indices of the non-zero elements in each non-zero column.
13: $\quad$ **for** $i = 1$ to L_f **do**
14: $\quad\quad \mathbf{p}_l(i) = \texttt{circ}[\mathbf{p}_1(i), \Delta\mathbf{p}^t(l-1)]_{M_{G_x}}$
$\quad\quad \mathbf{q}_l(i) = \texttt{circ}[\mathbf{q}_1(i), \Delta\mathbf{q}^t(l-1)]_{M_{G_y}}$
15: $\quad$ **end for**
16: $\quad \mathbf{s}_l = (\mathbf{p}_l - 1)M_{G_x} + \mathbf{q}_l$
17: $\quad \tilde{\mathbf{h}}_l = \tilde{\mathbf{W}}_k(:, \mathbf{s}_l)^{\dagger}\tilde{\mathbf{y}}_l$
18: $\quad \hat{\tilde{\mathbf{H}}}_k(\mathbf{s}_l, \mathbf{\Omega}^c(l)) = \tilde{\mathbf{h}}_l$
19: **end for**

2) The Proposed SS-OMP (Step 12 in Alg. 1): The proposed SS-OMP targets for CE when $\mathbf{G}_k$ remains unchanged. According to the row-structure B unveiled in Section III-C, $\{\Delta p_{l_g',l_g}^t, \Delta q_{l_g',l_g}^t\}$ are common for the sparse channels with the same RIS-BS channel. Hence, it is justifiable to leverage this feature to enhance the CE process further. Details of SS-OMP are listed in **Algorithm** 3. The shared spacing vectors $\{\Delta\mathbf{p}^t, \Delta\mathbf{q}^t\}$ required in SS-OMP are obtained via step 24 of TP-OMP in former blocks. After initialization, there are also two main parts of SS-OMP.

- **Estimation of $\tilde{\mathbf{h}}_1$:** SS-OMP starts by estimating a reference column via OMP in steps 2-10 (the same as Phase 1 of TP-OMP).
- **Estimation of the remaining columns by structured shift:** With the acquisition of $\tilde{\mathbf{h}}_1$ and the spacings $\{\Delta\mathbf{p}^t, \Delta\mathbf{q}^t\}$), steps 13-15 of SS-OMP successively obtain the indices of the non-zero elements in each of the remaining non-zero columns by cyclically shifting the positions according to $\{\Delta\mathbf{p}^t, \Delta\mathbf{q}^t\}$.

Remark 2: It should be noted that there is no need to first obtain the reference non-zero element ($h_{l,1}$) in estimating $\tilde{\mathbf{h}}_l (l = 2, \cdots, L_G)$ in SS-OMP. This is because the acquisition of $\tilde{\mathbf{h}}_1$ and the spacings are enough to obtain the indices of the remaining non-zero elements. For example, $p^{2,1}, \cdots, p^{2,L_f}$ can be directly calculated via (17) with known $p^{1,1}, \cdots, p^{1,L_f}$ and $\Delta p_{1,2}^t$. As a result, the complex coherence calculation (step 4

in Alg. 3) only needs to be repeated L_f times in SS-OMP.

C. Computational Complexity

The comparison of the computational complexity is listed in Table I[5]. To evaluate the computational complexity of TP-OMP and SS-OMP within a single block, the estimation of the non-zero columns supports is also included. In obtaining the non-zero columns' support, complexity mainly comes from calculating the power of N_G columns of $\tilde{\mathbf{Y}} \in \mathbb{C}^{N_G \times T}$. The resulting complexity is $\mathcal{O}(TN_G)$.

The coherence calculation step contributes the most to the complexity of OMP [29]. In TP-OMP, phase 1 requires L_f times of coherence calculation, leading to a complexity of $\mathcal{O}(L_f T M_G)$. The modification of coherence in step 15 adds a complexity of $\mathcal{O}((L_f - 1)M_G)$. For the recovery of the remaining non-zero columns in phase 2, the coherence calculation only needs to be conducted once for each column. In total, the computational complexity of TP-OMP is

$$\mathcal{O}((L_G + L_f - 1)TM_G) + \mathcal{O}(TN_G) + \mathcal{O}((L_f - 1)M_G). \quad (26)$$

SS-OMP only needs to repeat the coherence calculation L_f times to obtain $\tilde{\mathbf{H}}(:, \mathbf{\Omega}^c(1))$, which contributes to the complexity of $\mathcal{O}(L_f T M_G)$. Hence, the computational complexity of SS-OMP within a single block is $\mathcal{O}(L_f T M_G) + \mathcal{O}(TN_G)$.

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY

Algorithms	Complexity
Direct OMP [12]	$\mathcal{O}(L_G L_f T N M_G N_G)$
CS-EST [11]	$\mathcal{O}(L_G T N N_G) + \mathcal{O}(L_G^2 L_f T N M_G)$
DS-OMP [15]	$\mathcal{O}(TN_G) + \mathcal{O}(L_G L_f T M_G)$
BALS [10]	$\mathcal{O}(M(TK_M)) + \mathcal{O}(N(KT + KM + M^2))$[a]
FBMP [14]	$\mathcal{O}(TNM_G N_G P D)$[b]
TP-OMP (proposed)	(26)
SS-OMP (proposed)	$\mathcal{O}(L_f T M_G) + \mathcal{O}(TN_G)$

[a] The listed computational complexity is required for each iteration in BALS. Also, for BALS, $K \geq M, T \geq N$ should be satisfied.
[b] In FBMP, P, D respectively represent the stopping parameter and search parameter. Please refer to [14] for more details.

D. Comparisons with Existing CE Schemes

A comparison between our proposed adaptive CE scheme and other schemes is presented in Table II. In terms of pilot transmission, the two-timescale scheme [21] adopted the dual-link transmission between BS and RIS, requiring beam search for the RIS to reflect the pilots towards the BS. In the anchor-assisted approach [25], training pilots are exchanged between the BS, anchor-1 (A1), and anchor-2 (A2) to obtain the CSI. In the TAD-CE scheme [22], uplink pilots from adjacent BS are required to conduct the RIS-BS angle estimation before the UE connection. Pilots from UE are then needed for the following stages. The CE scheme proposed in [24] requires the RIS elements to have signal processing capabilities, and thus semi-passive RIS elements are utilized. The BS first sends pilots to the RIS in a large timescale to estimate the RIS-BS channel. To estimate the UE-RIS channel in each block, UE sends pilots to RIS. In our proposed methods, only the uplink pilots from the UE are needed.

While the aforementioned CE schemes leverage the two-timescale property of RIS-assisted channels to lower the pilot overhead, a common limitation is that they assume prior knowledge of the timescales for channel variation and lack adaptivity. On the contrary, the proposed adaptive scheme in this paper does not make ideal assumptions about the unchanged duration of $\mathbf{G}$. The inclusion of the check process allows for the timely selection of the appropriate CE algorithms. Moreover, this process does not require any additional hardware or involve more complex transmission schemes.

[5]Please refer to specific references for detailed derivations of the computational complexity.

V. EXTENSION TO OFF-GRID METHODS

The CE algorithms discussed above rely on the assumption that the AoAs and AoDs align precisely with discrete predetermined grids. Nonetheless, in reality, angles are continuously distributed. Consequently, when mapping the continuous angles onto discrete grids, off-grid errors (also referred to as quantization errors) occur, which inevitably lead to a degradation in CE performance.

There are numerous methods available to mitigate off-grid errors [30]–[32]. While utilizing a finer grid interval can lower the off-grid errors. However, increasing the number of grid points significantly raises computational complexity. Moreover, it leads to a higher correlation within the sensing matrix, which subsequently degrades the performance of compressed sensing (CS)-based algorithms. Gridless approaches, such as atomic norm minimization [30], suffer from computational complexity that restricts their practical applicability in RIS-assisted time-varying CE. A simpler alternative is to incorporate optimizations into on-grid methods to alleviate the impact of off-grid errors. For instance, we have proposed a discrete-continuous OMP (DC-OMP) [32] has been proposed, which addresses this problem with relatively lower complexity. Therefore, we adopt the idea presented in [32] to improve and show the availability of our proposed methods in this paper.

A. Optimized Column Estimation

For the column estimation in Algorithm 1, due to the off-grid error, obtaining the indices by simply calculating each columns energy may no longer be effective. Instead, we estimate the L_G indices of the non-zero columns in $\tilde{\mathbf{H}}$ (equivalent to the non-zero rows in $\bar{\mathbf{H}}$ since $\tilde{\mathbf{H}} = \bar{\mathbf{H}}^H$) one by one with an optimized OMP. Applying the VAD representation, (7) can be rewritten as

$$\mathbf{Y} = \mathbf{U}_N \bar{\mathbf{H}} \mathbf{U}_M^{\mathsf{T}} \mathbf{W} = \mathbf{U}_N \mathbf{X} + \mathbf{N}, \quad (27)$$

where $\mathbf{X} \triangleq \bar{\mathbf{H}} \mathbf{U}_M^{\mathsf{T}}$. It is easy to obtain that $\mathbf{X}$ is a row-sparse matrix with L_G non-zero rows. Consequently, (27) can be viewed as a CS model, where $\mathbf{U}_N$ is the sensing matrix, $\mathbf{Y}$ is the measurement matrix, $\mathbf{X}$ is the row-sparse signal to be recovered. To mitigate the quantization error, we include an optimization step when applying OMP to estimate the indices of the non-zero columns in $\tilde{\mathbf{H}}$. Details of the optimized column estimation are summarized in Algorithm 4.

In step 3, similar to the conventional OMP, the index (s_l) of the most correlated column in the sensing matrix

TABLE II
COMPARISON OF DIFFERENT CE SCHEMES

Different CE schemes	Features			
	Pilot transmission	Processing nodes	No assumption on τ	CE algorithm adjustment
The two-timescale scheme [21]	BS-RIS-BS and UE-RIS-BS	BS	×	×
The anchor-assisted scheme [25]	A1-RIS-BS, A1-RIS-A2 and A2-RIS-UE	BS, UE and the anchors	×	×
TAD-CE [22]	Adjacent BS-RIS-BS, UE-RIS-BS	BS	×	×
Semi-passive elements scheme [24]	BS-RIS and UE-RIS	RIS	×	×
The adaptive scheme (proposed)	UE-RIS-BS	BS	✓	✓

Algorithm 4 Optimized column estimation in the k-th block

Input: $\mathbf{Y}_k$, $\tilde{\mathbf{U}}_N$, L_G
Output: estimated $\mathbf{\Omega}^c$ and modified dictionary $\tilde{\mathbf{U}}_N$.

% initialization
1: $\mathbf{R} = \mathbf{Y}_k$, $\mathbf{\Omega}_k^c = \varnothing$
2: **for** $l = 1 : L_G$ **do**
3: $s_l = \arg\max_{1 \le s_l \le N_G} \|(\tilde{\mathbf{U}}_N^{\mathsf{H}} \mathbf{R})_{s_l,:}\|_F^2$
4: $\mathbf{\Omega}_k^c = \mathbf{\Omega}_k^c \cup s_l$
5: $\tilde{\theta}_l = \arg\max_{\tilde{\theta}_l} f_1(\tilde{\theta}_l)$ ▷ Refine the estimated angle
6: $\tilde{\mathbf{U}}_N(:, s_l) = \mathbf{b}_N(\tilde{\theta}_l)$ ▷ Update the dictionary
7: $\mathbf{A} = \tilde{\mathbf{U}}_N(:, \mathbf{\Omega}_k^c)$
8: $\hat{\mathbf{X}} = \mathbf{A}^{\dagger}\mathbf{Y}$
9: $\mathbf{R} = \mathbf{Y} - \mathbf{A}\hat{\mathbf{X}}$
10: **end for**

with the residual is found. In step 4, the estimated column set is augmented. Consequently, the corresponding estimated spatial angle at the l-th iteration is $\theta_{s_l} = -\frac{1}{2} + \frac{(s_l - 1)}{N_G}$. Due to the off-grid effect, in step 3, there may exist a larger coherence value that corresponds to a continuous-distributed angle near the discrete grid. Specifically, it is easy to obtain that $(\tilde{\mathbf{U}}_N \mathbf{R})_{s_l,:} = \mathbf{b}_N^{\mathsf{H}}(\theta_{s_l})\mathbf{R}$. The off-grid angle $\tilde{\theta}_l$ satisfy $\|\mathbf{b}_N^{\mathsf{H}}(\tilde{\theta}_l)\mathbf{R}\|_F^2 > \|\mathbf{b}_N^{\mathsf{H}}(\theta_{s_l})\mathbf{R}\|_F^2$, where $\tilde{\theta}_l$ is not in the pre-determines grids.

Thus, an optimized angle can be obtained by solving a non-linear constrained continuous optimization problem in step 5. Similar to [32], the optimization problem can be formulated as

$$\begin{aligned} \max_{\tilde{\theta}_l} \quad & f_1(\tilde{\theta}_l) = \|\mathbf{b}_N^H(\tilde{\theta}_l)\mathbf{R}\|_F^2 \\ s.t. \quad & \begin{cases} \theta_{s_l} = -\frac{1}{2} + \frac{(s_l - 1)}{N_G} \\ |\tilde{\theta}_l - \theta_{s_l}| \le \frac{1}{2N_G}. \end{cases} \end{aligned} \tag{28}$$

With the obtained optimized angle $\tilde{\theta}_l$, the sensing matrix can be updated with the corrected angles, i.e., $\theta_{s_l} = \tilde{\theta}_l$. Note that $\tilde{\mathbf{U}}_N$ is initialized as $\tilde{\mathbf{U}}_N = \mathbf{U}_N$. In this way, we adjust the grid point and sensing matrix in every iterative step to find more accurate angles and corresponding VAD representation of $\mathbf{H}$. Following the optimized column estimation, the check process can be adjusted accordingly.

B. Optimized TP-OMP and SS-OMP

Similar to the optimized column estimation in Section V-A, the estimated on-grid angles in TP-OMP and SS-OMP can be corrected to the off-grid ones. For example, in the optimized TP-OMP (O-TP-OMP), the results obtained from step 7 and step 16 in Algorithm 2 can be modified by solving the following optimization problem.

$$\begin{aligned} \max_{\tilde{\upsilon}_i^j, \tilde{\rho}_i^j} \quad & f_2(\tilde{\upsilon}_i^j, \tilde{\rho}_i^j) = \left| \left(\mathbf{W}^{\mathsf{H}} \mathbf{b}_{M_x}(\tilde{\upsilon}_i^j) \otimes \mathbf{b}_{M_y}(\tilde{\rho}_i^j) \right)^{\mathsf{H}} \tilde{\mathbf{r}} \right| \\ s.t. \quad & \begin{cases} \upsilon_i^j = -\frac{1}{2} + \frac{(p_i^j - 1)}{M_{G_x}} \\ \rho_i^j = -\frac{1}{2} + \frac{(q_i^j - 1)}{M_{G_y}} \\ |\tilde{\upsilon}_i^j - \upsilon_i^j| \le \frac{1}{2M_{G_x}} \\ |\tilde{\rho}_i^j - \rho_i^j| \le \frac{1}{2M_{G_y}}, \end{cases} \end{aligned} \tag{29}$$

where p_i^j, q_i^j are respectively the j-th entries in $\mathbf{p}_i, \mathbf{q}_i$. With the modified angles $\tilde{\upsilon}_i^j, \tilde{\rho}_i^j$, the corresponding column of $\mathbf{U}_M$ should be updated to $\mathbf{b}_{M_x}(\tilde{\upsilon}_i^j) \otimes \mathbf{b}_{M_y}(\tilde{\rho}_i^j)$ to form the updated dictionary $\tilde{\mathbf{U}}_{M,i}$ accordingly.

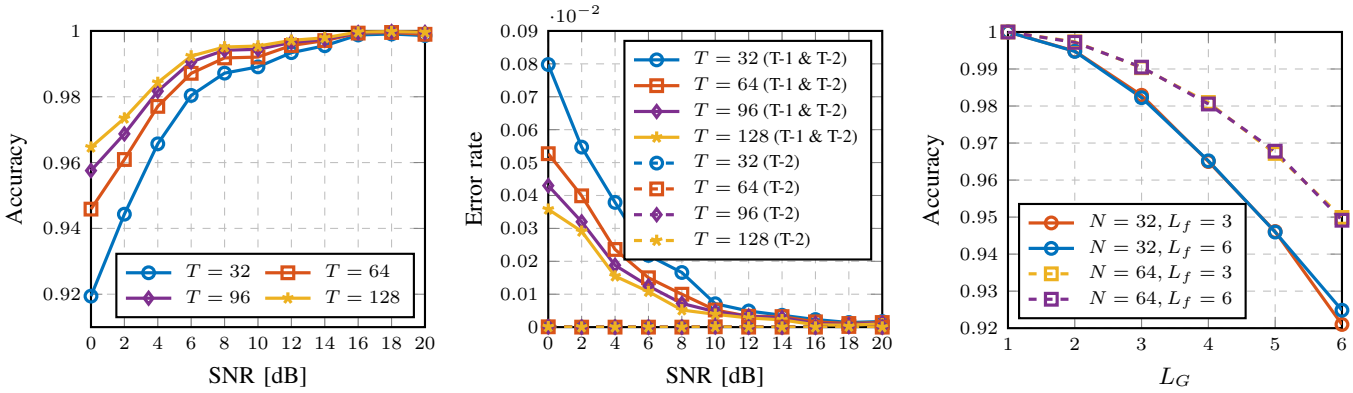
In O-TP-OMP, angular differences in (13) and (14) should be computed. Subsequently, the coherence modification of O-TP-OMP can then be adjusted according to the obtained angular differences. The cyclic shift can be replaced by adding the coherence computed from shifted dictionaries. Importantly, with the optimized dictionaries, $\mathbf{H}$ should be constructed as

$$\mathbf{H} = \sum_{i=1}^{L_G} \tilde{\mathbf{U}}_N(:, \mathbf{\Omega}^c(l)) \hat{\mathbf{h}}_i^{\mathsf{H}} \mathbf{U}_{M,i}^{\mathsf{H}}, \tag{30}$$

where $\mathbf{U}_{M,i}$ represents the matrix constructed by the steering vectors of the estimated angles for each non-zero column. Optimized SS-OMP (O-SS-OMP) can be similarly designed to obtain the off-grid optimized channel estimation.

VI. SIMULATION RESULTS

In this section, the effectiveness of the proposed CE approaches is evaluated through numerical simulations in terms of the accuracy of the check process, CE performance of the TP-OMP and SS-OMP in a single block, the long-term CE performance of the adaptive CE scheme, and the performance of the optimized off-grid methods. In our simulations, the reflection pattern of RIS is randomly selected from $[-\frac{1}{\sqrt{M}}, \frac{1}{\sqrt{M}}]$. Other simulation parameters are listed in Table III. $N_G = N$, $M_{G_x} = M_x$, $M_{G_y} = M_y$ are assumed in the simulations. The normalized mean squared error (NMSE) is defined as NMSE $= \mathbb{E}\{\|\hat{\mathbf{H}}_k - \mathbf{H}_k\|_F^2 / \|\mathbf{H}_k\|_F^2\}$ for CE performance comparison. The signal-to-noise ratio (SNR) is defined as SNR $= \mathbb{E}\{\|\tilde{\mathbf{W}}_k \tilde{\mathbf{H}}_k\|_F^2 / \|\tilde{\mathbf{N}}_k\|_F^2\}$.



(a) Check accuracy vs. SNR. (b) Error rate of check vs. SNR. (c) Check accuracy vs. L_G.

Fig. 4. Performances of the check process. In Fig. 4(b), T-1 and T-2 respectively represent the type-1 and type-2 errors. In Fig. 4(c), SNR=5 dB.

TABLE III
SIMULATION PARAMETERS

Parameters	Values
(M, M_x, M_y)	$(256, 16, 16)$
(d_{RB}, d_{UR})[a]	$(20m, 20m)$
$(\lvert\alpha_{l_g}\rvert, \lvert\beta_{l_f}\rvert)$	$(10^{-2}d_{RB}^{-2.1}, 10^{-2}d_{UR}^{-4.2})$

[a] d_{RB} and d_{UR} respectively represent the distance between the BS and the RIS and the distance between the RIS and the UE.

A. Evaluation of the Check Process

Fig. 4(a) illustrates the accuracy of the proposed check process at various SNRs with different pilot overhead. The number of paths for $\mathbf{G}_k$ and $\mathbf{f}_k$ are respectively assumed to be $L_G = 3$ and $L_f = 6$. The BS is considered to be equipped with $N = 32$ antennas. It can be found a higher accuracy of the check process at higher SNRs. In addition, with an increased number of pilots, the accuracy grows. But even under relatively unfavorable conditions, such as $T = 32$, SNR=0 dB, the accuracy remains approximately 92%. When SNR=20 dB, the accuracy approaches 100% for all pilot conditions. In Fig. 4(b), the error rate of the check process is depicted. Specifically, when the check process fails, there are two types of errors:

- **Type-1 Error**: If the check process regards an unchanged RIS-BS channel as changed, TP-OMP should be utilized. In this case, the computational complexity cannot be further reduced (as TP-OMP exhibits higher complexity than SS-OMP), but the CE performance can be guaranteed.
- **Type-2 Error**: If the RIS-BS channel has changed but the check process fails to detect the variation, then the type-2 error will occur. Note that this type of error can not be 100% eliminated and it also exists in other existing works (e.g., the two-timescale scheme [21]). However, the likelihood of such errors is significantly lower compared to Type-1 errors. As depicted in Fig. 4(b), Type-2 error rates for $T = \{32, 64, 96, 128\}$ are all very low. The majority of errors made by the check process are Type-1, whereas the probability of Type-2 error is nearly zero. This is because we have designed strict rules for the check process to make the "unchanged" verdict. More specifically, the check process will make the "$\mathbf{G}_k$ is unchanged" decision only when $\texttt{diff}(\mathbf{\Omega}_k^c, \mathbf{\Omega}^c) = \varnothing$.

Fig. 4(c) reveals a clear relationship between the number of antennas, the number of RIS-BS and UE-RIS channel paths, and the corresponding accuracy. Firstly, the results indicate that, with the same N and L_f, the accuracy tends to decline when increasing L_G. Because it is more difficult for the column estimation to distinguish more non-zero columns from the total N columns. Additionally, due to the same reason, there is a noticeable improvement in the accuracy when N increases with constant L_G. Fortunately, it is reasonable that $L_G \ll N$ in the sparse mmWave channels. Consequently, the accuracy of the check process can be guaranteed.

In summary, as shown in Fig. 4, the proposed check process is effective in detecting the changing status of $\mathbf{G}_k$.

B. Performances of the Proposed Algorithms in Single Block

In this subsection, we compare the NMSE performance of the two proposed algorithms (TP-OMP and SS-OMP) with the direct OMP [12], CS-EST [11], DS-OMP [15], BALS [10], and FBMP [14]. Specifically, the SS-OMP is exploited with known $\Delta\mathbf{p}^t$ and $\Delta\mathbf{q}^t$ (estimated in the CE for former blocks). In our simulations, the oracle LS provides the best achievable CE performance with the perfect knowledge of the supports of non-zero elements. To better illustrate the performance of the algorithms, we also derived the theoretical lower bound of estimation mean square error (MSE) for OMP-related methods as follows.

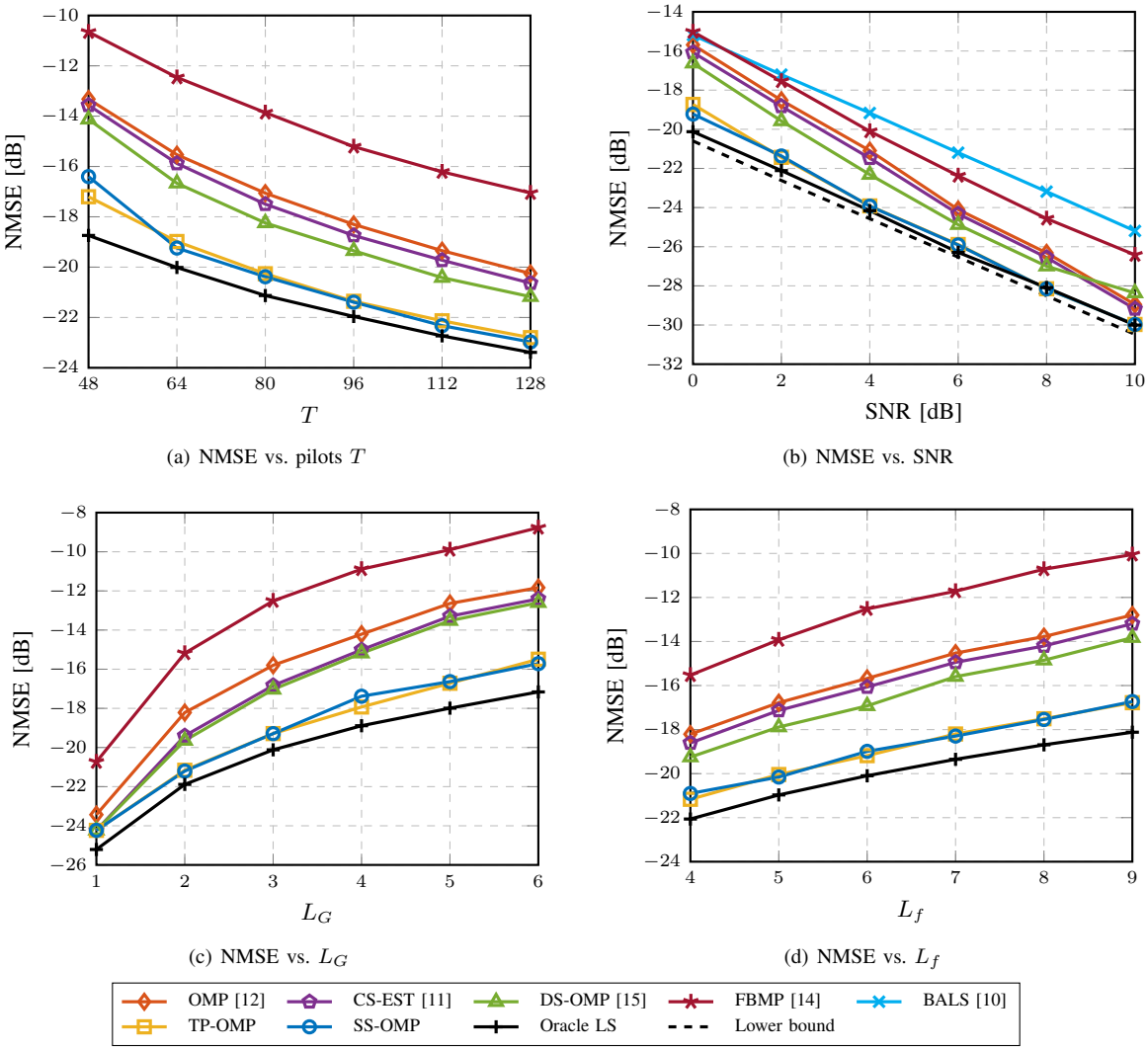
Proposition 3: For the OMP-related CE methods, supposing the ideal VAD representations of the channels are known, the lower bound of estimation mean square error (MSE) is

$$\mathbb{E}\{\|\texttt{vec}(\hat{\mathbf{H}}) - \texttt{vec}(\mathbf{H})\|_F^2\} = \frac{\lambda_{min}(\mathbf{D}^{\mathsf{H}}\mathbf{D})\sigma_n^2 L_G^2 L_f^2}{\texttt{Tr}(\mathbf{A}^{\mathsf{H}}\mathbf{A})}, \quad (31)$$

where $\mathbf{D} \triangleq (\mathbf{U}'_M \otimes \mathbf{U}'_N)$, $\mathbf{U}'_M, \mathbf{U}'_N$ are the corresponding ideal dictionaries, λ_{min} represents the minimum eigenvalue of $(\mathbf{D}^{\mathsf{H}}\mathbf{D})$. $\mathbf{A} \triangleq (\mathbf{W}^{\mathsf{T}}\mathbf{U}'_M \otimes \mathbf{U}'_N)$.

Proof: Please refer to Appendix C.

In Fig. 5(a) and Fig. 5(b), the simulation parameters are set as $N = 32, L_G = 3, L_f = 6$. Fig. 5(a) shows the NMSE performance versus pilots of different algorithms, where SNR=0 dB.



(a) NMSE vs. pilots T

(b) NMSE vs. SNR

(c) NMSE vs. L_G

(d) NMSE vs. L_f

Fig. 5. CE performances within a single block.

It can be found that both the proposed TP-OMP and SS-OMP outperform the compared algorithms (OMP, CS-EST, DS-OMP, and FBMP). Specifically, the TP-OMP and SS-OMP require about 33% fewer pilots compared with DS-OMP to achieve an NMSE of -19 dB. This is because TP-OMP and SS-OMP respectively utilize the inherent row structure A and B of the cascaded channel as the prior information. Consequently, the CE performance can be lifted.

In Fig. 5(b), $T = 64$ are provided for the CS-based algorithms, and a total of $MN = 8192$ pilots are given for BALS. Obviously, compared with the tensor-based methods, leveraging the sparsity of mmWave channels contributes significantly to the reduction of pilots. In Fig. 5(b), TP-OMP and SS-OMP show a more pronounced superiority at lower SNRs. However, even in high SNRs, both TP-OMP and SS-OMP consistently achieve lower NMSE compared to other benchmark methods. Notably, the NMSE performance of TP-OMP and SS-OMP can approach that of Oracle LS at higher SNRs. Considering that SNR is typically unknown in communication, using TP-OMP and SS-OMP is recommended to achieve better performance with fewer pilots. It can also be calculated from Table I that TP-OMP and SS-OMP respectively obtain about 55% and 66% reduction of the computational complexity compared with DS-OMP in this setup. Considering the requirements of communication system applications, the lower complexity of TP-OMP and SS-OMP ensures their practical feasibility.

Fig. 5(c) and Fig. 5(d) respectively show the NMSE performance against the path number of RIS-BS (L_G) and UE-RIS (L_f) channel, where $N = 32$, SNR= 0 dB, $T = 64$. In Fig. 5(c), it can be seen that TP-OMP and SS-OMP achieve more significant improvement with larger L_G. Besides, both Fig. 5(c) and Fig. 5(d) present similar trends, highlighting the efficiency of exploiting row structures to achieve improved CE performance when dealing with increasing path numbers.

C. Long-term Performance of the Adaptive CE Scheme

To evaluate the adaptivity of the proposed scheme, simulations for multiple blocks are included with comparative analysis against the TP-OMP scheme. Specifically, the TP-OMP scheme employs the TP-OMP in each block during

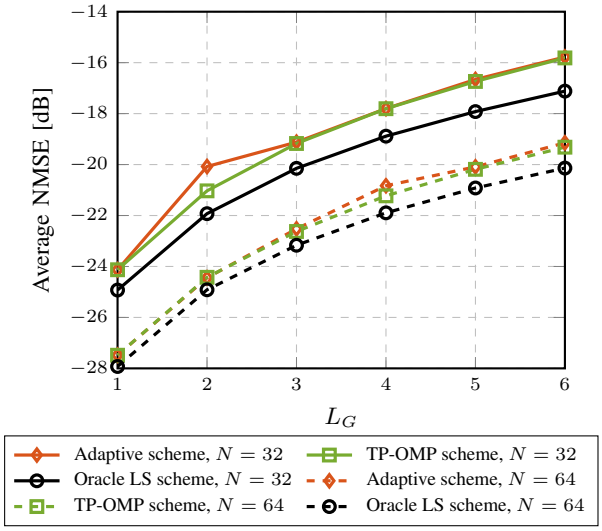
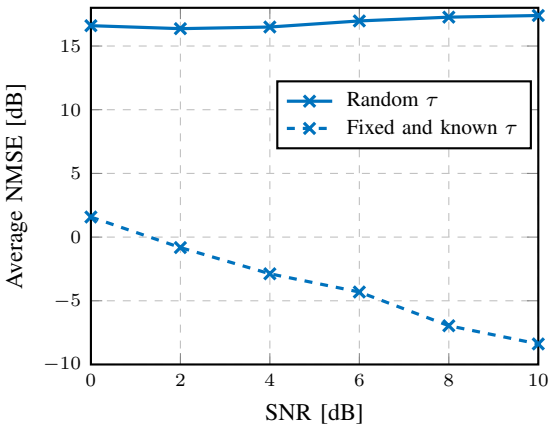


Fig. 6. Average NMSE vs. L_G.

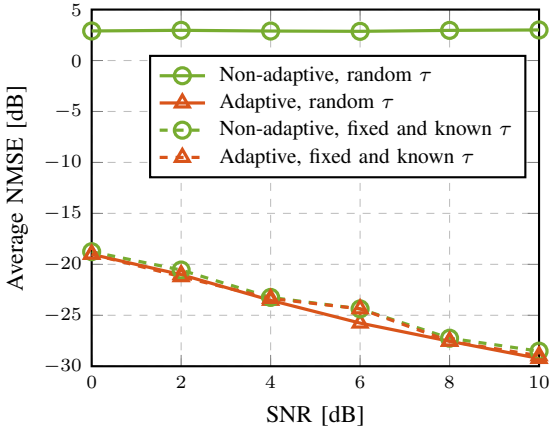
the long-term CE process. In contrast, the proposed adaptive scheme can select the CE algorithm (TP-OMP or SS-OMP) according to the check result. Additionally, we include the Oracle LS scheme, which is assumed to know the support of the channel matrix in each block. Average NMSE is defined as the average of multiple block CE NMSE and is regarded as a metric to evaluate the long-term performance of the compared schemes. In the simulations, if not specified, we consider the probability of $\mathbf{G}_k$ and $\mathbf{f}_k$ varying in each block to be 20% and 100%, respectively.

Fig. 6 presents the average NMSE performance of different schemes concerning the RIS-BS channel paths L_G with various numbers of BS antennas, where $T = 64$, SNR= 0 dB and $L_f = 6$. It can be found that the adaptive scheme and TP-OMP scheme have the same performance when $L_G = 1$. Because they exploit the same estimation process of a single non-zero column. According to the analysis of Fig. 4(c), it is known that the larger L_G results in lower accuracy of the check process. As a result, when L_G grows, the performances of all schemes degrade. Notably, the performance of the adaptive scheme is very close to the TP-OMP scheme. This is because there does exist little error propagation among blocks of the adaptive scheme. Also in Fig. 6, it can be seen that even with a decreased accuracy of the check process (larger L_G according to Fig. 4(c)), the performance of the adaptive scheme still approaches that of the TP-OMP scheme. This further indicates that the errors resulting from the check process can be interrupted and corrected by choosing the TP-OMP algorithm which does not rely on CSI from the former blocks. Fig. 6 confirms the necessity of the strict check rule and the reliability of the proposed adaptive scheme.

Fig. 7 presents the average NMSE performance of the two-timescale scheme [21] and the proposed adaptive schemes against different SNRs, where $N = 32$. The pilots for the large-timescale and small-timescale CE are respectively $ML = 512$ and $M/N = 8$, where $L = 2$ is the transmitter antennas at the BS [21]. For the proposed schemes, 64 pilots are used.



(a) The two-timescale scheme



(b) The proposed methods

Fig. 7. Average NMSE vs. SNR.

It should be noted that to demonstrate the adaptability of the proposed adaptive scheme, we also compare the proposed adaptive scheme with another scheme named the non-adaptive scheme. The non-adaptive scheme utilizes TP-OMP or SS-OMP at a fixed mode, i.e., TP-OMP is employed in the $(n\tau + 1)$-th block ($n \geq 0$) and SS-OMP is employed in the remaining blocks. For the two-timescale and non-adaptive scheme, we assume the high-dimensional estimation of $\mathbf{G}$ and SS-OMP are conducted every 5 block in the simulations. From Fig. 7, it can be seen that when the channel variation follows the assumed mode, both the two-timescale scheme and the non-adaptive scheme can obtain desirable average NMSE performance. However, in practical communications, the channels do not change regularly. As a result, when the assumption does not hold (random τ in the simulations), the performances of these two schemes drop dramatically. For a fair comparison, Fig. 7(b) illustrates the NMSE performance of the proposed adaptive scheme with random τ and fixed $\tau = 5$. The performances of the adaptive scheme under different channel changing modes (fixed $\tau = 5$ and random τ) are very close, which demonstrates the stability and practicality of the proposed adaptive scheme.

In conclusion, for the long-term CE, the proposed adaptive CE scheme can achieve desirable performance at the cost

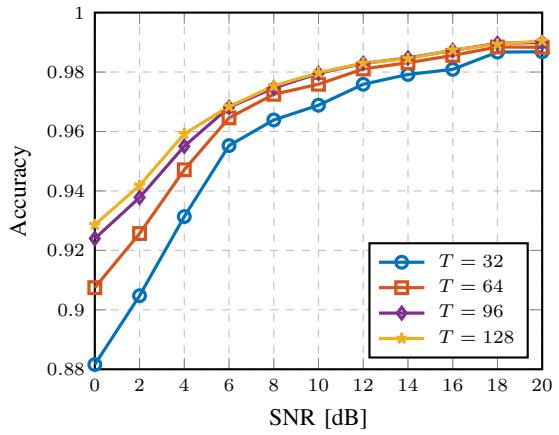


Fig. 8. Accuracy of the optimized check vs. SNR.

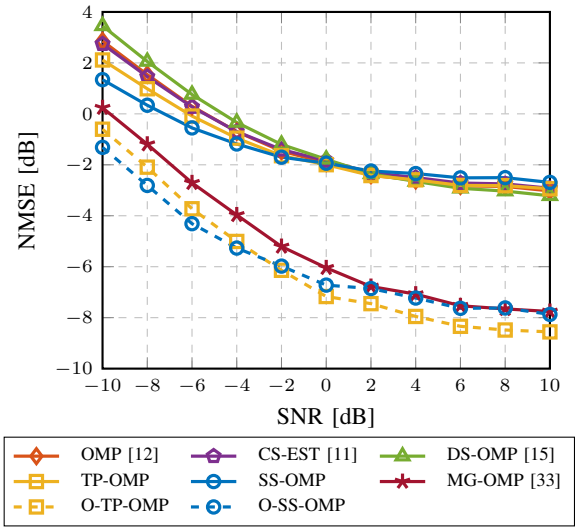


Fig. 9. NMSE vs. SNR (off-grid channels).

of much lower computational complexity and fewer pilots compared with existing methods. More importantly, the adaptive scheme can maintain good performance even when the variation rate of $\mathbf{G}_k$ is unknown.

D. Evaluation of the off-grid methods

Simulation results of the proposed methods with optimizations to combat the off-grid errors are depicted in Fig. 8 and Fig. 9. It can be seen from Fig. 8 that the accuracy of the optimized column estimation and check process still exceeds 88% and approaches 100% with increased SNR and pilots.

To evaluate the performance of the proposed O-TP-OMP and O-SS-OMP, we include the widely used multi-grid OMP (MG-OMP) [33] for comparison. Specifically, in each iteration, MG-OMP starts with a coarse grid estimated by OMP and then adopts a finer grid interval around the coarse grid to search for a more accurate grid point. Fig. 9 reveals that TP-OMP and SS-OMP still outperform other on-grid algorithms at lower SNRs in off-grid scenarios. It can also be seen from Fig. 9 that the O-TP-OMP and O-SS-OMP can alleviate quantization errors and achieve superior performance compared with MG-OMP. Additionally, it is viable for our proposed methods to incorporate alternate off-grid mitigation methods to achieve enhanced performance.

In terms of computational complexity, for MG-OMP, similar to OMP [12], each coherence calculation in MG-OMP results in a complexity of $\mathcal{O}(TNM_GN_G)$. To obtain an estimated angle, MG-OMP requires multiple times coherence calculations in different angular resolutions. Furthermore, the sensing matrix has to be updated in each iteration by calculating the high-dimensional Kronecker product, which also leads to additional cost of computational complexity. Both O-TP-OMP and O-SS-OMP have higher computational complexity compared with their on-grid versions. For O-TP-OMP and O-SS-OMP, by dividing the CE into multiple stages, the coherence calculation in each iteration has a complexity of $\mathcal{O}(M_GT)$. In contrast to the MG-OMP, O-TP-OMP and O-SS-OMP eliminate the need for repeated coherence calculation. Thus, by solving the optimization problem within a relatively small range, both O-TP-OMP and O-SS-OMP have much lower computational complexity than MG-OMP.

VII. Conclusion

In this paper, we proposed an adaptive CE scheme for the RIS-assisted systems in time-varying mmWave channels. Firstly, we analyzed the inherent structures of the UE-RIS-BS channel in the angular domain, which provides the theoretical foundation for the design of the proposed methods. Within the proposed two-stage adaptive CE scheme, a check process is included in the first stage to detect if the RIS-BS channel has changed among blocks. By utilizing the result of the check process, the adaptive scheme can dynamically select the appropriate CE algorithms from two proposed algorithms: TP-OMP and SS-OMP. The adaptive scheme maintains robust performance even when the channel variation rate is unknown. Simulation results confirm the adaptive CE scheme's ability to achieve desirable long-term CE performance with reduced computational complexity and pilot overhead.

Appendix A
The Proof of Proposition 1

Specifically,

$$\begin{aligned}\tilde{\mathbf{b}}_N^{\mathsf{H}}(\theta_{l_g}^{G_r}) =& \mathbf{b}_N^{\mathsf{H}}(\theta_{l_g}^{G_r})[\mathbf{b}_N(\theta_1),\cdots,\mathbf{b}_N(\theta_{N_G})] \\ \overset{(a)}{=}& [x_1,\cdots,|\mathbf{b}_N(\theta_n)|^2,\cdots,x_N],\end{aligned} \tag{32}$$

where in (a),

$$x_n = \frac{1}{N}\sum_{i=1}^{N} e^{j2\pi(i-1)(\upsilon_{l_g}^{G_r}-\upsilon_n)} = \frac{1}{N}\frac{1-e^{j2\pi N(\upsilon_{l_g}^{G_r}-\upsilon_n)}}{1-e^{j2\pi(\upsilon_{l_g}^{G_r}-\upsilon_n)}}. \tag{33}$$

And

$$\lim_{N\to\infty} x_n = \begin{cases} 0, & \text{if } \upsilon_{l_g}^{G_r} \neq \upsilon_n \\ 1, & \text{if } \upsilon_{l_g}^{G_r} = \upsilon_n \end{cases}. \tag{34}$$

Thus, $x_i(i = 1,\cdots,N_G, i \neq n)$ is negligible compared with $x_n = |\mathbf{b}_N(\theta_n)|^2 = 1$. Hence, index of the only one non-zero element in $\tilde{\mathbf{b}}_N^{\mathsf{H}}$ is determined by $\theta_{l_g}^{G_r}$.

For UPA, the steering vector in the angular domain is

$$\begin{aligned}\tilde{\mathbf{a}}_M(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r},\rho_{l_g}^{G_t}+\rho_{l_f}^{r}) &= \mathbf{U}_M^{\mathsf{H}}\mathbf{a}_M(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r},\rho_{l_g}^{G_t}+\rho_{l_f}^{r})\\ &= (\mathbf{U}_{M_x}^{\mathsf{H}}\otimes\mathbf{U}_{M_y}^{\mathsf{H}})\left(\mathbf{b}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})\otimes\mathbf{b}_{M_y}(\rho_{l_g}^{G_t}+\rho_{l_f}^{r})\right)\\ &\overset{(a)}{=}\left(\mathbf{U}_{M_x}^{\mathsf{H}}\mathbf{b}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})\right)\otimes\left(\mathbf{U}_{M_y}^{\mathsf{H}}\mathbf{b}_{M_y}(\rho_{l_g}^{G_t}+\rho_{l_f}^{r})\right)\\ &\overset{(b)}{=}\tilde{\mathbf{b}}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})\otimes\tilde{\mathbf{b}}_{M_y}(\rho_{l_g}^{G_t}+\rho_{l_f}^{r}),\end{aligned} \tag{35}$$

where (a) is obtained by following $\mathbf{AB}\otimes\mathbf{CD}=(\mathbf{A}\otimes\mathbf{C})(\mathbf{B}\otimes\mathbf{D})$. In (b), $\tilde{\mathbf{b}}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})\triangleq\mathbf{U}_{M_x}^{\mathsf{H}}\mathbf{b}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})$, $\tilde{\mathbf{b}}_{M_y}(\rho_{l_g}^{G_t}+\rho_{l_f}^{r})\triangleq\mathbf{U}_{M_y}^{\mathsf{H}}\mathbf{b}_{M_y}(\rho_{l_g}^{G_t}+\rho_{l_f}^{r})$. As been proved, both $\tilde{\mathbf{b}}_{M_x}$ and $\tilde{\mathbf{b}}_{M_y}$ have one non-zero element respectively depending on $\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r}$ and $\rho_{l_g}^{G_t}+\rho_{l_f}^{r}$. Therefore, there exists only one non-zero element in $\tilde{\mathbf{a}}_M(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r},\rho_{l_g}^{G_t}+\rho_{l_f}^{r})$. ■

APPENDIX B
THE PROOF OF PROPOSITION 2

Based on the proof of Proposition 1, it is easy to find that the column index of the non-zero element related to the l_gl_f-th subpath is n_{l_g}. The following discusses how the row coordinate is determined. According to (2), the steering vector of the cascaded azimuth angle at the RIS is

$$\begin{aligned}\mathbf{b}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r}) &= \frac{1}{\sqrt{M_x}}[e^{j2\pi\mathbf{m}_x(-\frac{1}{2}+\frac{m_x^{l_g}-1}{M_{G_x}}-\frac{1}{2}+\frac{m_x^{l_f}-1}{M_{G_x}})}]^{\mathsf{T}}\\ &= \frac{1}{\sqrt{M_x}}[e^{j2\pi\mathbf{m}_x(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}})}]^{\mathsf{T}},\end{aligned} \tag{36}$$

where $\mathbf{m}_x=[0,\cdots,M_x-1]$. Further, $\tilde{\mathbf{b}}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})=\mathbf{U}_{M_x}^{\mathsf{H}}\mathbf{b}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})$ can be represented as

$$\begin{aligned}\tilde{\mathbf{b}}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r}) &= [\mathbf{b}_{M_x}(\upsilon_1),\cdots,\mathbf{b}_{M_x}(\upsilon_{M_{G_x}})]^{\mathsf{H}}\\ &\quad\times\mathbf{b}_{M_x}(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}}),\end{aligned} \tag{37}$$

where $\upsilon_{m_x}=-\frac{1}{2}+\frac{m_x-1}{M_{G_x}}$, $m_x\in[1,M_{G_x}]$. $(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}})\in[-1,1-\frac{2}{M_{G_x}}]$ since $(m_x^{l_g}+m_x^{l_f})\in[2,2M_{G_x}]$.

Assuming the index of the non-zero element in $\tilde{\mathbf{b}}_{M_x}(\upsilon_{l_g}^{G_t}+\upsilon_{l_f}^{r})$ is $p^{l_gl_f}$, according to (34), it must satisfy

$$\mathbf{b}_{M_x}(\upsilon_{p^{l_gl_f}})=\mathbf{b}_{M_x}(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}}). \tag{38}$$

Since $e^{j2\pi\mathbf{m}_x(\cdot)}$ is a periodic function with a period of 1, we can always find a $\upsilon_{p^{l_gl_f}}$ to meet the requirement in (38) by circular shifting for any $M_x=2n, n\geq1$. We discuss the value of $p^{l_gl_f}$ in different cases as follows.

Case 1: If $(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}})\in[-1,-\frac{1}{2})$, i.e., $(m_x^{l_g}+m_x^{l_f})\in[2,\frac{M_{G_x}+2}{2})$, then

$$\begin{aligned}\mathbf{b}_{M_x}(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}}) &= \frac{1}{\sqrt{M_x}}e^{j2\pi\mathbf{m}_x(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_x})}\\ &\overset{(a)}{=}\frac{1}{\sqrt{M_x}}e^{j2\pi\mathbf{m}_x(-\frac{1}{2}+\frac{m_x^{l_g}+m_x^{l_f}+\frac{M_{G_x}-2}{2}-1}{M_{G_x}})}\end{aligned} \tag{39}$$

where (a) can be obtained since $e^{j2\pi\mathbf{m}_xx}=e^{j2\pi\mathbf{m}_x(x+1)}$. Conseuquently, $p^{l_g,l_f}=(m_x^{l_g}+m_x^{l_f}+\frac{M_{G_x}-2}{2})$ in this case.

Case 2: If $(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}})\in[-\frac{1}{2},\frac{1}{2}-\frac{1}{M_{G_x}}]$, i.e., $(m_x^{l_g}+m_x^{l_f})\in[\frac{M_{G_x}}{2}+2,\frac{3M_{G_x}}{2}+1]$, then

$$\begin{aligned}\mathbf{b}_{M_x}(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}}) &= \frac{1}{\sqrt{M_x}}e^{j2\pi\mathbf{m}_x(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}})}\\ &= \frac{1}{\sqrt{M_x}}e^{j2\pi\mathbf{m}_x(-\frac{1}{2}+\frac{m_x^{l_g}+m_x^{l_f}-\frac{M_{G_x}+2}{2}-1}{M_{G_x}})}\end{aligned} \tag{40}$$

Thus, $p^{l_g,l_f}=(m_x^{l_g}+m_x^{l_f}-\frac{M_{G_x}+2}{2})$ in case 2.

Case 3: If $(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}})\in(\frac{1}{2}-\frac{1}{M_{G_x}},1-\frac{2}{M_{G_x}}]$, i.e., $(m_x^{l_g}+m_y^{l_f})\in(\frac{3M_{G_x}}{2}+1,2M_{G_x}]$, then

$$\begin{aligned}\mathbf{b}_{M_x}(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}}) &= \frac{1}{\sqrt{M_x}}e^{j2\pi\mathbf{m}_x(-1+\frac{m_x^{l_g}+m_x^{l_f}-2}{M_{G_x}})}\\ &\overset{(a)}{=}\frac{1}{\sqrt{M_x}}e^{j2\pi\mathbf{m}_x(-\frac{1}{2}+\frac{m_x^{l_g}+m_x^{l_f}-\frac{3M_{G_x}+2}{2}-1}{M_{G_x}})}\end{aligned} \tag{41}$$

where (a) can be obtained since $e^{j2\pi\mathbf{m}_xx}=e^{j2\pi\mathbf{m}_x(x-1)}$. In this case, $p^{l_g,l_f}=(m_x^{l_g}+m_x^{l_f}-\frac{3M_{G_x}+2}{2})$.

With the discussion above, (11a) can be obtained. Similaly, (11b) is attainable. ■

APPENDIX C
THE PROOF OF PROPOSITION 3

With an ideal VAD representation, the cascaded channel $\mathbf{H}$ can be decomposed as

$$\mathbf{H}=\mathbf{U}_N'\mathbf{\Pi}\mathbf{U}_M', \tag{42}$$

where $\mathbf{U}_N'\in\mathbb{C}^{N\times L_G}$, $\mathbf{U}_M'\in\mathbb{C}^{M\times L_f}$ are the corresponding steering matrices. $\mathbf{\Pi}\in\mathbb{C}^{L_G\times L_f}$ is the channel gain matrix. The received signal (7) can be rewritten as

$$\mathbf{y}=\mathtt{vec}(\mathbf{Y})=(\mathbf{W}^{\mathsf{T}}\mathbf{U}_M'\otimes\mathbf{U}_N')\mathtt{vec}(\mathbf{\Pi})+\mathtt{vec}(\mathbf{N}). \tag{43}$$

In OMP, the oracle estimator of $\mathbf{H}$ aims to obtain $\mathbf{\Pi}$ by assuming perfect knowledge of the steering vectors. It then proceeds to estimate $\mathtt{vec}(\mathbf{\Pi})$ via the LS method: $\mathtt{vec}(\hat{\mathbf{\Pi}})=(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{H}}\mathbf{y}$, where $\mathbf{A}\triangleq(\mathbf{W}^{\mathsf{T}}\mathbf{U}_M'\otimes\mathbf{U}_N')$. Define $\mathbf{X}\triangleq\mathbf{A}^{\mathsf{H}}\mathbf{A}$, which satisfies

$$\mathtt{rank}(\mathbf{X}^{-1})=\mathtt{rank}((\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1})=\mathtt{rank}(\mathbf{A})=L_GL_f. \tag{44}$$

The mean square error (MSE) of the estimated $\hat{\mathbf{\Pi}}$ can be calculated as

$$\begin{aligned}\mathbb{E}\{\|\mathbf{e}\|_F^2\} &= \mathbb{E}\{\|\mathtt{vec}(\hat{\mathbf{\Pi}})-\mathtt{vec}(\mathbf{\Pi})\|_F^2\}\\ &= \mathbb{E}\{\mathbf{X}^{-1}\mathbf{A}^{\mathsf{H}}(\mathbf{A}\mathtt{vec}(\mathbf{\Pi})+\mathbf{n})-\mathtt{vec}(\mathbf{\Pi})\|_F^2\|\\ &= \mathbb{E}\{\|(\mathbf{X}^{-1}\mathbf{A}^{\mathsf{H}}\mathbf{n})^H\mathbf{X}^{-1}\mathbf{A}^{\mathsf{H}}\mathbf{n}\|_F^2\}\\ &= \sigma_n^2\mathtt{Tr}(\mathbf{X}^{-1}),\end{aligned} \tag{45}$$

where σ_n^2 is the variance of the noise. Furthermore,

$$\begin{aligned}\mathtt{Tr}(\mathbf{X}^{-1}) &= \sum_{l=1}^{L_G L_f} \lambda_l\left(\mathbf{X}^{-1}\right) = L_G L_f \frac{1}{L_G L_f} \sum_{l=1}^{L_G L_f} \lambda_l\left(\mathbf{X}^{-1}\right) \\ &\geq \frac{L_G^2 L_f^2}{\sum_{l=1}^{L_G L_f} \frac{1}{\lambda_l(\mathbf{X}^{-1})}} = \frac{L_G^2 L_f^2}{\mathtt{Tr}(\mathbf{X})},\end{aligned} \tag{46}$$

where $\lambda_l(\mathbf{X}^{-1})$ denotes the l-th eigenvalue of $\mathbf{X}^{-1}$. With the obtained $\mathtt{vec}(\hat{\mathbf{\Pi}})$, oracle estimate of $\mathbf{H}$ can be composed. Exploiting the property of Kronecker product, we have

$$\mathtt{vec}(\hat{\mathbf{H}}) = \underbrace{(\mathbf{U}'_M \otimes \mathbf{U}'_N)}_{\mathbf{D}} \mathtt{vec}(\hat{\mathbf{\Pi}}). \tag{47}$$

Consequently, MSE of $\mathtt{vec}(\hat{\mathbf{H}})$ can be given by

$$\begin{aligned}&\mathbb{E}\{\|\mathtt{vec}(\hat{\mathbf{H}}) - \mathtt{vec}(\mathbf{H})\|_F^2\} = \mathbb{E}\{\|\mathbf{De}\|_F^2\} \\ &\quad = \mathbb{E}\{\|\mathbf{e}^{\mathsf{H}}\mathbf{D}^{\mathsf{H}}\mathbf{De}\|_F^2\} \\ &\quad \geq \lambda_{min}(\mathbf{D}^{\mathsf{H}}\mathbf{D})\mathbb{E}\{\|\mathbf{e}\|_F^2\} = \frac{\lambda_{min}(\mathbf{D}^{\mathsf{H}}\mathbf{D})\sigma_n^2 L_G^2 L_f^2}{\mathtt{Tr}(\mathbf{X})},\end{aligned} \tag{48}$$

where the inequality holds due to the positive definiteness of $\mathbf{D}^{\mathsf{H}}\mathbf{D}$. Similar derivation can also be found in [31]. ■

REFERENCES

[1] M. Jian *et al.*, "Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques," *Intell. Converg. Netw.*, vol. 3, no. 1, pp. 1–32, Mar. 2022.

[2] J. Xu *et al.*, "Reconfiguring wireless environments via intelligent surfaces for 6G: reflection, modulation, and security," *Sci. China Inf. Sci.*, vol. 66, no. 3, Mar. 2023, Art. No. 130304.

[3] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkashlan, "Active RIS versus passive RIS: Which is superior with the same power budget?" *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1150–1154, May 2022.

[4] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44 304–44 321, 2021.

[5] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan, and M. Li, "Channel estimation with reconfigurable intelligent surfacesA general framework," *Proc. IEEE*, vol. 110, no. 9, pp. 1312–1338, Sep. 2022.

[6] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1035–1071, 2nd Quart. 2022.

[7] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 4659–4663.

[8] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 5000–5004.

[9] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.

[10] G. T. de Araújo, A. L. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 3, pp. 789–802, Apr. 2021.

[11] T. Lin, X. Yu, Y. Zhu, and R. Schober, "Channel estimation for IRS-assisted millimeter-wave MIMO systems: Sparsity-inspired approaches," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4078–4092, Jun. 2022.

[12] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 905–909, May 2020.

[13] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.

[14] P. Schniter, L. C. Potter, and J. Ziniel, "Fast bayesian matching pursuit," in *Proc. Inf. Theory Appl. Workshop*, San Deigo, CA, USA, Jan./Feb. 2008, pp. 326–333.

[15] 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, May 2021.

[16] X. Shi, J. Wang, and J. Song, "Triple-structured compressive sensing-based channel estimation for RIS-aided MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 11 095–11 109, Dec. 2022.

[17] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 1478–1492, Mar. 2022.

[18] Z. Peng *et al.*, "Channel estimation for RIS-aided multi-user mmwave systems with uniform planar arrays," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8105–8122, Dec. 2022.

[19] Z. Peng *et al.*, "Two-stage channel estimation for RIS-aided multi-user mmWave systems with reduced error propagation and pilot overhead," *IEEE Trans. Signal Process.*, vol. 71, pp. 3607–3622, Sep. 2023.

[20] Y. You, F. Chen, L. Zhang, Y. Huang, X. You, and C. Zhang, "Improved channel estimation for IRS-assisted mmwave systems exploiting beam leakage," *IEEE Wireless Commun. Lett.*, vol. 12, no. 5, pp. 769–773, Feb. 2023.

[21] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736–7747, Nov. 2021.

[22] S. Kim, J. Wu, and B. Shim, "Efficient channel probing and phase shift control for mmWave reconfigurable intelligent surface-aided communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 231–246, Jan. 2024.

[23] V. Jamali, G. C. Alexandropoulos, R. Schober, and H. V. Poor, "Low-to-zero-overhead IRS reconfiguration: Decoupling illumination and channel estimation," *IEEE Commun. Lett.*, vol. 26, no. 4, pp. 932–936, Apr. 2022.

[24] X. Hu, R. Zhang, and C. Zhong, "Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1132–1142, Feb. 2022.

[25] X. Guan, Q. Wu, and R. Zhang, "Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 3764–3778, Nov. 2021.

[26] R. Faqiri, C. Saigre-Tardif, G. C. Alexandropoulos, N. Shlezinger, M. F. Imani, and P. Del Hougne, "PhysFad: Physics-based end-to-end channel modeling of RIS-parametrized environments with adjustable fading," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 580–595, Jan. 2023.

[27] A. Rabault *et al.*, "On the tacit linearity assumption in common cascaded models of RIS-parametrized wireless channels," *IEEE Trans. Wireless Commun.*, early access, Feb. 2024, doi: 10.1109/TWC.2024.3367953.

[28] Y. You, L. Zhang, M. Yang, Y. Huang, X. You, and C. Zhang, "Structured OMP for IRS-assisted mmWave channel estimation by exploiting angular spread," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4444–4448, Apr. 2022.

[29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[30] 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, Apr. 2021.

[31] 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.

[32] Y. You, Y. Xue, L. Zhang, X. You, and C. Zhang, "Channel estimation for RIS assisted millimeter wave systems via OMP with optimization," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16 783–16 787, Jul. 2023.

[33] 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 Glob. Commun. Conf.*, Austin, TX, USA, 2014, pp. 3326–3331.