Active 3D Double-RIS-Aided Multi-User Communications: Two-Timescale-Based Separate Channel Estimation via Bayesian Learning

Songjie Yang®, Wanting Lyu®, Yue Xiu®, Zhongpei Zhang®, Member, IEEE, and Chau Yuen®, Fellow, IEEE

Abstract—Double-reconfigurable intelligent surface (RIS) is a promising technique, achieving a substantial gain improvement compared to single-RIS techniques. However, in double-RIS-aided systems, accurate channel estimation is more challenging than in single-RIS-aided systems. This work solves the problem of double-RIS-based channel estimation based on active RIS architectures with only one radio frequency (RF) chain. Since the slow time-varying channels, i.e., the BS-RIS 1, BS-RIS 2, and RIS 1-RIS 2 channels, can be obtained with active RIS architectures, a novel multi-user two-timescale channel estimation protocol is proposed to minimize the pilot overhead. First, we propose an uplink training scheme for slow time-varying channel estimation, which can effectively address the double-reflection channel estimation problem. With channels’ sparsity, a low-complexity Singular Value Decomposition Multiple Measurement Vector-Based Compressive Sensing (SVD-MMV-CS) framework with the line-of-sight (LoS)-aided off-grid MMV expectation maximization-based generalized approximate message passing (M-EM-GAMP) algorithm is proposed for channel parameter recovery. For fast time-varying channel estimation, based on the estimated large-timescale channels, a measurements-augmentation-estimate (MAE) framework is developed to decrease the pilot overhead. Additionally, a comprehensive analysis of pilot overhead and computing complexity is conducted. Finally, the simulation results demonstrate the effectiveness of our proposed multi-user two-timescale estimation strategy and the low-complexity Bayesian CS framework.

Index Terms—Active double-reconfigurable intelligent surface, multi-user two-timescale channel estimation, measurements-augmentation-estimate, low-complexity Bayesian compressive sensing.

I. INTRODUCTION

RECONFIGURABLE intelligent surface (RIS) has been recognized as a potential future technology for sixth-generation (6G) communications [1], [2], [3]. It can overcome obstacles, improve channel capacity, and lower transmit power for wireless communications. With the fast advancement of the RIS technique in wireless communications, novel RIS structures have evolved, such as semi-passive/active RIS [4], [5], [6], [7], [8], [9], double-/multi-RIS [10], [11], [12], [13], simultaneous transmitting and reflecting (STAR) RIS [14], and holographic multiple-input multiple-output (MIMO) surfaces [15], [16]. To achieve the beamforming gain provided by RIS, accurate channel estimation is very essential.

As of now, a lot of researches have documented single-RIS-based channel estimation, with a focus on cascaded and separate channels. For cascaded channel estimation, the fully-passive RIS is considered. One straightforward approach is the ON/OFF switching method [17], for successive channel estimation of each RIS element. Following this, the full-ON methods, based on the discrete Fourier transform (DFT) and Hadamard matrices, were proposed for better estimation [18]. Moreover, by exploiting the spatial sparsity of high-frequency channels, the pilot overhead can be reduced using compressive sensing (CS) methods. In [19], the authors transformed the cascaded channel estimation into a sparse recovery problem based on the Katri-Rao and/or Kronecker products, and used the orthogonal matching pursuit (OMP) algorithm [20] to solve it. Furthermore, the cascaded channel’s structural sparsity, i.e., row-column-block sparsity, was discussed in [21], for OMP-based multi-user channel estimation. Along with this, more complex CS methods [22], [23], [24], e.g. message passing and atomic norm minimization, were used for RIS-based channel estimation. Another popular method is the matrix factorization/decomposition formulating cascaded channel estimation as a bilinear estimation problem [25], [26], [27]. In this way, the cascaded channel estimation problem is decoupled into two separate channel estimation problems. Notably, the scaling ambiguity problem cannot escape. That is, the matrix factorization/decomposition method is hard to acquire accurate separate channels. Moreover, deep learning methods have been applied for RIS-based channel estimation, such as deep neural network (DNN)- and convolutional neural network (CNN)-based cascaded channel estimation [28], [29], [30], [31].
On the other hand, the separate channel estimation issue must be handled to fulfill the requirements of situations requiring separate channels, such as [32] and [33]. To this end, a few publications have investigated separate channel estimation with the semi-passive/active RIS architecture [4], [5], [6], [7], [8]. In [4], a hybrid RIS with numerous active elements attached to radio frequency (RF) chains was designed to ease the channel estimation process. On this premise, the authors of [5] evaluated the estimation performance with varying numbers of RF chains, each connected to a single active element, using the CS approach. Due to the many RF chains, this design automatically results in a significant cost. In order to decrease hardware costs, [6], [7] used an architecture that connected a single RF chain to numerous active elements. Notably, all these strategies employed a two-way method. Specifically, both the base station (BS) and the user transmit pilots to the RIS in order to estimate the channels. The RIS then transmits the channel information to the BS over the wired control. This technique focuses on the active RIS’s capacity to receive signals, but ignores the fact that it can also reflect signals, resulting in additional pilot overhead. To solve this, the authors of [8] and [9] presented a one-way training strategy that reduced training overhead by taking into account the RIS’s capacity to reflect.

Recently, the double-RIS design was widely studied in [10], [11], and [12], with one RIS set near the BS and the other RIS set near the user, for capacity enhancement. This architecture could achieve an $L^4$-fold cooperative power gain (a substantial improvement compared with the single-RIS aided system’s $L^2$-fold gain) by the double-reflection link [10], with $L$ representing the number of RIS elements. Notably, in such a double-RIS-aided system, channel estimation will become more problematic due to two single-reflection channels and one double-reflection channel. In [10], two semi-passive RISs were assumed to receive signals for estimating the channels between them and the BS/user. To deal with the double-reflection channel, the assumption is made that the inter-RIS channel consisted only of the known LoS path. This may be impractical in outdoor communications. To tackle the double-reflection channel, the authors in [34], [35], [36], [37], and [38] have presented effective cascaded channel estimation methods for fully-passive double-RIS systems. Nevertheless, these methods will result in a high amount of training complexity, and obtaining separate channels will be challenging. Earlier-mentioned semi-passive/active RIS can make up for this by incurring low hardware expenses. In this sense, the constraint of [10] with semi-passive RISs should be overcome. Additionally, as in the case of single-RIS-based channel estimation, the sparsity of the high-frequency channel could be utilized to reduce the pilot overhead.

Last but not least, the assumption that both the BS and the RIS are equipped with uniform planar arrays (UPAs) is more realistic than the case of uniform linear arrays considered in the majority of CS-based channel estimation [5], [8], [19], [21], [22], [23]. This UPA-based system setup [6], [7], however, will make it more challenging to estimate the channel’s parameters due to the three-dimensional (3D) beampattern, especially in double-RIS-aided systems. In this case, the conventional CS framework, i.e. Kronecker CS, is relatively time-consuming when used to the multi-parameter recovery issue due to the immense dictionary generated by the vectorization of multidimensional signals. Such a recovery framework, with a large computational cost, is impracticable for UPA-based systems when recovering the 3D beamprobe (the BS’s and RIS’s angles-of-departure (AoDs)/angles-of-arrival (AoAs) in the elevation and azimuth directions). Therefore, it is essential to investigate estimation methods with minimal complexity for 3D double-RIS-aided systems.

Taking into account the above, an efficient separate channel estimation strategy for active 3D double-RIS-aided systems is investigated in this paper, in which the BS and the RISs are all equipped with UPAs. Compared with [34], [35], [36], [37], and [38], this work can obtain separate channels owing to the RF chain set at the RIS. Besides, the channel’s sparsity is exploited to effectively reconstruct the channel. Based on the separate channels, the two-timescale estimation technique, which has been employed in relay- and single-RIS-aided systems [39], [40], is used for pilot minimization. In comparison to the most related work [10] with semi-passive RISs, the double-reflection channel estimation problem is addressed.

Before presenting our work, we first introduce three significant channel properties of double-RIS-aided systems.

- **Two-Timescale Channel Property:** Owing to the fixed locations of the BS and RISs, the channels among them are slow time-varying. On the contrary, the channel between the RIS and the mobile user is fast time-varying.
- **LoS-Path Property:** Owing to the known locations of the BS and RISs, the LoS paths’ AoA/AoD of channels among them can be easily obtained.
- **Multi-User Channel Property:** For each RIS, all users share the same BS-RIS channel. This indicates that the BS’s AoD and the RIS’s AoA are common for all users’ channels.

According to the above properties, the following contributions are exhibited with respect to hardware cost, pilot overhead and computational complexity:

- **Cost-Minimized Active RIS Structure:** Considering the hardware cost of the active RIS, we use only one RF chain to receive signals. Different from the active RIS structure with each active element connected to a RF chain [4], [5], [8], the single-RF-chain-based RIS makes each RF chain connects to all elements. Since the proposed channel estimation method makes no power requirements of the RIS, all active elements can be set to unit power, which is similar to the fully-passive RIS, without incurring additional power consumption.
- **Pilot-Minimized Channel Estimation Protocol:** Considering that the single-RF-chain is bound to incur higher training overhead for parameter recovery compared with multiple RF chains, a novel multi-user two-timescale estimation technique is proposed to minimize the pilot overhead based on the CS theory\(^1\) and Two-Timescale

\(^1\)When using the CS method for high-frequency channels, the channel estimation problem can be converted into a beamspace recovery problem.
Channel Property. In this sense, we first estimate the slow time-varying channels in an uplink training manner, which can effectively solve the double-reflection channel estimation problem in [10]. Leveraging on Multi-User Channel Property, a measurements-augmentation-estimate (MAE) CS strategy is proposed for better estimation. Based on the estimated slow time-varying channels, the fast time-varying channels can be estimated at the BS with more RF chains, so as to develop a MAE-based solution with low pilot overhead.

- Computational Complexity-Minimized CS Framework: To avoid the unacceptable computational complexity induced by recovering the 3D beamspace, we propose a novel CS framework for multi-parameter recovery. Specifically, inspired by the decoupling of two parameter recovery problems using SVD in [41], we exploit the common sparsity of decomposition vectors to develop a more effective singular value decomposition multiple measurement vector-based compressive sensing (SVD-MMV-CS) Framework. Most importantly, it is much faster and has better performance than the common Kronecker CS framework, as shown in simulation results. Besides, the LoS-aided dictionary based on LoS-Path Property is designed, which can resist the basis mismatch problem to some extent.

- Efficient Bayesian Recovery Algorithm: With our proposed CS framework, the selection of the recovery algorithm is very crucial. To achieve an excellent recovery performance, the Bayesian learning-based method is considered for the 3D beamspace recovery. Given the great computing cost produced by the inverse operation of the standard Bayesian recovery method, i.e., EM-based sparse Bayesian learning, the expectation maximization-based generalized approximate message passing (EM-GAMP) [43] algorithm is used, which utilizes the GAMP procedure to approximate the true posterior distribution with low complexity. To employ the EM-GAMP algorithm for our proposed CS framework, we extend it to the MMV version (M-EM-GAMP) by sharing a sparse rate factor across all measurement vectors. Considering that M-EM-GAMP cannot handle the mismatch problem, we propose a fast off-grid MMV method, based on the Taylor expansion, to refine the on-grid recovered parameters.

The rest of this paper is organized as follows: in Section II, we model the two-timescale channels and describe the received signals at the BS and RISs via uplink training. Section III shows the two-timescale channel estimation protocol, and design the CS recovery problems for the large-timescale and small-timescale estimation, respectively. Section IV proposes a LoS-aided dictionary, a low-complexity CS framework, an M-GAMP algorithm, an EM procedure for updating the hyperparameters of M-GAMP, and the Taylor expansion-based off-grid refinement method. Section V evaluates the large-timescale and small-timescale channel estimation methods via various experiments. Additionally, the analysis of pilot overhead and computational complexity of different schemes is provided. Finally, Section VI concludes the paper.

Notations: $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote conjugate, transpose, and conjugate transpose, respectively. $A^\dagger$ is the Moore-Penrose pseudo-inverse matrix. $\| \cdot \|$ and $\| \cdot \|_2$ represent $\ell_0$ norm and $\ell_2$ norm, respectively. $\|A\|_F$ denotes the Frobenius norm of matrix $A$. Furthermore, $\otimes$ and $\circ$ are the Kronecker product and the Hadamard product, respectively. $[a]_i$ and $[A]_{i,j}$ denote the $i$-th element of vector $a$, the $(i, j)$-th element of matrix $A$, respectively. vec represents the vectorization operation. $\text{Re}\{\alpha\}$ denotes the real part of complex $a$. $E\{\cdot\}$ and $\mathbb{V}\{\cdot\}$ denote the expectation and variance operations, respectively. $\mathbf{I}_M$ denotes the $M$-by-$M$ identity matrix. Moreover, diag($a$) is a square diagonal matrix with entries of $a$ on its diagonal. Finally, $\mathcal{CN}(a, A)$ is the complex Gaussian distribution with mean $a$ and covariance matrix $A$.

II. SYSTEM AND CHANNEL MODEL

We consider a multi-user double-RIS-aided system with central frequency $28$ GHz, shown in Fig. 1, where the BS with the assistance of the RIS serves $U$ users. The BS is equipped with $J$ antenna elements, and each of RISs is equipped with $L$ reflective elements. The users are all equipped with a single antenna. As shown in Fig. 1, we denote the channels from the BS to RIS $i$ as $F_i$, from RIS 1 to RIS 2 as $D$, and from the RIS $i$ to the $u$-th user as $h_{i,u} \in \mathbb{C}^{L \times 1}$, where $i = 1, 2$ and $u = 1, \cdots, U$. Since there is no requirement that the power of each element is controlled separately for channel estimation, the situation shown in Fig. 2(a) is unnecessary. Consequently, we utilize a unique low-cost active RIS structure with one RF chain and one power amplifier (PA), as shown in Fig. 2(b). This is a

---

2In a CS framework, it may be recognized that the observed values rise for some underlying reasons.
specific instance of the active RIS with the sub-connected architecture introduced in [44]. Next, Section II-A models the large-timescale and small-timescale channels, and Section II-B gives the received signal model via uplink training.

### A. Two-Timescale Channel Model

Quasi-static block-fading channels are considered, we adopt the widely used spatial channel model, characterizing the geometrical structure. Assuming that the BS and RISs are all equipped with UPAs, then the slow and fast time-varying channels are modeled as follows.

1) **Slow Time-Varying Channel**: The slow time-varying channels between the BS and RIS 1, the BS and RIS 2, and RIS 1 and RIS 2, i.e., \( F_1, F_2 \) and \( D \), are respectively expressed as

\[
F_i = \sum_{p_i=1}^{P_i} \alpha_{p,i} \mathbf{a}_i \left( \frac{2d}{\lambda} \cos(\varphi_{p,i}^e), \frac{2d}{\lambda} \sin(\varphi_{p,i}^e) \right) \\
D = \sum_{b=1}^{P_d} \beta_b \mathbf{a}_{L_d} \left( \frac{2d}{\lambda} \cos(\phi_b^e), \frac{2d}{\lambda} \sin(\phi_b^e) \right),
\]

where \( i \in \{1,2\} \) denote the indices of the two RISs, \( P_i \), and \( P_d \) are the number of paths of \( F_i \) and \( D \), respectively. \( \{\varphi_{p,i}^e\} \) and \( \{\phi_b^e\} \) are AoDs and AoAs of the channel between the BS and RIS 1, respectively. Notably, the superscript \( e \) and \( a \) denote the elevation and azimuth directions, respectively. \( \{\varphi_{p,i}^e\} \) and \( \{\phi_b^e\} \) are AoDs and AoAs of the channel between RIS 2 and RIS 1, respectively. \( P_{i,a} \) are the path gains of the above channels. In addition, \( \lambda \) is the antenna inter-element spacing, generally \( d = \lambda/2 \). Ignoring the subscript, the UPA steering vector \( \mathbf{a}(x_1, x_2) \) is given by

\[
\mathbf{a}(x_1, x_2) = \frac{1}{N_y N_z} \begin{bmatrix} e^{j\pi x_1}, \ldots, e^{j\pi(N_y-1)x_1} \\
1, e^{j\pi x_2}, \ldots, e^{j\pi(N_z-1)x_2} \end{bmatrix}^T,
\]

where \( N_y \) and \( N_z \) are the number of antennas on the \( y \) and \( z \) axes, respectively.

2) **Fast Time-Varying Channel**: Due to the mobility of users, the channels with respect to (w.r.t.) users are thought of fast time-varying. The channels between the BS and users, and between RISs and users are \( h_{i,u} = \sum_{c=1}^{C_i} \tilde{\gamma}_{i,u,c} \mathbf{a}_L \left( \frac{2d}{\lambda} \cos(\chi_{i,u,c}), \frac{2d}{\lambda} \sin(\chi_{i,u,c}) \sin(\chi_{i,u,c}) \right) \)

where \( \tilde{\gamma}_{i,u,c} = \sqrt{L_{i,u}} \gamma_{i,u,c} P_{h,i,u} \) is the number of paths of \( h_{i,u} \), \( \{\tilde{\gamma}_{i,u,c}\}_{c=1}^{C_i} \) denote path gains and AoDs from RIS \( i \) to user \( u \), respectively.

3) **Virtual Channel Representation**: Using the virtual channel representation, the BS-RIS, RIS 1-2 and RISs-user channels can be rewritten by sampling the 3D beamspace, i.e.,

\[
F_i \approx A_B \Xi_i, A_{L_1}^H, D \approx A_{L_2} \Delta A_{L_1}^H, h_{i,u} \approx A_{L_i} \zeta_{i,u},
\]

where \( A_B \in \mathbb{C}^{J \times G_I} \) and \( A_{L_i} \in \mathbb{C}^{L \times G_R} \) denote the dictionaries of the BS’s AoD, RISs’ AoA/AoD, respectively. \( \Xi_i \in \mathbb{C}^{G_i \times G_r} \) denotes the sparse matrix, where the non-zero entries correspond to the BS’s AoDs and the \( i \)-th RIS’s AoAs. In addition, each AoD/AoA consists of the elevation and azimuth angles. This resembles both \( D \) and \( h_{i,u} \). The later section will provide specific dictionaries appropriate for RIS systems.

### B. Uplink Training

We arrange the uplink training into two parts based on the estimates of \( \{F_i,D\}_{i=1}^2 \) and \( \{h_{i,u}\}_{i=1}^2 \). For RIS-based multi-user channel estimation, the key is to design orthogonal pilot sequences and reflection coefficients. Assume that there are \( Q \) sub-frames for uplink training and \( T \) symbol durations for users in each sub-frame, where \( T \geq U \). Owing to orthogonal pilot sequences, we have \( s_{u_1}^H s_{u_2} = 0 \), \( u_1 \neq u_2 \) and \( s_{u_1}^H s_{u_1} = \sigma_u^2 T, \forall i \), where \( \sigma_u^2 \) is the transmit power. Denoting the reflection coefficients of RIS 1 and RIS 2 in the \( q \)-th sub-frame and the pilot sequence of the \( u \)-th user by \( v_{1,q} \in \mathbb{C}^{Q \times 1} \), \( v_{2,q} \in \mathbb{C}^{Q \times 1} \) and \( h_{u}^H \in \mathbb{C}^{1 \times T} \), respectively.

In the \( q \)-th sub-frame, the received signals at RIS 1, RIS 2 and the BS are denoted by \( y_{RIS_1}^q \), \( y_{RIS_2}^q \) and \( y_{BS}^q \). They are given by

\[
y_{RIS_1} = \sum_{u=1}^{U} v_{1,q}^H h_{1,u} s_{u}^H + v_{1,q}^H N_{RIS_1}^q,
\]

\[
y_{RIS_2} = \sum_{u=1}^{U} v_{2,q}^H D_{1,q} h_{2,u} s_{u}^H + v_{2,q}^H N_{RIS_2}^q,
\]

\[
y_{BS} = \sum_{u=1}^{U} (F_1 V_{1,q} h_{1,u} + F_2 V_{2,q} h_{2,u} + F_2 V_{2,q} D_{1,q} h_{1,u}) s_{u}^H + N_q^B,
\]

where \( V_{1,q} = \text{diag}(v_{1,q}) \in \mathbb{C}^{Q \times 1} \), \( v_{2,q} \in \mathbb{C}^{Q \times 1} \), \( \sigma_q^{RIS_1} \in \mathbb{C}^{L_x \times L} \), \( N_{RIS_1}^{q} \in \mathbb{C}^{L_x \times L} \), \( N_{RIS_2}^{q} \in \mathbb{C}^{L_x \times L} \) and \( N_{BS}^{q} \in \mathbb{C}^{L_x \times L} \) are the Gaussian white noise.

In contrast to the received signal from the RIS, the signal delivered to the BS is comprised of the RIS’s reflected signal and the direct signal. Since the direct channel between the BS and the user can be estimated simply by turning off all RISs, the direct signal can be removed from the BS’s received signal. As a result, this study need not consider the direct channel. This, too, is built on the same considerations as cascaded channel estimation literature.

### III. Two-Timescale-Based Separate Channel Estimation Framework

To efficiently accomplish channel estimation for double-RIS systems, this section proposes a multi-user channel estimation framework with two timescales.

For large-timescale estimation, a three-phase scheme is proposed to estimate \( F_1, F_2 \) and \( D \) separately. Different from
the approach in [4], [5], [6], and [7] that the BS transmits the pilot to estimate the BS-RIS channel at the RIS, our proposed method exploits uplink training sequences sent from users to estimate \( F_1, F_2 \) and \( D \) at the BS. There are two benefits of this approach. First, the measurement dimension of the signal received by the BS is greater than that received by the RIS due to the number of the BS’s RF chains are much larger than that of RISS’s. The second is joint multi-user channel estimation. As shown in Fig. 1, all users share the common channels \((F_1, F_2, D)\), hence the signals received by the BS from all users may be thought of as observations in the same space. This finding suggests that signals from several users can be handled in concert to improve recovery performance. In this way, a MAE strategy based on Multi-User Channel Property is proposed. Specifically, since the channel \( F_i \) is common for all users, the signals of different users received by the BS can be utilized to jointly estimate their shared channel \( F_i \). For small-timescale estimation, we also present a MAE strategy. Based on the channels estimated in the large-timescale, \( \{\{h_{1,u}\}^2\}_{i=1}^{U} \) can be estimated at the BS with numerous RF chains as opposed to the RIS, in order to improve recovery performance.

We assume that there are \( Q_1, Q_2, Q_3, Q_1, Q_2 \) sub-frames for the estimation of \( F_1, F_2, D \), \( \{h_{1,u}\}^U_{i=1} \) and \( \{h_{2,u}\}^U_{i=1} \), respectively. Similar to the previous description, each sub-frame consists of \( T (T \geq U) \) symbol durations for orthogonal pilot transmission. Following that, the proposed multi-user two-timescale channel estimation protocol, accommodating large-timescale channel estimation, small-timescale channel estimation, data transmission, and uplink orthogonal pilot sequence training, is shown in Fig. 3. It can be noticed that in a large-timescale period, the slow time-varying channels just need to be estimated only once. Therefore, the whole pilot overhead is reduced, compared with the traditional estimation protocol that needs to estimate the slow and fast time-varying channels with the same number of times. Next, we will introduce the large-scale and small-scale estimation methods in detail.

A. Large-Timescale Estimation

1) Estimation of \( F_i \): For simplicity, here we use the time-division mode, in which one RIS is active, while another is inactive. Utilizing the receiving/processing signal capacity of the RIS to acquire the RIS-user channel and subsequently the reflection capability of the RIS to estimate \( F_i \) at the BS using the estimated RIS-user channel is essential to our suggested MAE strategy.

First, the \( i \)-th RIS’s received signal \( y_{RIS}^{i} \in \mathbb{C}^1 \times T \) in the \( q_i \)-th sub-frame, with \( q_i \in \{1, \ldots, Q_i\} \), is given by:

\[
y_{RIS}^{i} = \sum_{u=1}^{U} v_{i,u}^H h_{i,u} s_u^H + v_{i,q_i}^H N_{RIS_i}^u,
\]

where \( v_{i,q_i} \) is the \( i \)-th RIS’s reflect coefficients in the \( q_i \)-th sub-frame, and \( N_{RIS_i}^u \in \mathbb{C}^{1 \times T} \) is the Gaussian noise vector following \( \mathcal{CN}(0, \sigma_n^2 I_T) \). The collected received signal \( \tilde{y}_{RIS}^{i} \in \mathbb{C}^{Q_i \times T} \) in \( Q_i \) sub-frames at RIS \( i \) can be expressed as

\[
\tilde{y}_{RIS}^{i} = \sum_{u=1}^{U} \tilde{V}_i^H h_{i,u} s_u^H + \tilde{N}_{RIS}^{i},
\]

where \( \tilde{V}_i = [v_{i,1}, \ldots, v_{i,Q_i}] \in \mathbb{C}^{L \times Q_i} \) and \( \tilde{N}_{RIS}^{i} = [v_{i,1}^H N_{RIS_i,1}, \ldots, v_{i,Q_i}^H N_{RIS_i,Q_i}]^T \in \mathbb{C}^{Q_i \times T} \). Owing to the orthogonal pilot sequences, we obtain the \( u \)-th user’s received signal \( \tilde{y}_{RIS}^{u} \in \mathbb{C}^{Q_i \times 1} \), i.e.,

\[
\tilde{y}_{RIS}^{u} = \frac{1}{\sigma_n^2 T} \tilde{y}_{RIS}^{i} s_u,
\]

\[
\approx \tilde{V}_i^H A_{L_i} \tilde{c}_{i,u} + \tilde{n}_{RIS}^{i}.
\]

where \( (a) \) holds because of (4), \( \tilde{n}_{RIS}^{i} = \frac{1}{\sigma_n^2 T} \tilde{N}_{RIS}^{i} s_u \in \mathbb{C}^{Q_i \times 1} \) and \( \tilde{V}_i^H A_{L_i} \in \mathbb{C}^{Q_i \times G_i} \) is the sensing matrix. \( \{h_{1,u}\}^U_{i=1} \) can be easily estimated by a variety of recovery algorithms. Then they are sent to the BS as the auxiliary information for estimating \( F_i \) through the wired control link.\(^3\)

Similar to the previous process, the BS’s received signal reflected by RIS \( i \) in the \( q_i \)-th sub-frame is given by

\[
\tilde{y}_{RIS}^{B_S} = F_{i} V_{i,q_i} h_{i,u} + \tilde{n}_{RIS}^{B_S}
\]

\[
\approx F_{i} \text{diag}(h_{i,u}) v_{i,q_i} + \tilde{n}_{RIS}^{B_S},
\]

where \( \tilde{n}_{RIS}^{B_S} \) is noise. Hence, the BS’s received signal in \( Q_i \) sub-frames is expressed as

\[
\tilde{y}_{B_S}^{i} = F_{i} \text{diag}(h_{i,u}) \tilde{V}_i + \tilde{N}_{B_S}^{i},
\]

\[
\approx A_{B} \Xi_{i} A_{L_i}^H \text{diag}(h_{i,u}) \tilde{V}_i + \tilde{N}_{B_S}^{i},
\]

\(^3\) Also, there are two additional types of feedback: pilot feedback used to estimate \( \{h_{1,u}\}^U_{i=1} \) at the BS and estimated spatial parameter (angle and gain information) feedback used to rebuild the channel at the BS.
where $\tilde{Y}_i$ is the same as the previous definition, $\tilde{N}_{i,u}^{BS} = [\tilde{n}_{i,u,1}, \ldots, \tilde{n}_{i,u,Q_i}] \in \mathbb{C}^{J \times Q_i}$.

Assuming that $\tilde{h}_{i,u}$ is the estimated channel, we denote $A_l^H \text{diag}(\tilde{h}_{i,u}) V_i$ by $\Phi_i \in \mathbb{C}^{G_r \times Q_i}$. As stated before, the common channel $F$ can be estimated by joint processing of multi-user signals. By collecting all users’ signals, we attain

$$\tilde{Y}_i^{BS} \approx A_B \Xi_i \Phi_i + \tilde{N}_i,$$  \hspace{1cm} (12)

where $\tilde{Y}_i^{BS} = [\tilde{Y}_{i,1}^{BS}, \ldots, \tilde{Y}_{i,U}^{BS}] \in \mathbb{C}^{J \times Q_i}, \Phi_i = [\Phi_i, \ldots, \Phi_i,u] \in \mathbb{C}^{G_r \times Q_i}$ and $\tilde{N}_i = [\tilde{N}_{i,1}, \ldots, \tilde{N}_{i,U}] \in \mathbb{C}^{J \times Q_i}$. According to the Kronecker CS framework, vectorizing $\tilde{Y}_i^{BS}$ yields

$$\tilde{y}_i^{BS} = \text{vec}(\tilde{Y}_i^{BS}) \approx (\Phi_i^T \otimes A_B) \xi_i + \text{vec}(\tilde{N}_i),$$  \hspace{1cm} (13)

where $\xi_i = \text{vec}(\Xi_i) \in \mathbb{C}^{G_r \times 1}$ and $\Phi_i \otimes A_B \in \mathbb{C}^{Q_r \times J \times G_r \times Q_i}$. Since the vectors in the sensing matrix $\Phi_i \otimes A_B$ created by the Kronecker product are excessively lengthy, direct processing of (13) will result in high computing cost regardless of the recovery algorithm used. In the later section, we will offer a fast solution for (12) that avoids the Kronecker product.

2) Estimation of $D$: Here, both RISs are turned on simultaneously to estimate $D$ based on the estimated $\{\hat{F}_1, \hat{F}_2\}$. By removing the received signals related to the two RISs’ cascaded channels, the remaining signal at the BS in the $q_3$-th sub-frame is approximated as $y_{q_3}^{BS} \approx \hat{F}_2 V_{q_3} D V_{q_3} \hat{h}_{i,u} + N_{q_3}^{BS}, \forall u$. According to the proposed MAE strategy, we jointly process all users’ signals to estimate $D$. By collecting all users’ signals in the $q_3$-th sub-frame, we have $\tilde{Y}_{q_3}^{BS} = \hat{F}_2 V_{q_3} D \hat{H}_1 + \tilde{N}_{q_3}^{BS}$, where $\hat{H}_1 = [\hat{h}_{i,1}, \ldots, \hat{h}_{i,U}] \in \mathbb{C}^{L \times U}$. Since $\tilde{N}_{q_3}^{BS} = [\tilde{N}_{q_3}^{BS}, \ldots, \tilde{N}_{q_3}^{BS}] \in \mathbb{C}^{J \times U}$.

Following that, we divide the total sub-frames $Q_3$ into $N_X$ and $N_Y$ ($Q_3 = N_X \times N_Y$) for phase adjustment of RIS 1 and 2, respectively. Then the collected signal in $Q_3$ sub-frames $\tilde{Y}_{q_3}^{BS} \in \mathbb{C}^{N_Y \times N_X \times U}$ is written as

$$\tilde{Y}_{q_3}^{BS} = \begin{bmatrix} \hat{F}_2 V_{1} \hat{H}_1 \quad \cdots \quad \hat{F}_2 V_{N_X} \hat{H}_1 \end{bmatrix} \begin{bmatrix} D \hat{V}_{1,1} \hat{H}_1 \quad \cdots \quad \hat{V}_{1,N_X} \hat{H}_1 \end{bmatrix} + \tilde{N}_{q_3}^{BS},$$  \hspace{1cm} (14)

where $\tilde{N}_{q_3}^{BS} \in \mathbb{C}^{N_Y \times N_X \times U}$ is the noise matrix. Via kronecker CS, the above equation can be converted to

$$\tilde{y}_{q_3}^{BS} = \text{vec}(\tilde{Y}_{q_3}^{BS}) \approx (A_{H}^T \otimes \delta) \xi + \text{vec}(\tilde{N}_{q_3}^{BS}),$$  \hspace{1cm} (15)

where $A_{H} = A_{H}^H \begin{bmatrix} D \hat{V}_{1,1} \hat{H}_1 \quad \cdots \quad \hat{V}_{1,N_X} \hat{H}_1 \end{bmatrix} \in \mathbb{C}^{G_r \times N_X \times U}$, $A_{F} = \begin{bmatrix} A_{F}^T \hat{V}_{1,1} \hat{F}_2 \quad \cdots \quad A_{F}^T \hat{V}_{N_Y} \hat{F}_2 \end{bmatrix}^T \in \mathbb{C}^{N_Y \times G_r}$ and $\delta = \text{vec}((\Delta)) \in \mathbb{C}^{G_r \times 1}$.

Evidently, the kronecker product imposes the same computing burden on (15) as (13). As indicated previously, this requires a computationally efficient framework for estimating several parameters $\theta_e, \theta_z, \phi_e, \phi_z$ and $\beta$ inherent in (2), which will be presented in the later section.

### B. Small-Timescale Estimation

As the channel $F_i$ known, channel estimates of $\{h_{i,u}\}_{u=1}^U$ can be performed at the BS with little training overhead, as contrast to the case of (9).

Recalling (10), the BS’s collected signal reflected by RIS $i$ in $Q_i$ sub-frames can be given by

$$[\tilde{y}_{i,u,1}^{BS}, \ldots, \tilde{y}_{i,u,Q_i}^{BS}] = \begin{bmatrix} \hat{F}_i V_{i,1} \quad \cdots \quad \hat{F}_i V_{i,Q_i} \end{bmatrix} h_{i,u} + \tilde{n}_{i,u}^{BS} \approx \Psi \xi_{i,u} + \tilde{n}_{i,u}^{BS},$$  \hspace{1cm} (16)

where $\Psi = \begin{bmatrix} A_{F}^T V_{1,1}^T \hat{F}_1^T \quad \cdots \quad A_{F}^T V_{Q_i,1}^T \hat{F}_1^T \end{bmatrix} \in \mathbb{C}^{Q_i \times G_r}$ is the sensing matrix. Compared to the sensing matrix used to recover $\xi_{i,u}$ in (9), the number of measurements of $\Psi$ is greater owing to the multiplication by the number of BS antennas. This implies that training costs for estimating $h_{i,u}$ can be significantly reduced if $F_i$ is known, i.e., $Q_i \ll Q_i$.

### IV. LOW-COMPLEXITY BAYESIAN COMPRESSIVE SENSING SOLUTION

This section offers low-complexity Bayesian learning-based solutions to the problems outlined in the preceding section. One is the basic MAE-based CS issue for the RIS-user channel, like (9) and (16), while the other is a more complicated CS issue in the face of more recovery parameters, as (12) and (14). We will focus on the latter problem, since the former problem can be addressed using the off-grid single measurement vector (SMV)-based EM-GAMP method, which is a simplified version of the latter problem’s solution.

#### A. LoS-Aided Dictionary Development

Let’s give the necessity of the LoS-aided dictionary first. Considering a one-dimensional bin, i.e., $\{-1 + \frac{2g}{G} | g = 0, \ldots, G-1\}$, when we sample multiple true angles $\sin(z)$ in this bin, they will be impacted by the mismatch problem due to the limited bin size $G$. Although off-grid techniques can overcome this problem, the algorithmic error is inevitable. If we know one of the true angles $\pi$, we can design a $\pi$-based sampling method to ensure that recovering $\pi$ is not affected by the mismatch problem, i.e.,

$$\sin(x) = \begin{cases} \pi + \frac{2g}{G}, & \pi + \frac{2g}{G} \leq 1, \\ -2 + \frac{2g}{G}, & \pi + \frac{2g}{G} > 1, \end{cases} \quad g = 1, \ldots, G-1.$$

(17)

Recalling the UPA steering vector of (3), the two-dimensional bin should be considered. Given the known LoS path’s angle $\theta_{ele}/\theta_{azi}$, the elevation and azimuth angles can be sampled similar to (17). Thus, a UPA dictionary $A_x$ can be constructed by $A_x = \{a(\sin(x_{ge}), \sin(x_{ga})) | g_{x} = 1, \ldots, G_{x}, g_{y} = 1, \ldots, G_{y}\}$.

(18)
where \( \sin(x_{g_i}) \) and \( \sin(x_{g_j}) \) are determined according to (17) with \( \{G_x, \Phi^{\text{ele}}\} \) and \( \{G_y, \Phi^{\text{azi}}\} \), respectively, and \( G_x \) and \( G_y \) are the dictionary sizes.

According to the locations of the BS, RIS 1 and RIS 2, the LoS path’s angle \( \theta_{\text{ele}}^{\text{azi}} \), \( \theta_{\text{azi}}^{\text{ele}} \), and \( \theta_{\text{azi}}^{\text{azi}} \) are easily obtained. When recovering \( \{\theta_{p,1}^{\text{ele/azi}}, \theta_{p,2}^{\text{ele/azi}}, \theta_{b}^{\text{ele/azi}}\} \), \( p = 1, \ldots, P_f \), \( b = 1, \ldots, P_d \), the corresponding dictionary can be established as above. Thus, the recovery of the LoS path will not be affected by the dictionary size. Along with this, the off-grid procedure is simplified since the on-grid LoS path’s parameters have been optimal. In particular, the dictionary of the BS/RIS may be different when representing different channels. For instance, \( A_{1,2} \) depends on \( \Psi \) for \( F_2 \) but \( \Phi \) for \( D \).

### B. SVD-MMV-CS Framework

(12) and (14) belong to the same class of problems, we shall derive the formula based on (14) without losing generality. As described previously, we have \( Y^{\text{BS}} \approx A_F A_{1,2} A_{H} + N^{\text{BS}} \). Via a rank decomposition of \( \Delta = \sum_{k=1}^{r} d_{1,k} d_{2,k}^H \), with \( \text{rank}(\Delta) = r_\Delta \), we can obtain

\[
Y^{\text{BS}} \approx \sum_{k=1}^{r} (A_F A_{1,2} d_{1,k})(A_{H} d_{2,k})^H + N^{\text{BS}}. \tag{19}
\]

Note that \( \Delta \) is a matrix with \( P_{\Delta} \)-sparsity; hence, \( \text{rank}(Y^{\text{BS}}) \leq \text{rank}(\Delta) = r_\Delta \leq P_d \). We denote \( C(\Delta) \) by the column space of \( \Delta \), which is the vector subspace whose elements are \( P_{\Delta} \)-sparse, as well as \( C(\Delta^H) \). Notably, we assume that the AoA and AoD are one-to-one mapping and the sensing matrices meet the null space property (NSP) [45], then \( \text{dim}(A_F A_{1,2} C(\Delta)) = \text{dim}(A_{H}^H C(\Delta^H)) = \text{rank}(\Delta) = r_\Delta \). Thus, the decomposition of \( Y^{\text{BS}} \) in (19) is a rank- \( r_\Delta \) decomposition due to \( d_{1,k} \in C(\Delta) \) and \( d_{2,k} \in C(\Delta^H) \). According to the above, the following proposition can be developed.

**Theorem 1:** Let two sensing matrices be \( \Phi_1 \in \mathbb{C}^{N_1 \times M_1} \) and \( \Phi_2 \in \mathbb{C}^{N_2 \times M_2} \), respectively. If \( \Delta \in \mathbb{C}^{M_1 \times M_2} \) is \( P_{\Delta} \)-sparse, then \( \Delta \) can be recovered uniquely via rank decomposition, such as SVD. We assume \( Y = (U^H \Sigma)^{1/2} V^H = \sum_{k=1}^{\text{rank}(Y)} \mathbf{e}_i \mathbf{e}_k^H \cdot \mathbf{e}_i \mathbf{e}_k^H \), where \( U, V \) are unitary matrices, and \( \Sigma \) is the singular value matrix. Let \( \delta_{i,k} \in \mathbb{C}^{M_1 \times 1} \), in which \( i \in \{1,2\} \) and \( k \in \{1,2,\ldots,\text{rank}(Y)\} \), be the solution of \( \min \{ \|\delta_{i,k}\|_0 + \|\mathbf{e}_i - \Phi_{i,k} \delta_{i,k}\|_2^2 \} \). Since each \( \delta_{i,k} \) is \( P_{\Delta} \)-sparse, MMV issues can be naturally formulated as follows. Re-writing that \( Y = E_1 E_2^H \) with \( E_1 = [\mathbf{e}_{1,k}, \ldots, \mathbf{e}_{1,\text{rank}(Y)}] \in \mathbb{C}^{N_1 \times \text{rank}(Y)} \) and \( E_2 = [\mathbf{e}_{2,k}, \ldots, \mathbf{e}_{2,\text{rank}(Y)}] \in \mathbb{C}^{N_2 \times \text{rank}(Y)} \). For \( i \in \{1,2\} \), the MMV issue is

\[
\arg \min_{\Delta^M} \|\Delta^M\|_0, \quad \Delta^M \in \mathbb{C}^{M_1 \times M_2}, \quad \text{s.t. } \|E_i - \Phi_i \Delta^M\|_F^2 \leq \epsilon, \quad \text{supp}(t_{i,1}) = \cdots = \text{supp}(t_{i,R}), \tag{21}
\]

where \( \Delta^M \in [t_{i,1}, \ldots, t_{i,R}] \in \mathbb{C}^{N_1 \times R} \) is the matrix with group sparsity, \( R = \text{rank}(Y) \), and \( \text{supp}(\cdot) \) denotes the sparsity support. Finally, the recovered signal can be written as

\[
\hat{\Delta} = \Delta^M (\Delta^2_H)^H. \tag{22}
\]

**Proof:** See Appendix A.

Based on **Theorem 1**, the recovery issue of (14) is solved as follows. First, decomposing \( Y^{\text{BS}} \) into \( \sum_{k=1}^{\text{rank}(Y)} \mathbf{e}_1 \mathbf{e}_2^H \) by SVD. Next, stacking \( \{e_{1,k}\}_{k=1}^{\text{rank}(Y)} \) into \( E_i \). Subsequently, using \( \Phi_1 = A_F \) and \( \Phi_2 = A_H \) to recover \( \Delta^M \) according to (21). Finally, the sparse signal \( \Delta \) in (14) is estimated by

\[
\hat{\Delta} = \Delta^M (\Delta^2_H)^H. \tag{22}
\]

### C. M-GAMP

Similar to the GAMP algorithm, M-GAMP consists of the two main components: prior distribution setup and posterior distribution approximation. According to **Theorem 1**, the whole recovery problem in (14) is decoupled into two separate subproblems as (21). Due to the independence of these two identical subproblems, we ignore the subscript \( i \) to obtain

\[
\arg \min_{\Delta^M} \|\Delta^M\|_0 \quad \text{s.t. } \|E - \Phi \Delta^M\|_F^2 \leq \epsilon, \quad \text{supp}(t_1) = \cdots = \text{supp}(t_R). \tag{23}
\]

We assume that \( E \in \mathbb{C}^{M \times R} \), \( \Phi \in \mathbb{C}^{M \times G} \) and \( \Delta^M = [t_{1,1}, \ldots, t_{R}] \in \mathbb{C}^{G \times R} \).

1) **Prior Distribution Setup:** To characterize the channel’s sparsity, the Gaussian mixture model (GMM)-based prior distribution is adopted, which is the extension of the Bernoulli-Gaussian distribution. Assume that the elements in \( \Delta^M \) follow the i.i.d. Gaussian mixture distribution, i.e.,

\[
p(t_{g,r}) = (1 - \kappa_g)\delta(t_{g,r}) + \kappa_g \sum_{l=1}^{L} \omega_{r,l} \mathcal{CN}(t_{g,r} | \nu_{r,l}, \varsigma_{r,l}), \tag{24}
\]

where \( \delta(\cdot) \) is the Dirac function, \( \kappa_g \) is the common sparsity rate for \( \forall r \), \( \nu_{r,l} \) and \( \varsigma_{r,l} \) denote the weight coefficient, mean and variance of the \( l \)-th GM component, respectively. In particular, the weight coefficients satisfy \( \sum_{l=1}^{L} \omega_l = 1 \). Additionally, we suppose that the noise follows the complex Gaussian distribution with variance \( \rho \), i.e.,

\[
p(n_{m,r} | \rho) = \mathcal{CN}(n_{m,r} | 0, \rho). \tag{25}
\]

Finally, let \( \Theta = \{\kappa_g, \omega_{r,l}, \nu_{r,l}, \varsigma_{r,l}, \rho_r | \forall g, r \} \) be the hyperparameter set mentioned above.
2) Posterior Distribution Approximation via M-GAMP:

In the construction of the GAMP algorithm, \( p(\eta_{m,r}|E) \) and \( p(t_{g,r}|E) \) approximations are adopted.

Firstly, to establish the relationship between the noiseless output and the measurement matrix \( E \), we have \( \eta_{m,r} = \{ \Phi_{m,i} \}_{i=1}^{M} | r \). The true posterior \( p(\eta_{m,r}|E, \Theta) \) is approximated by GAMP:

\[
\hat{p}(\eta_{m,r}|E, \Theta, \hat{\mu}_{m,r}, \tau_{m,r}^{\mu}) = \frac{p(e_{m,r}|\eta_{m,r}, \Theta)CN(\eta_{m,r}|\hat{\mu}_{m,r}, \tau_{m,r}^{\mu})}{\int_{\eta_{m,r}} p(e_{m,r}|\eta_{m,r}, \Theta)CN(\eta_{m,r}|\hat{\mu}_{m,r}, \tau_{m,r}^{\mu})},
\]

where \( \hat{\mu}_{m,r} \) and \( \tau_{m,r}^{\mu} \) denote the mean and variance of \( \eta_{m,r} \), respectively, which are iteratively updated during the GAMP procedure. Under the additive white Gaussian noise assumption, we obtain \( p(e_{m,r}|\eta_{m,r}, \Theta) = CN(e_{m,r}|\eta_{m,r}, \rho) \). The mean and variance of (25) can be expressed as

\[
\hat{\eta}_{m,r} = \hat{\mu}_{m,r} + \frac{\tau_{m,r}^{\mu}}{\tau_{m,r} + \rho} (e_{m,r} - \hat{\mu}_{m,r}), \quad \tau_{m,r}^{\eta} = \frac{\tau_{m,r}^{\mu} \rho}{\tau_{m,r} + \rho},
\]

where \( \hat{\eta}_{m,r} \) and \( \tau_{m,r}^{\eta} \) are iteratively updated during the GAMP procedure.

Since GAMP assumes posterior independence among hidden variables \( \{ t_{g,r} \}_{g=1}^{G} \), \( \{ \eta_{r,l} \}_{r=1}^{R} \), it approximates the true posterior \( p(t_{g,r}|E, \Theta) \) by

\[
\hat{p}(t_{g,r}|E, \Theta, \hat{\psi}_{g,r}, \tau_{g,r}^{\psi}) = \frac{p(t_{g,r}|\Theta)CN(t_{g,r}|\hat{\psi}_{g,r}, \tau_{g,r}^{\psi})}{\int_{t_{g,r}} p(t_{g,r}|\Theta)CN(t_{g,r}|\hat{\psi}_{g,r}, \tau_{g,r}^{\psi})}.
\]

Using the multiplication rule of the Gaussian distribution, i.e.,

\[
CN(x|\mu_{a}, \tau_{a})CN(x|\mu_{b}, \tau_{b}) = CN(x|\frac{\mu_{a} \mu_{b}}{\tau_{a} + \tau_{b}} + \frac{1}{\tau_{a} + \tau_{b}}CN(x|\mu_{a} - \mu_{b}, \tau_{a} + \tau_{b}),
\]

(27) with (24) substituted is given by

\[
\hat{p}(t_{g,r}|E, \Theta, \hat{\psi}_{g,r}, \tau_{g,r}^{\psi}) = \frac{(1 - \kappa_{g})\delta(t_{g,r})CN(t_{g,r}|\hat{\psi}_{g,r}, \tau_{g,r}^{\psi})}{\int_{t_{g,r}} p(t_{g,r}|\Theta)CN(t_{g,r}|\hat{\psi}_{g,r}, \tau_{g,r}^{\psi})} + \kappa_{g}\sum_{l=1}^{L} \omega_{l}CN(t_{g,r}|\theta_{l}, \xi_{l})CN(t_{g,r}|\hat{\psi}_{g,r}, \tau_{g,r}^{\psi})
\]

\[
= (1 - \omega_{g,r})\delta(t_{g,r}) + \omega_{g,r}\sum_{l=1}^{L} \nu_{g,r,l}CN(t_{g,r}|\theta_{g,r,l}, \xi_{g,r,l})
\]

with \( \theta_{g,r,l} = \hat{\psi}_{g,r}/\tau_{g,r}^{\psi} + \nu_{g,r,l}/\nu_{g,r,l} \), \( \xi_{g,r,l} = 1/\tau_{g,r}^{\psi} + 1/\nu_{g,r,l} \), and

\[
\omega_{g,r} = \frac{1}{1 + \frac{\sum_{l=1}^{L} \omega_{g,r,l}CN(\hat{\psi}_{g,r}|\theta_{g,r,l}, \xi_{g,r,l}) + \tau_{g,r}^{\psi}}{\omega_{g,r}CN(\eta_{g,r}|\hat{\mu}_{g,r}, \tau_{g,r}^{\mu})}},
\]

\[
\nu_{g,r,l} = \frac{\omega_{g,r}CN(\hat{\psi}_{g,r}|\theta_{g,r,l}, \xi_{g,r,l}) + \tau_{g,r}^{\psi}}{\sum_{l=1}^{L} \omega_{g,r,l}CN(\hat{\psi}_{g,r}|\theta_{g,r,l}, \xi_{g,r,l}) + \tau_{g,r}^{\psi}}.
\]

3) Input and Output Functions: With the above approximations, two scalar functions are defined: \( g_{in}(\cdot) \) and \( g_{out}(\cdot) \). Using the minimum mean squared error (MMSE) mode, the input function \( g_{in}(\cdot) \) is expressed as \( g_{in}(\hat{\psi}_{g,r}, \tau_{g,r}^{\psi}, \Theta) = \mathbb{E}(t_{g,r}|E, \Theta, \hat{\psi}_{g,r}, \tau_{g,r}^{\psi}) \), and the scaled partial derivative of \( \tau_{g,r}^{\psi}g_{in}(\hat{\psi}_{g,r}, \tau_{g,r}^{\psi}, \Theta) \) w.r.t. \( \hat{\psi}_{g,r} \) is \( \tau_{g,r}^{\psi}\tau_{g,r}^{\psi}g_{in}(\hat{\psi}_{g,r}, \tau_{g,r}^{\psi}, \Theta) = \mathbb{V}(t_{g,r}|E, \Theta, \hat{\psi}_{g,r}, \tau_{g,r}^{\psi}) \), where \( \mathbb{E}\{\cdot\} \) and \( \mathbb{V}\{\cdot\} \) denote the expectation and variance operations, respectively.

The output scalar function \( g_{out}(\cdot) \) is expressed as \( g_{out}(\hat{\mu}_{m,r}, \tau_{m,r}^{\mu}, \Theta) = \frac{\hat{g}_{m,r} - \mu_{m,r}}{\tau_{m,r}^{\mu} + \rho} \). Moreover, the partial derivative of \( g_{out}(\hat{\mu}_{m,r}, \tau_{m,r}^{\mu}, \Theta) \) w.r.t. \( \hat{\mu}_{m,r} \) is \( \delta_{\hat{\mu}_{m,r}}g_{out}(\hat{\mu}_{m,r}, \tau_{m,r}^{\mu}, \Theta) = \frac{1}{\tau_{m,r}^{\mu} + \rho} \). Based on the GAMP flow, this can be summed up as Algorithm 1 lines 5-18 given definitions of \( g_{in}(\cdot) \) and \( g_{out}(\cdot) \).

D. Hyperparameter Learning via EM

It is well established that the EM algorithm aims at iteratively maximizing the evidence lower bound (ELBO). We describe briefly the expectation and maximization procedures that derive and optimize ELBO based on the MMV problem.

1) Expectation-Step: From the perspective of maximum likelihood estimate (MLE), \( \log(E|\Theta) \) tends to be maximized w.r.t. \( \Theta \). This is challenging due to the existence of hidden variables. Thus, EM uses a lower bound w.r.t. ELBO to approach the likelihood, i.e.,

\[
\log(E|\Theta) = \int_{\Delta^{M}} q(\Delta^{M})\log\frac{p(\Delta^{M}|E, \Theta)}{q(\Delta^{M})}d\Delta^{M}
\]

\[
- \int_{\Delta^{M}} q(\Delta^{M})\log\frac{p(\Delta^{M}|E, \Theta)}{q(\Delta^{M})}d\Delta^{M} \leq \text{ELBO} - \text{KL}(q(\Delta^{M})||p(\Delta^{M}|E, \Theta)) \leq \text{ELBO} - \text{KL}(q(\Delta^{M})||p(\Delta^{M}|E, \Theta)).
\]

Since the Kullback-Leibler (KL) divergence is nonnegative, in the \( o \)-th iteration, \( \log(E|\Theta) \) takes the maximum value when \( \text{KL}(q(\Delta^{M})||p(\Delta^{M}|E, \Theta)) = 0 \), i.e., \( q(\Delta^{M}|E) = p(\Delta^{M}|E, \Theta) \).

2) Maximization-Step: On the basis of the E-step, \( \Theta^{o+1} \), where \( o + 1 \) is the current iteration, is solved by

\[
\Theta^{o+1} = \text{arg max}_{\Theta} \int_{\Delta^{M}} p(\Delta^{M}|E, \Theta^{o})\log\frac{p(\Delta^{M}|E, \Theta^{o})}{p(\Delta^{M}|E, \Theta)}d\Delta^{M}
\]

\[
= \text{arg max}_{\Theta} \mathbb{E}_{\Delta^{M}|E, \Theta^{o}}\log p(\Delta^{M}|E, \Theta) \Omega
\]

Due to the difficulty of calculating \( p(\Delta^{M}|E, \Theta^{o}) \), GAMP is used to attain the approximated \( \hat{p}(\Delta^{M}|E, \Theta^{o}) \). Similar to [43], the updating rule of \( \Theta^{o+1} \) is given by

\[
\Theta^{o+1} = \frac{1}{M} \sum_{m=1}^{M} \left( \frac{\epsilon_{m,r} - \hat{\mu}_{m,r}}{\hat{\mu}_{m,r}^{2} + 1} + \frac{\tau_{m,r}^{\mu} \rho_{r}^{2}}{\tau_{m,r}^{\mu} + \rho_{r}^{2}} \right).
\]

\footnote{The reference [46] provides details on the origin of the GAMP algorithm.}
 Particularly, the sparsity factor is determined by the average of $R$ measurement vectors, i.e.,

\[
\kappa_{g,r}^{o+1} = \frac{1}{R} \sum_{r=1}^{R} \frac{1}{1 + \left( \frac{\sum_{l=1}^{T} \kappa_{g,r}^{o+1} \psi_{\theta_{b}} \theta_{g,r,l}^{o+1} + \tau_{g,r,l}^{o+1}}{1 - \kappa_{r}^{o}} \right)^{2}}.
\]  

(38)

In summary, the EM procedure can be found in Algorithm 1 lines 21-24.

E. LoS-Aided Off-Grid Refinement

Note that the sensing matrix $A_{F/H}$ in (15) is comprised of two matrices. Hence, we suppose that $\Phi = CA$, where $C \in \mathbb{C}^{M \times N}$ is the measurement matrix (corresponding to the transmit/receive beams), and $A \in \mathbb{C}^{N \times G}$ is the partial dictionary regarding (3) for virtual channel representation, in which each column represents a recovered atom.

Assuming that (23) is solved by the previous recovery procedure, and that the angle set $\{\theta_{b}^{o+1}\}_{b=1}^{P}$ of $A$ is estimated, the following perturbation-based off-grid problem is formed:

\[
\begin{align*}
\text{arg min}_{\{\theta_{b}^{o+1}\}_{b=1}^{P}, \Delta^{M}} & \quad \|E - C \sum_{b=1}^{P} a_b \cos(\theta_{b}^{o+1}) (1 + \mathcal{P}_{\theta_{b}^{o+1},b}) \|_{2}^{2}, \\
\text{subject to} & \quad \text{sin}(\theta_{b}^{o+1}) \text{sin}(\theta_{a}^{o+1}) (1 + \mathcal{P}_{\theta_{b}^{o+1},b}) \|_{2}^{2},
\end{align*}
\]

(39)

where $2 \theta_{b}^{o+1} = 1$ is set, $\mathcal{P}_{\theta_{b}^{o+1}} = \{\mathcal{P}_{\theta_{b}^{o+1},1}, \cdots, \mathcal{P}_{\theta_{b}^{o+1},P}\}$ and $\mathcal{P}_{\theta_{b}^{o+1},a} = \{\mathcal{P}_{\theta_{b}^{o+1},a,1}, \cdots, \mathcal{P}_{\theta_{b}^{o+1},a,p}\}$ are the perturbations, and $\Delta^{M} = [\delta_{1}^{M}, \cdots, \delta_{P}^{M}]^{T} \in \mathbb{C}^{P \times R}$ is the estimated sparse signal.

Note that the LoS path does not need to be refined due to LoS-aided dictionaries used. Thus, the off-grid procedure is simplified. For clarity, the path index $b = 1$ is assumed to be the LoS path. Denoting $\cos(\theta_{b}^{o+1})$ and $\text{sin}(\theta_{b}^{o+1})$ by $\theta_{b}^{o+1}$ and $\theta_{a}^{o+1}$ respectively. Based on the first-order Taylor expansion, we have

\[
\begin{align*}
\text{arg min}_{\{\theta_{b}^{o+1}\}_{b=1}^{P}, \Delta^{M}} & \quad \|R - C \sum_{b=2}^{P} \left( \mathcal{P}_{\theta_{b}^{o+1},b} \Delta_{b}^{o+1} \right) \|_{2}^{2}, \\
\text{subject to} & \quad \|R - C \sum_{b=2}^{P} \left( \mathcal{P}_{\theta_{b}^{o+1},b} \Delta_{b}^{o+1} \right) \|_{2}^{2},
\end{align*}
\]

(40)

where $R = E - Ca \left( \theta_{b}^{o+1}, \theta_{a}^{o+1} \right) \hat{r} = C \sum_{b=2}^{P} a_{b} \left( \theta_{b}^{o+1}, \theta_{a}^{o+1} \right) \hat{r}_{b}$ is the residual. For this coupled optimization, the alternating iteration algorithm is used. We first present the following theorem w.r.t. the matrix-based least squares (MLS) algorithm.

\textbf{Theorem 2:} Postulate that $Y \in \mathbb{C}^{M \times R}$, $A_{b} \in \mathbb{C}^{M \times R}$, $b \in \{1, 2, \cdots, P\} \quad (P \leq MR)$ and $w = [w_{1}, w_{2}, \cdots, w_{P}] \in \mathbb{C}^{P \times 1}$, the closed-form solution w.r.t. $w$ of

\[
\begin{align*}
\text{min} & \quad \| \hat{Y} - X_{0}^{w}A_{0} \|_{2}^{2}, \\
\text{subject to} & \quad \| \hat{Y} - X_{0}^{w}A_{0} \|_{2}^{2},
\end{align*}
\]
arg min \|Y - \sum_{b=1}^{P} w_{b} A_{b}\|_{F}^{2} \text{ is } w = (G^{H} G)^{-1}G^{H} \text{vec}(Y)

with \( G = \{ \text{vec}(A_{1}), \text{vec}(A_{2}), \cdots, \text{vec}(A_{P}) \} \in \mathbb{C}^{MR \times P} \).

**Proof:** See Appendix B.

Based the above theorem, the updating rules w.r.t. \( \bar{\theta}_{b}^{\text{ele}} \) and \( \bar{\theta}_{b}^{\text{azi}} \) \((\forall b, b \neq 1)\) are given by

\[
\bar{\theta}_{b}^{\text{ele}, t+1} = \frac{\partial a(\bar{\theta}_{b}^{\text{ele}, t}, \bar{\theta}_{b}^{\text{azi}, t})}{\partial \bar{\theta}_{b}^{\text{ele}, t}} = \sqrt{\frac{1}{N_{b}N_{z} \left[ 1, \cdots, e^{j\pi(N_{b}-1)\bar{\theta}_{b}^{\text{azi}, t}} \right]^{T}} \circ \left[ 1, \cdots, j\pi(N_{b}-1) \right]^{T}}
\]
\[
\bar{\theta}_{b}^{\text{azi}, t+1} = \frac{\partial a(\bar{\theta}_{b}^{\text{ele}, t}, \bar{\theta}_{b}^{\text{azi}, t})}{\partial \bar{\theta}_{b}^{\text{azi}, t}} = \sqrt{\frac{1}{N_{b}N_{z} \left[ 1, \cdots, e^{j\pi(N_{b}-1)\bar{\theta}_{b}^{\text{azi}, t}} \right]^{T}} \circ \left[ 1, \cdots, j\pi(N_{b}-1) \right]^{T}}
\]

After \( \bar{\theta}_{b}^{\text{ele}} \) and \( \bar{\theta}_{b}^{\text{azi}} \) are updated, the partial dictionary matrix w.r.t. angles is also updated as \( \Delta^{M+1} \). Then the coefficient matrix and the residual are renewed by

\[
\Delta^{M+1} = \Delta^{M+1,H} C^{H} \Delta^{M+1} - 1 \Delta^{M+1,H} C^{H} \bar{E}
\]
\[
R^{I+1} = \bar{E} - \Delta^{M+1,H} \Delta^{M+1}
\]

where \( \bar{E} = E - \text{vec}(\bar{\theta}_{1}^{\text{ele}}, \bar{\theta}_{1}^{\text{azi}}) \) is the signal after the LoS path is removed. Then the off-grid procedure performs \( T \) iterations. For a stable solution, we set an iteration termination condition that stops the off-grid procedure when \( R^{I+1} \geq R^{I} \). Finally, the estimated sparse signal \( \Delta^{M} \) can be obtained by filling zero rows into \( \Delta^{M,T} \). In summary, the whole two-timescale channel estimation process is shown in Algorithm 2.

### V. Simulation Results

This section supports simulations to determine the effectiveness of our proposed strategies. To begin, we define the normalized mean square error (NMSE) as \( E\{ \| \hat{D} - D \|_{F}^{2} \} \), where \( \hat{D} \) is the estimated RIS 1-RIS 2 channel. The NMSE definition of other channels is similar to that of the above. The system setup considered for our simulations consists of \( J = 36 \), \( L_{1} = L_{2} = 64 \), and system frequency \( f = 28 \) GHz with bandwidth \( B = 100 \) kHz. The BS and RISs are all assumed to be equipped with uniform square arrays. The noise figure \( (NF) \) is set to 9. Hence, the noise power is given by \( \sigma_{n}^{2} = -174 \text{dBm/Hz} + 10\log_{10}(B) + NF \). For channel parameters, we employ the 28 GHz mmWave channel setting in [47]. Following Table I in [47], we assume that that path gains follow \( CN(0, 0.1^{0.1}) \), where \( B = K_{1} \times 10^{0.1k} \) with \( K_{1} \sim \mathcal{U}(0, 1) \) and \( K_{2} \sim \mathcal{C}N(0, 0.16) \). Besides, \( PL = a_{1} + 10a_{2}\log_{10}(d) + N(0, \sigma) \) with \( d \) denoting the distance in meters, where \( a_{1} = 72, a_{2} = 2.92 \) and \( \sigma = 8.7 \) for NLoS paths, and \( a_{1} = 61.4, a_{2} = 2 \) and \( \sigma = 5.8 \) for LoS paths.

For the deployment of the BS and RISs, we assume the three-dimensional Cartesian coordinates of the BS, RIS 1 and RIS 2 are \((0, 0, 5), (0, \sqrt{2}, \sqrt{2}, 6), (0, \sqrt{2}, 0, 100, \sqrt{2}, 6), \) and the users are uniformly distributed around RIS 2 at a distance of 1 to 30 meters. Moreover, we assume \( P_{d} = P_{f} = P_{h,i,u} = 3 \) for \( \forall i \), the LoS paths of BS-RISs and RIS 1-RIS 2 can be calculated by their locations. The LoS paths of users’ channels and NLoS paths of all channels are uniformly distributed in elevation and azimuth coverage. All users’ uplink power \( \sigma_{u}^{2} \) is identical. All the measurement matrices, i.e., the RIS phase shifts, are generated by making each element of them follow the complex Gaussian distribution \( \mathcal{CN}(0, 1) \) with modulus 1. Lastly, the size of dictionaries is set as \( G_{t} = J \) and \( G_{r} = L \).
A. Evaluation of Large-Timescale Channel Estimation

Here, we will highlight the excellence of our proposed large-timescale channel estimation method. To this end, we will evaluate our proposed low-complexity Bayesian framework and the MAE method for large-timescale estimation. Moreover, we use the greedy methods, OMP and simultaneous orthogonal matching pursuit (SOMP) [48], as the algorithm benchmark.

1) CS Framework Comparison: We first evaluate the effectiveness of our proposed CS frameworks, i.e., SVD-MMV-CS and off-grid SVD-MMV-CS frameworks. For a good comparison, the Kronecker CS and SVD-CS frameworks are employed as benchmarks. Note that (12) and (14) depend on the RIS-user channels, i.e., $h_{i,u}$, $\forall i, u$, here we assume they are accurately estimated for CS framework comparison. Certainly, the impact of these channels’ estimation accuracy will be explored later. Firstly, Fig. 4 exhibits the NMSE performance of different schemes regarding the estimation of the BS-RIS 1 channel, BS-RIS 2 channel and RIS 1-RIS 2 channel, where uplink power $\sigma_p^2 = 30$ dBm and the pilot overhead $Q_1 = Q_2$ ranges from 16 to 64. Particularly, the pilot overhead $Q_3 = N_X \times N_Y$ for RIS 1-RIS 2 channel estimation follows that $N_X = N_Y$ ranges from 16 to 64. It can be observed that the proposed SVD-MMV-CS framework outperforms the SVD-CS and Kronecker CS frameworks in terms of the NMSE. Besides, the NMSE performance of on-grid EM-GAMP-based methods is much better than that of on-grid OMP-based methods. Furthermore, the off-grid technique based on the proposed framework can achieve a significant performance improvement (please see the red and black curves in Fig. 4). Notably, the NMSE of Fig. 4. (a) is higher than that of Fig. 4. (b). This is because the path loss of the BS-RIS 2-user channel is generally smaller than that of the BS-RIS 1-user channel. To evaluate the impact of the uplink power on the NMSE performance of different schemes, we set $\sigma_p^2 = 15$ dBm to repeat the above experiments in Fig. 5. It can be shown that all schemes’ NMSEs become larger. Despite this, the off-grid M-EM-GAMP algorithm based on the SVD-MMV-CS framework has a considerable performance at a low pilot overhead value.

2) Estimation Strategy Evaluation: Note that the uplink training of the BS-RIS $i$ channel is reliant on the previous step of estimating $h_{i,u}, V_i, u$. Besides, the estimation of $D$ depends on $h_{1,u}$ and $F_2$ according to (15). Hence, we will explore the impact of $h_{i,u}$ and $F_2$ on the estimation of the BS-RIS 1, BS-RIS 2 and RIS 1-RIS 2 channels. Toward this end, we compare the NMSE performance of different schemes under the perfect channel assumption and the imperfect channel assumption, respectively. The imperfect channel assumption means that the channel required to estimate the large-timescale channel is obtained by the specified method. For example, when estimating $D$, the required channel $h_{1,u}$ is obtained by using the off-grid EM-GAMP method to solve (9) instead of being perfectly known. It is noteworthy that when estimating $D$, the required channel $F_2$ is obtained under the imperfect channel assumption. As shown in Fig. 6 with uplink power $\sigma_p^2 = 30$ and 15 dBm, the NMSE performance under the perfect channel assumption is much better than that under

![Fig. 4. NMSE performance of different schemes versus pilot overhead at $\sigma_p^2 = 30$ dBm.](image1)

![Fig. 5. NMSE performance of different schemes versus pilot overhead at $\sigma_p^2 = 15$ dBm.](image2)
the imperfect channel assumption when the pilot overhead is small. As the pilot overhead increases, the NMSE gap between the two assumptions becomes narrow, indicating that the impact of the required channels on large-timescale channel estimation becomes small as the pilot overhead becomes large. This is because the estimation accuracy of the required channel increases with the increase of the pilot overhead.

**B. Evaluation of Small-Timescale Channel Estimation**

To show the effectiveness of the MAE method for small-timescale channel estimation, we use the method of directly estimating the RIS-user channel at the RIS using the receive RF chain as the benchmark, which is the first step of the estimation of large-timescale channels. Seeing Fig. 7, in which the pilot overhead ranges from 16 to 64, and $\sigma_p^2$ is set to 15 and 30 dBm, it can be concluded that the NMSE performance of the MAE-based methods is better than the direct estimation-based methods due to the former’s higher dimensional sensing matrix. When using the direct estimation strategy, the NMSE performance of Fig. 7 (a) is slightly better than that of Fig. 7 (b) since users are close to RIS 1 than to RIS 2. Furthermore, Fig. 7 indicates that our proposed MAE strategy used in the small-timescale stage can reduce the pilot overhead when estimating the RIS-user channel. That is, when achieving an NMSE performance index, the MAE strategy requires less pilot overhead than the direct estimation strategy, $Q_i < Q_1$, $i \in \{1, 2\}$. Additionally, we evaluate the impact of the large-timescale estimation performance on the MAE-based small-timescale estimation strategy. In Fig. 8, the NMSE performance of different schemes under the perfect and imperfect channel assumptions is shown with different uplink power. Here, the imperfect channel assumption denotes the large-scale channel required for RIS-user channel estimation is obtained by the performance-optimal SVD-MMV-CS-based off-grid M-EM-GAMP algorithm. It can be observed that the perfect channel assumption-based NMSE performance is better than the imperfect channel assumption-based NMSE performance. As the pilot overhead increases, the NMSE performance of the two assumptions-based schemes is gradually approaching.

**C. Complexity Analysis**

Table. I provides a comparison of various schemes regarding pilot overhead and computational complexity, where the recovery algorithm options ① and ② denote the the variant algorithms of OMP and EM-GAMP, respectively. Notably, all schemes are conducted by two algorithms, i.e., ① and ②. We will discuss the computing cost of OMP-based and EM-GAMP-based methods’ successively. First, we denote the iteration number for the two algorithms and the off-grid method by $I_1$, $I_2$ (both the GAMP and EM loops are considered) and $I_3$, respectively. Since the channel is assumed to be sparse, the iteration number $I_1$ is generally small and set to 10. $I_2$ is set as [43], where the maximum
Two-Timescale Estimation | Average Pilot Overhead | CS Framework | Recovery Algorithm | Computational Complexity
---|---|---|---|---
Large-Timescale Estimation | $S_1^{*}(Q_1 + Q_2 + Q_3)$ | Kronecker CS | $\mathcal{O}(I_1 N_X N_Y U J G_2)$ |
 | | SVD-CS | $\mathcal{O}(I_2 N_Y G_s)$ |
 | | Proposed On-Grid | $\mathcal{O}(I_1 N_Y G_s)$ |
 | | Proposed Off-Grid | $\mathcal{O}(I_1 N_Y G_s)$ |
 | | Proposed On-Grid | $\mathcal{O}(I_2 N_Y G_s)$ |
 | | Proposed Off-Grid | $\mathcal{O}(I_2 N_Y G_s)$ |
Small-Timescale Estimation | $\bar{Q}_1 + \bar{Q}_2$ | Proposed On-Grid | $\mathcal{O}(I_1 Q_1 J G_s)$ |
 | | Proposed Off-Grid | $\mathcal{O}(I_1 Q_1 J G_s)$ |
 | | Standard CS | $\mathcal{O}(I_2 Q_1 J G_s)$ |
 | | Standard CS | $\mathcal{O}(I_2 Q_1 J G_s)$ |

Table I

A Comparison of Pilot Overhead and Computational Complexity of Various Schemes

Fig. 8. NMSEs of small-timescale channel estimation with perfect/imperfect channel.

M-GAMP-iteration number is 20 with tolerance $10^{-5}$, and the maximum EM-iteration number is 20 with tolerance $10^{-5}$. Thus, the maximum iteration number for $I_2$ is set to 400. Moreover, iteration number $I_3$ is set to 50.

1) Large-Timescale: We first consider the pilot overhead of large-timescale estimation. Due to the fact that the large-timescale channels are estimated only once in a timescale period, the average pilot overhead is $S_r^{*}(Q_1 + Q_2 + Q_3)$, with $S_1$ and $S_2$ denoting the large-timescale and small-timescale channel coherence time, respectively. Since the computing cost of estimating $\mathbf{D}$ dominates the whole large-scale channel estimation procedure, the computational complexity of estimating $\mathbf{D}$ is analyzed in detail. Recalling (15), the kronecker CS framework will process the large-scale sensing matrix. It is well established that the complexity of OMP is dominated by the atom matching step, which calculates the projection of $\mathbf{A}_H^T \otimes \mathbf{A}_Y$ onto $\tilde{\mathbf{y}}^{\text{BS}}$, comprised of a $(G_s^2 \times N_X N_Y U J) \times (N_X N_Y U J \times 1)$ matrix-vector multiplication. Thus, the whole complexity of Kronecker OMP is $\mathcal{O}(I_1 N_X N_Y U J G_2^2)$. On the contrary, the SVD-CS framework deals with two parallel subproblems with small-scale sensing matrices. Thus, this two subproblems need to calculate two simple projections with $N_X U G_s R$ and $N_Y J G_s R$, where $R$ is the rank of the received signal and is set to 3.\(^5\) On the other hand, $N_X$ and $N_Y$ are the number of measurements of the two RISs, they are approximately equal. Generally, the number of access users is not greater than that of BS’s antennas, i.e., $U \leq J$. Hence, $N_X U$ and $N_Y J$ are on the same order of magnitude. Thus, the whole complexity can be expressed as $\mathcal{O}(I_1 N_Y G_s J)$. Furthermore, our proposed off-grid SVD-MMV-CS incurs the same complexity as SVD-CS. The difference between them is when processing all measurement vectors’ atom projections, SVD-CS can calculate them in parallel, while SVD-MMV-CS requires all measurement vectors to calculate them jointly. Next, we will discuss the complexity of our proposed off-grid method based on the SVD-MMV-CS framework. Let us consider (41) for the off-grid solutions of the two subproblems, it will produce a complexity of $N_Y J R (P - 1)^2 + (P - 1)^3$ and $N_X U R (P - 1)^2 + (P - 1)^3$, respectively. $(P - 1)^3$ produced by the inverse operation can be ignored because of the small estimated number of paths $P \geq 1$.\(^6\) Considering (45) yields a different complexity of $N_Y J L (P - 1)$ or $N_X U L (P - 1)$. Note that $(P - 1) R$ is generally smaller than $L$ since both $P - 1$ and $R$ are related to the low rank of the channel. In addition, according to the relation between $N_X U$ and $N_Y J$ discussed before, the whole complexity of the proposed off-grid method is summarized as $\mathcal{O}(I_2 Q_1 J G_s)$.

\(^5\)Indeed, $R$ exactly equals to the number of paths, i.e., the channel’s rank, with the noiseless received signal. Although considering the noise effect, we set $R$ to the number of paths, i.e., $R = 3$, for simplicity.

\(^6\)Due to the known LoS path, the number of paths requiring off-grid refinement is reduced to $P - 1$. 
2) Small-Timescale: The pilot overhead of this timescale is \( \bar{Q}_1 + \bar{Q}_2 \). Generally, \( \bar{Q}_1 + \bar{Q}_2 \ll Q_1 + Q_2 \) owing to the proposed MAE strategy. Similar to the above analysis, recalling (16), the per-iteration complexity of OMP is given by \( O(Q_1JG_r + Q_2JG_r) \). Since \( Q_1 \) and \( Q_2 \) are comparable, the complexity is simplified as \( O(Q_1JG_r) \). Further, according to the previous analysis, the per-iteration complexity of off-grid refinement is \( O((\bar{Q}_1 + \bar{Q}_2)PLJ) \). Hence, the total complexity of the off-grid EM-GAMP algorithm for small-timescale estimation is \( O(\bar{Q}_1J(I_1G_r + I_3PL)) \). Along with this, the complexity of the off-grid EM-GAMP algorithm for small-timescale estimation is \( O(\bar{Q}_1J(I_2G_r + I_3PL)) \).

VI. CONCLUSION

This paper has effectively addressed the problem of 3D double-RIS-based separate channel estimation from two perspectives: 1) a low hardware cost and pilot overhead estimation protocol and 2) an efficient channel parameter recovery framework. For 1), we first propose a multi-user two-timescale channel estimation protocol based on the active RIS architecture with only one RF chain, where the multi-user uplink training strategy is proposed for performance improvement. Notably, our proposed two-timescale estimation protocol decreases the pilot overhead from two aspects. Using the conventional estimation protocol, the demanding times for estimating the large-timescale channels are equal to that for estimating the small-timescale channels. In contrast, the two-timescale estimation protocol requires less times to estimate the large-timescale channels, so as to decrease the pilot overhead. Moreover, simulations have demonstrated that our proposed MAE-based strategy can effectively reduce the pilot overhead for small-timescale channel estimation. For 2), the SVD-MMV-CS framework has been proposed for computing efficiently for the large-timescale estimation problem. Not only that, simulation results demonstrate that this framework has better NMSE performance than the conventional Kronecker framework and the SVD-CS framework. Furthermore, the LoS-aided M-EM-GAMP algorithm and its off-grid version are presented to estimate the channel parameters based on the SVD-MMV-CS framework. Simulation results have shown that the NMSE performance of EM-GAMP-based methods is better than that of OMP-based methods in different simulation scenarios. The future work will focus on deep learning-based channel estimation for double-RIS systems.

APPENDIX A

Recalling (20), we know that \( C(Y) \subset C(\Phi_1 \Delta) \subset \Phi_2 C(\Delta) \), as well as \( C(Y^H) \subset C(\Phi_2 \Delta^H) \subset \Phi_2 C(\Delta^H) \). Since \( e_{1,k} \) and \( e_{2,k} \) are the column vectors of the SVD of \( Y \) and \( Y^H \), respectively, such that \( e_{1,k} \in C(Y) \subset \Phi_1 C(\Delta) \) and \( e_{2,k} \in C(Y^H) \subset \Phi_2 C(\Delta^H) \). Therefore, the solution of \( \min_\Delta \|Y - \Phi_1 \Delta\|_F \) is \( \Delta \approx \Phi_2 e_{2,k} \approx \Phi_2 e_{2,k} \in C(\Delta^H) \). In particular, the relation of \( \delta_{1,k} = \delta_{2,k} \in C(\Delta) \) indicates that \( \delta_{1,k} = \delta_{2,k} \approx \delta_{e_{2,k}} \in C(\Delta^H) \).

APPENDIX B

Here, we give the proof of Theorem 2. Since
\[
\left\| Y - \sum_{b=1}^{P} w_b A_b \right\|_F^2 \geq \left\| \text{vec}(Y - \sum_{b=1}^{P} w_b A_b) \right\|_2^2,
\]
then the MLS issue is expressed as \( \arg \min_w \left\| \text{vec}(Y - \sum_{b=1}^{P} w_b A_b) \right\|_2^2 \). This can be solved by the conventional LS algorithm with \( \text{vec}(Y - \sum_{b=1}^{P} w_b A_b) = \text{vec}(Y) - Gw \), where \( G = [\text{vec}(A_1), \text{vec}(A_2), \ldots, \text{vec}(A_P)] \in \mathbb{C}^{MR \times P} \). Thus, we have \( w = (G^H G)^{-1}G^H \text{vec}(Y) \). Theorem 2 has been proved.

ACKNOWLEDGMENT

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore.

REFERENCES

[1] W. Tang et al., “Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement,” IEEE Trans. Wireless Commun., vol. 20, no. 1, pp. 421–439, Jan. 2021.
[2] X. Cao, B. Yang, H. Zhang, C. Huang, C. Yuen, and Z. Han, “Reconfigurable-intelligent-surface-assisted MAC for wireless networks: Protocol design, analysis, and optimization,” IEEE Internet Things J., vol. 8, no. 18, pp. 14171–14186, Sep. 2021.
[3] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, “Reconfigurable intelligent surfaces for energy efficiency in wireless communication,” IEEE Trans. Wireless Commun., vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
[4] A. Taha, M. Alraiebiah, and A. Alkhateeb, “Enabling large intelligent surfaces with compressive sensing and deep learning,” IEEE Access, vol. 9, pp. 44304–44321, 2021.
[5] R. Schroeder, J. He, and M. Juntti, “Passive RIS vs. hybrid RIS: A comparative study on channel estimation,” in Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring), Apr. 2021, pp. 1–7.
[6] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, “Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces,” IEEE Trans. Veh. Technol., vol. 69, no. 8, pp. 9223–9228, Aug. 2020.
[7] G. C. Alexandropoulos and E. Vlachos, “A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2020, pp. 9175–9179.
[8] R. Schroeder, J. He, G. Brante, and M. Juntti, “Two-stage channel estimation for hybrid RIS assisted MIMO systems,” IEEE Trans. Commun., vol. 70, no. 7, pp. 4793–4806, Jul. 2022.

[9] S. Yang, W. Lyu, D. Wang, and Z. Zhang, “Separate channel estimation with hybrid RIS-aided multi-user communications,” IEEE Trans. Veh. Technol., vol. 72, no. 1, pp. 1318–1324, Jan. 2023.

[10] Y. Han, S. Zhang, L. Duan, and R. Zhang, “Cooperative double-IRS aided communication design and power scaling,” IEEE Wireless Commun. Lett., vol. 9, no. 8, pp. 1206–1210, Aug. 2020.

[11] A. Papazafeiropoulos, P. Kourtessis, S. Chatzinotas, and J. M. Senior, “Coverage probability of double-IRS assisted communication systems,” IEEE Wireless Commun. Lett., vol. 11, no. 1, pp. 96–100, Jan. 2022.

[12] Y. Han, S. Zhang, L. Duan, and R. Zhang, “Double-IRS aided MIMO communication under LoS channels: Capacity maximization and scaling,” IEEE Trans. Commun., vol. 70, no. 4, pp. 2820–2837, Apr. 2022.

[13] C. Huang et al., “Multi-hop RIS-empowered terahertz communications: A DRL-based hybrid beamforming design,” IEEE J. Sel. Areas Commun., vol. 39, no. 6, pp. 1663–1677, Jun. 2021.

[14] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, “Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications,” IEEE Trans. Wireless Commun., vol. 21, no. 5, pp. 3083–3098, May 2022.

[15] C. Huang et al., “Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends,” IEEE Wireless Commun., vol. 27, no. 5, pp. 118–125, Oct. 2020.

[16] L. Wei et al., “Multi-user holographic MIMO surfaces: Channel modeling and spectral efficiency analysis,” IEEE J. Sel. Topics Signal Process., vol. 16, no. 5, pp. 1112–1124, Aug. 2022.

[17] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, “Intelligent reflecting surface meets OFDM: Protocol design and rate maximization,” IEEE Trans. Commun., vol. 70, no. 7, pp. 4552–4555, Jul. 2022.

[18] B. Zheng, C. You, and R. Fang, “Fast channel estimation for IRS-assisted OFDM,” IEEE Wireless Commun. Lett., vol. 10, no. 3, pp. 580–584, Mar. 2021.

[19] 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, 2020.

[20] Y. C. P. Rezaifar, P. S. Krishnarasadas, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in Proc. Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput., vol. 1. Pacific Grove, CA, USA, Nov. 1993, pp. 40–44.

[21] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, “Channel estimation for reconﬁgurable intelligent surface aided multi-user mmWave MIMO systems,” 2019, arXiv:1912.03619.

[22] Z.-Q. He and X. Yuan, “Cascaded channel estimation for large intelligent metasurface assisted massive MIMO,” IEEE Wireless Commun. Lett., vol. 9, no. 2, pp. 210–214, Feb. 2020.

[23] L. Wei et al., “Joint channel estimation and signal recovery for RIS-empowered multiuser communications,” IEEE Trans. Commun., vol. 70, no. 7, pp. 4640–4655, Jul. 2022.

[24] J. He, H. Wyneersch, 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, Sep. 2021.

[25] L. Wei, C. Huang, G. C. Alexandropoulos, and C. Yuen, “Parallel factor decomposition channel estimation in RIS-assisted multi-user MISO communication,” in Proc. IEEE 11th Sensor Array MultiChannel Signal Process. Workshop (SAM), Jun. 2020, pp. 1–5.

[26] G. T. de Araújo, A. L. F. de Almeida, and R. Boyer, “Channel estimation for reconfigurable intelligent surface aided wireless communications,” in Proc. IEEE Int. Conf. Commun., vol. 70, no. 4, pp. 2404–2416, Apr. 2022.

[27] Y. Sun et al., “Energy-efficient hybrid beamforming for multilayer RIS-aided secure integrated terrestrial-aerial networks,” IEEE Trans. Commun., vol. 70, no. 6, pp. 4189–4210, Jun. 2022.

[28] S. Bazzi and X. Wu, “IRS parameter optimization for channel estimation MSE minimization in double-IRS aided systems,” IEEE Wireless Commun. Lett., vol. 11, no. 10, pp. 2170–2174, Oct. 2022.

[29] B. Zheng, C. You, and R. Fang, “Uplink channel estimation for double-IRS assisted multi-user MIMO,” in Proc. IEEE Int. Conf. Commun., Jun. 2021, pp. 1–6.

[30] K. Ardah, S. Gheherkhou, A. L. F. de Almeida, and M. Haardt, “Double-RIS versus single-RIS aided systems: Tensor-based MIMO channel estimation and design perspectives,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), May 2022, pp. 5183–5187.

[31] C. You, B. Zheng, and R. Zhang, “Wireless communication via double IRS: Channel estimation and passive beamforming designs,” IEEE Wireless Commun. Lett., vol. 10, no. 2, pp. 431–435, Feb. 2021.

[32] B. Zheng, C. You, and R. Zhang, “Efficient channel estimation for double-IRS aided multi-user MIMO system,” IEEE Trans. Commun., vol. 69, no. 6, pp. 3818–3832, Jun. 2021.

[33] Z. Wang, Y. Cai, A. Li, L. Wang, and G. Yue, “Two-timescale uplink channel estimation for dual-hop MIMO relay multi-user systems,” IEEE Trans. Veh. Technol., vol. 70, no. 5, pp. 4724–4739, May 2021.

[34] 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.

[35] S. Friedland, Q. Li, and D. Schonfeld, “Compressive sensing of sparse tensors,” IEEE Trans. Image Process., vol. 23, no. 10, pp. 4438–4447, Oct. 2014.

[36] Y. Chi, A. Pezeshki, L. Scharf, and R. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Mar. 2010, pp. 3930–3933.

[37] J. P. Vila and P. Schniter, “Expectation-maximization Gaussian-mixture approximate message passing,” IEEE Trans. Signal Process., vol. 61, no. 19, pp. 4658–4672, 2013.

[38] K. Liu, Z. Zhang, L. Dai, S. Xu, and F. Yang, “Active reconﬁgurable intelligent surface: Fully-connected or sub-connected?” IEEE Commun. Lett., vol. 26, no. 1, pp. 167–171, Jan. 2022.

[39] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best K-term approximation,” J. Amer. Math. Soc., vol. 22, no. 1, pp. 211–231, 2009.

[40] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2011, pp. 2168–2172.

[41] M. R. Akdeniz et al., “Millimeter wave channel modeling and cellular capacity evaluation,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1164–1179, Apr. 2014.

[42] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Simultaneous sparse approximation via greedy pursuit,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Mar. 2005, pp. v721–v724.
Wanting Lyu received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China in 2021, where she is currently pursuing the Ph.D. degree with the National Key Laboratory of Science and Technology on Communications. Her research and study interests are in wireless communications, reconfigurable intelligent surfaces, age of information, cell-free communications, and convex programming.

Zhongpei Zhang (Member, IEEE) received the B.S. and M.S. degrees from the Department of Physics, Sichuan Normal University, in 1990 and 1993, respectively, and the Ph.D. degree from the School of Computer and Communication Engineering, Southwest Jiaotong University, in 2000. From 2001 to 2003, he was a Post-Doctoral Fellow with the National Key Laboratory of Microwave and Digital Communication, Tsinghua University; and the University of Oulu, from 2004 to 2005. He is currently a Professor and a Doctoral Tutor with the University of Electronic Science and Technology of China; and the Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, respectively. He has participated in many research projects. He has authored or coauthored more than 90 journal articles and conference papers. His research interests include channel coding, coordinated multiple points transmission, information theory, channel estimation based on compressive sensing, reconfigurable-reflecting-surface aided communications, and joint sensing and communication. He has chaired the National High-Tech Research and Development Program of China (863 Program) on Coordinated Multiple Points Transmission and the National Natural Science Foundation of China on Massive MIMO Channel Acquisition.

Yue Xiu received the master's and Ph.D. degrees from the University of Electronic Science and Technology of China, where he is currently pursuing the second Ph.D. degree with the National Key Laboratory of Science and Technology on Communications. In 2020, he became a Visiting Ph.D. Student with Nanyang Technology University (NTU). He is also a Communication Algorithm Engineer with the Department of Wireless Communication, Huawei, Chengdu, China. His research interests include convex optimization, manifold optimization, differential geometry, and wireless communication.

Chau Yuen (Fellow, IEEE) received the B.Eng. and Ph.D. degrees from Nanyang Technological University, Singapore, in 2000 and 2004, respectively. He was a Post-Doctoral Fellow with Lucent Technologies Bell Labs, Murray Hill, NJ, USA, in 2005; and a Visiting Assistant Professor with The Hong Kong Polytechnic University, in 2008. From 2006 to 2010, he was with the Institute for Infocomm Research, Singapore. From 2010 to 2023, he was a Associate Professor with the Engineering Product Development Pillar, Singapore University of Technology and Design. Since 2023, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University. He has three U.S. patents and published over 500 research articles in international journals or conferences.

Dr. Yuen was a recipient of the Lee Kuan Yew Gold Medal, the Institution of Electrical Engineers Book Prize, the Institute of Engineering of Singapore Gold Medal, the Merck Sharp and Dohme Gold Medal, and the Hewlett Packard Prize (twice). He received the IEEE Asia–Pacific Outstanding Young Researcher Award in 2012 and the IEEE VTS Singapore Chapter Outstanding Service Award in 2019. He was awarded as the Excellent Editor Award for IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING and the Top Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, from 2009 to 2015. He currently serves as an Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE SYSTEMS JOURNAL, and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING. He also served as the Guest Editor for several special issues, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE Wireless Communications Magazine, IEEE Communications Magazine, IEEE Vehicular Technology Magazine, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and Applied Energy (Elsevier). He is a Distinguished Lecturer of IEEE Vehicular Technology Society and also a Highly Cited Researcher by Clarivate Web of Science.