DOI: 10.1109/TVT.2023.3309950

Beam Pattern and Reflection Pattern Design for Channel Estimation in RIS-assisted mmWave MIMO Systems

You You, *Member, IEEE*, Wuqiong Zhao, *Student Member, IEEE*, Li Zhang, *Senior Member, IEEE*, Xiaohu You, *Fellow, IEEE*, and Chuan Zhang, *Senior Member, IEEE*

***Abstract*—Reconfigurable intelligent surface (RIS) is a revolutionary technology that can be applied in millimeter wave (mmWave) communications to reduce the high power consumption and propagation loss. However, channel estimation (CE) is challenging due to the large number of passive RIS elements without signal processing abilities. In this paper, the uplink CE for RIS-assisted mmWave multi-input multi-output (MIMO) systems is formulated as a sparse signal recovery problem in a novel way. Then, the beam pattern and reflection pattern design based on the compressed sensing (CS) theory are proposed to guarantee the efficient CE. Simulation results demonstrate that, for various CS-based CE algorithms, the proposed patterns can reduce more than 50% pilot overhead at 0 dB signal-to-noise ratio (SNR) while maintaining the same accuracy of CE compared with the existing patterns.**

***Index Terms*—Reconfigurable intelligent surface (RIS), channel estimation (CE), pilot beam pattern design, reflection design.**

I. Introduction

INTELLIGENT reflecting surface also referred to as reconfigurable intelligent surface (RIS) is a kind of electromagnetic metasurfaces consisting of passive reflecting units. It has been regarded as an appealing solution to the high hardware cost and power consumption problem in millimeter wave (mmWave) multi-input multi-output (MIMO) systems [1]. However, channel estimation (CE) for RIS-assisted mmWave MIMO systems poses a challenge due to the large number of passive elements that lack signal processing abilities.

Recently, there have been some works investigating the CE problem for the fully passive RIS system. A binary-reflection controlled least square (LS) CE protocol was proposed in [2] at cost of intractable computational complexity. To reduce the pilot overhead and the computational complexity, the compressive sensing (CS) algorithms are utilized in [3]–[5] for multi-input single-output (MISO) systems. However, the beam pattern and reflection pattern design have great impact on the performance of the CS-based CE methods and only few works take them into consideration.

To achieve better CS-based CE accuracy, the beam pattern and reflection pattern need to be designed. By assuming that the position of RIS is known to the BS, the pilot signals and reflection patterns are designed in [6], but only considering the single antenna at the receiver. For MISO systems, an optimized channel estimator in a closed form is proposed by leveraging the typical mean-squared error (MSE) criterion in [7]. However, the existing works do not take into account the beam pattern and reflection pattern design simultaneously and do not focus on the CS-based CE problem in MIMO systems.

In this paper, we consider both the beam pattern and reflection pattern design for RIS-assisted mmWave MIMO systems. The contributions are summarized as follows:

1) We propose a novel channel estimation formulation to significantly reduce the computational complexity of CE for RIS-assisted MIMO system.
2) The reflection pattern is designed by ensuring the sparsity of the cascaded channel and the beam pattern is designed by minimizing the coherence of the sensing matrix based on the CS theory.
3) Various CS-based algorithms are utilized to verify the superiority of the proposed scheme. And the benefit of the proposed scheme is especially evident for CS-based greedy algorithms.

Notations: In this paper, lower-case and upper-case boldface letter $\mathbf{x}$ and $\mathbf{X}$ denote a vector and a matrix respectively. $[\mathbf{x}]_i$ denotes the i-th element of vector $\mathbf{x}$. $\mathbf{X}^{\mathsf{T}}$, $\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}^*$ denote the transpose, the conjugate transpose, the conjugate. The ℓ_2-norm of vector $\mathbf{x}$ is indicated by $\|\mathbf{x}\|_2$, while the Frobenius norm of matrix $\mathbf{X}$ is denoted as $\|\mathbf{X}\|_F$. The diagonal matrix having vector $\mathbf{x}$ on its diagonal is denoted as $\mathrm{diag}(\mathbf{x})$. Vectorization of matrix $\mathbf{X}$ is given by $\mathrm{vec}(\mathbf{X})$, and the inverse operation, reshaping vector $\mathbf{x}$ into a $p \times q$ matrix, is represented by $\mathrm{vec}_{p,q}^{-1}(\mathbf{x})$. $\mathrm{tr}(\mathbf{X})$ denotes the trace of $\mathbf{X}$. The Kronecker product is denoted as $\mathbf{A} \otimes \mathbf{B}$. $\mathbf{I}_p$ is the identity matrix of size $p \times p$ and $\mathbf{O}_{p\times q}$ is a $p \times q$ matrix with all elements being zero. $\mathrm{mod}(a, b)$ calculates the remainder of a being divided by b.

This work was supported in part by National Key R&D Program of China under Grant 2020YFB2205503, in part by CPSF under Grants 2022M722433, in part by NSFC under Grants 62122020 and 61871115, in part by the Jiangsu Provincial NSF under Grant BK20211512, in part by the Major Key Project of PCL under Grant PCL2021A01-2, in part by Jiangsu Excellent Postdoctoral Program, and in part by the Fundamental Research Funds for the Central Universities. *(You You and Wuqiong Zhao contribute equally to this paper.) (Corresponding author: Chuan Zhang.)*

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

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

II. System Model

A. Cascaded Channel Model

In a single-user MIMO system with the carrier wavelength λ, the base station (BS) and the user are equipped with the uniform linear array (ULA) with N_r and N_t antennas, respectively, and the RIS has $M = M_x \times M_y$ elements in a uniform planar array (UPA). Following [3], a geometric channel model for narrowband systems is used to formulate the RIS–BS channel and user–RIS channel. These two channels are low-rank due to the sparse scattering nature of mmWave, i.e. L_1 and L_2 are small, where they denote the numbers of RIS–BS and user–BS paths, respectively. Specifically, the RIS–BS channel $\mathbf{G}$ can be modeled as

$$\mathbf{G} = \sqrt{\frac{MN_r}{L_1}} \sum_{l_1=1}^{L_1} \alpha_{l_1} \mathbf{a}_r \left(\theta_{l_1}^{G_{N_r}}\right) \mathbf{a}_f \left(\vartheta_{l_1}^{G_M}, \varphi_{l_1}^{G_M}\right)^{\mathsf{H}}, \quad (1)$$

where α_{l_1} is the complex gain consisting of path loss, $\theta_{l_1}^{G_{N_r}}$ is the angle of arrival (AoA) at the BS, and $\vartheta_{l_1}^{G_M}$ and $\varphi_{l_1}^{G_M}$ represent the azimuth and elevation angle of departure (AoD) at the RIS, respectively. $\mathbf{a}_r(\vartheta, \varphi) \in \mathbb{C}^{N_r \times 1}$ and $\mathbf{a}_f(\vartheta, \varphi) \in \mathbb{C}^{M \times 1}$ represent the normalized array steering vectors at the BS and the RIS. Similarly, the user–RIS channel $\mathbf{R}$ is represented as

$$\mathbf{R} = \sqrt{\frac{MN_t}{L_2}} \sum_{l_2=1}^{L_2} \alpha_{l_2} \mathbf{a}_f \left(\vartheta_{l_2}^{R_M}, \varphi_{l_2}^{R_M}\right) \mathbf{a}_u \left(\theta_{l_2}^{R_{N_t}}\right)^{\mathsf{H}}, \quad (2)$$

where α_{l_2} is the complex gain consisting of path loss, $\theta_{l_2}^{R_{N_t}}$ is the AoD at the user, and $\vartheta_{l_2}^{R_M}$ and $\varphi_{l_2}^{R_M}$ represent the azimuth and elevation AoA at the RIS. $\mathbf{a}_u(\theta) \in \mathbb{C}^{N_t \times 1}$ represents the normalized array steering vector at the user. Normalized array steering vectors for X-element UPA ($X = X_1 \times X_2$) and ULA can be formulated according to [3] as

$$\begin{cases} \mathbf{a}(\vartheta, \varphi) = \frac{1}{\sqrt{X}} \left(e^{-i \cdot \kappa \sin(\vartheta)\cos(\varphi)\mathbf{x}_1/\lambda}\right) \otimes \left(e^{-i \cdot \kappa \sin(\varphi)\mathbf{x}_2/\lambda}\right), \\ \mathbf{a}(\theta) = \frac{1}{\sqrt{X}} e^{-i \cdot \kappa \cos(\theta)\mathbf{x}/\lambda}, \end{cases} \quad (3)$$

where $\kappa = 2\pi d$, $\mathbf{x}_1 = [0, 1, 2, \ldots, X_1 - 1]^{\mathsf{T}}$, $\mathbf{x}_2 = [0, 1, 2, \ldots, X_2 - 1]^{\mathsf{T}}$, $\mathbf{x} = [0, 1, 2, \ldots, X - 1]^{\mathsf{T}}$, and d is the antenna spacing which is assumed to be $\lambda/2$ in this paper. We can further virtually represent the two channels as

$$\mathbf{G} = \mathbf{V}_{N_r} \mathbf{\Gamma} \mathbf{V}_M^{\mathsf{H}}, \quad \mathbf{R} = \mathbf{V}_M \mathbf{\Sigma} \mathbf{V}_{N_t}^{\mathsf{H}}, \quad (4)$$

where the beamspace matrices, denoted as $\mathbf{V}_{N_r} \in \mathbb{C}^{N_r \times N_r^G}$, $\mathbf{V}_M \in \mathbb{C}^{M \times M^G}$ and $\mathbf{V}_{N_t} \in \mathbb{C}^{N_t \times N_t^G}$, are composed of N_r^G, $M^G = M_x^G \times M_y^G$ and N_t^G steering vectors of predetermined grids at the BS, RIS and user. $\mathbf{\Gamma} \in \mathbb{C}^{N_r^G \times M^G}$ and $\mathbf{\Sigma} \in \mathbb{C}^{M^G \times N_t^G}$ are the L_1 and L_2-sparse beamspace channels corresponding to $\mathbf{G}$ and $\mathbf{R}$, respectively. For simplicity, we consider $M_x^G = M_x, M_y^G = M_y$ in this work, but it can be extended to general cases easily. Considering a uniform grid spanning from -1 to 1 for all spatial angles, we define the uplink cascaded channel as $\mathbf{H} \triangleq \mathbf{G}\mathrm{diag}(\mathbf{\Psi})\mathbf{R}$, where $\mathbf{\Psi} \in \mathbb{C}^{M \times 1}$ is the phase shift vector at the RIS assuming each RIS element does not change the signal amplitude. Thus, we can obtain

$$\mathbf{H} = \mathbf{V}_{N_r} \mathbf{\Gamma} \mathbf{V}_M^{\mathsf{H}} \mathrm{diag}(\mathbf{\Psi}) \mathbf{V}_M \mathbf{\Sigma} \mathbf{V}_{N_t}^{\mathsf{H}} = \mathbf{V}_{N_r} \mathbf{\Gamma} \mathbf{\Omega} \mathbf{\Sigma} \mathbf{V}_{N_t}^{\mathsf{H}}, \quad (5)$$

where $\mathbf{\Omega} \triangleq \mathbf{V}_M^{\mathsf{H}} \mathrm{diag}(\mathbf{\Psi}) \mathbf{V}_M$.

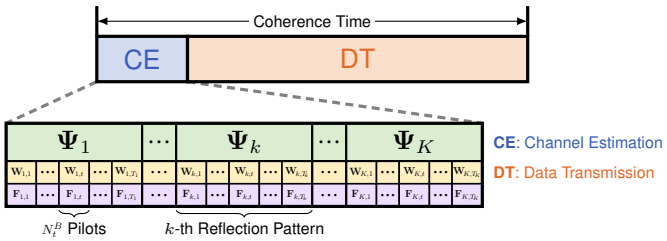


Fig. 1. Sounding procedure. [8]

B. Sounding Procedure

The CE scheme is shown as Fig. 1 similar to [8]. K different RIS reflection vector designs $\mathbf{\Xi} \triangleq [\mathbf{\Psi}_1, \mathbf{\Psi}_2, \ldots, \mathbf{\Psi}_K]$ ($k = 1, 2, \ldots, K$) are considered within the CE stage in one channel coherence time. T_k pilot blocks are transmitted for one reflection vector design $\mathbf{\Psi}_k$, and the corresponding cascaded channel is $\mathbf{H}_k \triangleq \mathbf{G}\mathrm{diag}(\mathbf{\Psi}_k)\mathbf{R}$. For the t-th ($t = 1, 2, \ldots, T_k$) pilot block, N_r^B and N_t^B beams are designed at the BS and the user respectively ($N_r^B < N_r$, $N_t^B < N_t$), so in one pilot block N_t^B pilots are transmitted. The beam pattern design for the t-th pilot block within the k-th reflection design is represented by precoding $\mathbf{F}_{k,t} \in \mathbb{C}^{N_t \times N_t^B}$ and combining $\mathbf{W}_{k,t} \in \mathbb{C}^{N_r \times N_r^B}$. Given the user's p-th ($p = 1, 2, \ldots, N_t^B$) transmitted beam $\mathbf{f}_{k,t,p}$ for the t-th pilot block, the received signal $\mathbf{y}_{k,t,p}$ can be described as

$$\mathbf{y}_{k,t,p} = \mathbf{W}_{k,t}^{\mathsf{H}} \mathbf{H}_k \mathbf{f}_{k,t,p} s_{k,t,p} + \mathbf{W}_k^{\mathsf{H}} \mathbf{n}_{k,t,p}, \quad (6)$$

where $s_{k,t,p}$ is the transmitted pilot signal with $|s_{k,t,p}| = 1$, $\mathbf{n}_{k,t,p} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{N_r^B})$ is the white Gaussian noise with mean μ and variance σ^2. Collect all N_t^B transmitted pilots within the pilot block as

$$\mathbf{Y}_{k,t} = \mathbf{W}_{k,t}^{\mathsf{H}} \mathbf{H}_k \mathbf{F}_{k,t} + \mathbf{N}_{k,t}, \quad (7)$$

where matrix $\mathbf{Y}_{k,t} \triangleq [\mathbf{y}_{k,t,1}, \mathbf{y}_{k,t,2}, \ldots, \mathbf{y}_{k,t,N_t^B}]$, $\mathbf{F}_{k,t} \triangleq [\mathbf{f}_{k,t,1}, \mathbf{f}_{k,t,2}, \ldots, \mathbf{f}_{k,t,N_t^B}]$, and the noise matrix $\mathbf{N}_{k,t} \triangleq [\mathbf{W}_{k,t}^{\mathsf{H}} \mathbf{n}_{k,t,1}, \ldots, \mathbf{W}_{k,t}^{\mathsf{H}} \mathbf{n}_{k,t,N_t^B}]$. The measurement number for the k-th reflection vector design is $Q_k = T_k N_t^B N_r^B$.

C. Problem Formulation

The cascaded channel with $\mathbf{\Psi}_k$ can be represented as

$$\begin{aligned} \mathrm{vec}(\mathbf{H}_k) &= (\mathbf{V}_{N_t}^* \otimes \mathbf{V}_{N_r}) \mathrm{vec}(\mathbf{\Gamma} \mathbf{\Omega}_k \mathbf{\Sigma}) \\ &\overset{(\star)}{=} (\mathbf{V}_{N_t}^* \otimes \mathbf{V}_{N_r})(\mathbf{\Sigma}^{\mathsf{T}} \otimes \mathbf{\Gamma}) \mathrm{vec}(\mathbf{\Omega}_k) \\ &= (\mathbf{V}_{N_t}^* \otimes \mathbf{V}_{N_r})(\mathbf{\Sigma}^{\mathsf{T}} \otimes \mathbf{\Gamma})(\mathbf{V}_M^{\mathsf{T}} \circledcirc \mathbf{V}_M^{\mathsf{H}}) \mathbf{\Psi}_k \\ &= (\mathbf{V}_{N_t}^* \otimes \mathbf{V}_{N_r}) \mathbf{J} \mathbf{D} \mathbf{\Psi}_k, \end{aligned} \quad (8)$$

where in $(\star)$ the mix-product property of Kronecker product is used. $\mathbf{J} \triangleq \mathbf{\Sigma}^{\mathsf{T}} \otimes \mathbf{\Gamma}$, and $\mathbf{D} \triangleq \mathbf{V}_M^{\mathsf{T}} \circledcirc \mathbf{V}_M^{\mathsf{H}}$ is the Khatri-Rao product of $\mathbf{V}_M^{\mathsf{T}}$ and $\mathbf{V}_M^{\mathsf{H}}$. (8) can be simplified as

$$\mathrm{vec}(\mathbf{H}_k) = (\mathbf{V}_{N_t}^* \otimes \mathbf{V}_{N_r}) \widetilde{\mathbf{J}} \widetilde{\mathbf{D}} \mathbf{\Psi}_k, \quad (9)$$

where $\widetilde{\mathbf{D}}$ corresponds to the initial M^G rows of $\mathbf{D}$, and $\widetilde{\mathbf{J}}$ is a merged form of $\mathbf{J}$ [3]:

$$\widetilde{\mathbf{J}}(:, i) = \sum_{n \in \mathcal{S}_i} \mathbf{J}(:, n), \quad (10)$$

where $\mathcal{S}_i$ designates the indices in $\mathbf{D}$ sharing the same row as the i-th row. Utilizing (9), the received signal in (7) can be further represented in the vectorized form as

$$\begin{aligned}\mathbf{y}_{k,t} &\overset{(a)}{=} (\mathbf{F}_{k,t}^{\mathsf{T}} \otimes \mathbf{W}_{k,t}^{\mathsf{H}})\mathrm{vec}(\mathbf{H}_k) + \mathbf{n}_{k,t} \\ &\overset{(b)}{=} (\mathbf{F}_{k,t}^{\mathsf{T}} \otimes \mathbf{W}_{k,t}^{\mathsf{H}})(\mathbf{V}_{N_t}^{*} \otimes \mathbf{V}_{N_r})\widetilde{\mathbf{J}}\widetilde{\mathbf{D}}\boldsymbol{\Psi}_k + \mathbf{n}_{k,t} \\ &\overset{(c)}{=} \left((\widetilde{\mathbf{D}}\boldsymbol{\Psi}_k)^{\mathsf{T}} \otimes \left((\mathbf{F}_{k,t}^{\mathsf{T}} \otimes \mathbf{W}_{k,t}^{\mathsf{H}})(\mathbf{V}_{N_t}^{*} \otimes \mathbf{V}_{N_r})\right)\right)\tilde{\mathbf{j}} + \mathbf{n}_{k,t},\end{aligned} \tag{11}$$

where (a) and (c) make use of the mixed-product property inherent to the Kronecker product with $\mathbf{y}_{k,t} \triangleq \mathrm{vec}(\mathbf{Y}_{k,t}) \in \mathbb{C}^{N_t^B N_r^B \times 1}$, $\mathbf{n}_{k,t} \triangleq \mathrm{vec}(\mathbf{N}_{k,t})$, $\tilde{\mathbf{j}} = \mathrm{vec}(\widetilde{\mathbf{J}}) \in \mathbb{C}^{N_t^G N_r^G M^G \times 1}$ as in [3]. (b) uses the result from (9). Given the sparse nature of $\tilde{\mathbf{j}}$, it is feasible to utilize CS algorithms. However, the huge size of $\tilde{\mathbf{j}}$ still makes it impractical to compute.

III. Proposed Method

A. Improved Channel Estimation Formulation

To reduce the computation complexity due to the huge size of $\tilde{\mathbf{j}}$, we further propose to formulate (11) as

$$\begin{aligned}\mathbf{y}_{k,t} &\overset{(d)}{=} (\mathbf{F}_{k,t}^{\mathsf{T}} \otimes \mathbf{W}_{k,t}^{\mathsf{H}})(\mathbf{V}_{N_t}^{*} \otimes \mathbf{V}_{N_r})\boldsymbol{\lambda}_k + \mathbf{n}_{k,t} \\ &\overset{(e)}{=} \mathbf{Q}_{k,t}\boldsymbol{\lambda}_k + \mathbf{n}_{k,t},\end{aligned} \tag{12}$$

where the deduction of (d) is illustrated in Fig. 2 and the merged $\boldsymbol{\lambda}_k \in \mathbb{C}^{N_t^G N_r^G \times 1}$ can be formulated as

$$[\boldsymbol{\lambda}_k]_i = \sum_{m=1}^{M^G} [(\widetilde{\mathbf{D}}\boldsymbol{\Psi}_k)^{\mathsf{T}}]_m [\tilde{\mathbf{j}}]_{N_t^G N_r^G \cdot (m-1)+i}, \tag{13}$$

for $i = 1, 2, \ldots, N_t^G N_r^G$. Without loss of generality, we assume $T_k = T$ and $Q_k = Q$ for $k = 1, 2, \ldots, K$. In (e), the equation is simplified by defining $\mathbf{Q}_{k,t} \triangleq (\mathbf{F}_{k,t}^{\mathsf{T}} \otimes \mathbf{W}_{k,t}^{\mathsf{H}})(\mathbf{V}_{N_t}^{*} \otimes \mathbf{V}_{N_r})$. Collecting all received signals from the T pilot blocks in (12), we can obtain

$$\mathbf{y}_k = \mathbf{Q}_k\boldsymbol{\lambda}_k + \mathbf{n}_k, \tag{14}$$

where $\mathbf{y}_k \triangleq [\mathbf{y}_{k,1}^{\mathsf{T}}, \ldots, \mathbf{y}_{k,T}^{\mathsf{T}}]^{\mathsf{T}}$, $\mathbf{Q}_k \triangleq [\mathbf{Q}_{k,1}^{\mathsf{T}}, \ldots, \mathbf{Q}_{k,T}^{\mathsf{T}}]^{\mathsf{T}}$ and $\mathbf{n}_k \triangleq [\mathbf{n}_{k,1}^{\mathsf{T}}, \ldots, \mathbf{n}_{k,T}^{\mathsf{T}}]^{\mathsf{T}}$. The problem can be solved by any CS algorithms where $\boldsymbol{\lambda}_k \in \mathbb{C}^{N_t^G N_r^G \times 1}$ is a sparse vector.

After the CE with K reflection vectors, the data transmission is considered within the channel coherence time. With a phase shift vector $\boldsymbol{\Psi}$ that satisfies

$$\widetilde{\mathbf{D}}\boldsymbol{\Psi} = \sum_{k=1}^{K} \beta_k \widetilde{\mathbf{D}}\boldsymbol{\Psi}_k, \tag{15}$$

the channel $\mathbf{H}$ for the data transmission can be formulated as

$$\mathbf{H} = \mathrm{vec}_{N_r,N_t}^{-1}\left((\mathbf{V}_{N_t}^{*} \otimes \mathbf{V}_{N_r})\boldsymbol{\Lambda}\boldsymbol{\beta}\right), \tag{16}$$

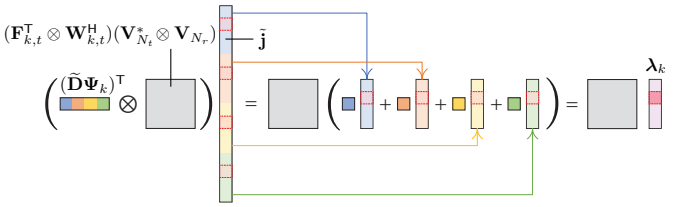


Fig. 2. Merging from $\tilde{\mathbf{j}}$ in Eq. (11) (c) to $\boldsymbol{\lambda}_k$ in Eq. (12) (d) with $M^G = 4$.

where $\boldsymbol{\Lambda} \triangleq [\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_K]$ is to be estimated via CE and coefficients $\boldsymbol{\beta} \triangleq [\beta_1, \beta_2, \ldots, \beta_K]^{\mathsf{T}}$ can be determined with the phase shift vector $\boldsymbol{\Psi}$ from (15). It is worth noting that with K ($K \leq M$) reflection vectors $\boldsymbol{\Psi}_k$ for $k = 1, 2, \ldots, K$, the phase shift vector $\boldsymbol{\Psi}$ used in data transmission can be well represented by considering (15) as a linear mapping.

B. Computational Complexity of CE Formulations

The computational complexity comparison with OMP algorithm is listed in Table I, where we assume $N_t^G = N_t, N_r^G = N_r$. L and $L' = L_1 L_2$ denote the sparsity of $\boldsymbol{\lambda}_k$ and $\widetilde{\mathbf{J}}$ respectively. $L \ll L'$ with our designed reflection patterns, and more details about the sparsity will be discussed in section III-C. With unmerged $\mathbf{J}$ in (8), The OMP algorithm solving (11) has the complexity $\mathcal{O}(L'Q'' \cdot M^2 N_t N_r)$ [9], where Q'' is the required number of measurement. With merged $\widetilde{\mathbf{J}}$ in (11) (c) which is proposed in [3], the complexity is reduced to $\mathcal{O}(L'Q' \cdot MN_t N_r)$, where Q' is the required number of measurement. The complexity of our proposed CE formulation using OMP to solve (12) (d) is $K \cdot \mathcal{O}(LQ \cdot N_t N_r) = \mathcal{O}(KL \cdot QN_t N_r)$. Considering the reduced dimension of the channel matrices to be estimated ($\boldsymbol{\lambda}_k$, $\tilde{\mathbf{j}}$, $\mathbf{j}$), the required number of measurements for OMP algorithm satisfies $Q \ll Q' \ll Q''$. Thus, the complexity of the proposed CE formulation is much lower than existing formulations [3].

TABLE I
Computational Complexity of Formulations with OMP.

Compressed CE Formulation	**Computational Complexity** (OMP)
Eq. (11) (c) with unmerged $\mathbf{J}$	$\mathcal{O}(L'Q'' \cdot M^2 N_t N_r)$
Eq. (11) (c) with merged $\widetilde{\mathbf{J}}$ [3]	$\mathcal{O}(L'Q' \cdot MN_t N_r)$
Eq. (12) (d) (proposed)	$K \cdot \mathcal{O}(LQ \cdot N_t N_r) = \mathcal{O}(LQ \cdot KN_t N_r)$

* $L \ll L', Q \ll Q' \ll Q'', K \leq M$.

C. RIS Reflection Pattern Design

The design of reflection patterns has a great impact on the performance of CE. In general, reflection patterns can be designed to maximize the received power to increase the received SNR or minimize the coherence among different $\boldsymbol{\lambda}_k$. However, different from other formulations such as (11) [3] where the RIS reflection vector is not part of the sparse channel to recover, the sparsity of $\boldsymbol{\Lambda}$ in (16) depends on the reflection pattern design as shown in (13). Obviously, not all $\boldsymbol{\Xi}$ that satisfies (15) will lead to a sparse $\boldsymbol{\Lambda}$. Considering that CS algorithms can only be applied when the sparsity is guaranteed, the aim of the reflection pattern design in this paper is to ensure the sparsity of $\boldsymbol{\Lambda}$. The K reflection patterns $\boldsymbol{\Xi}$ can be designed by choosing K unique columns from $\widehat{\boldsymbol{\Xi}}$:

$$\boldsymbol{\Xi} \subseteq \widehat{\boldsymbol{\Xi}} = \widetilde{\mathbf{D}}^{-1}, \tag{17}$$

where $\widehat{\boldsymbol{\Xi}}$ contains M candidates for the reflection pattern design. It can be readily verified that $\boldsymbol{\Lambda}$ becomes extremely sparse because $\widetilde{\mathbf{D}}\boldsymbol{\Psi}_k$ has only one non-zero element according to (13). For general cases where $M_x^G \neq M_x$ and/or $M_y^G \neq M_y$, to ensure the phase shift property, $\widehat{\boldsymbol{\Xi}} = \frac{1}{M}\max\{M_x^G, M_x\}\max\{M_y^G, M_y\}\widetilde{\mathbf{D}}^{\dagger}$, where $\widetilde{\mathbf{D}}^{\dagger}$ is the pseudo-inverse of $\widetilde{\mathbf{D}}$. When $M_x^G < M_x$, M_x^G needs to be

$$\begin{cases} \mathbf{W}_{k,t} = \left[\mathbf{O}_{N_r^B \cdot \left\lfloor \frac{(t-1)N_t^B}{N_t} \right\rfloor \times N_r^B} \quad \mathbf{I}_{N_r^B} \quad \mathbf{O}_{\left(N_r - N_r^B \cdot \left\lfloor \frac{(t-1)N_t^B}{N_t} \right\rfloor\right) \times N_r^B}\right]^{\mathsf{T}}, \\ \mathbf{F}_{k,t} = \left[\mathbf{O}_{N_t^B \cdot \mathrm{mod}\left(t-1, \frac{N_t}{N_t^B}\right) \times N_t^B} \quad \mathbf{I}_{N_r^B} \quad \mathbf{O}_{\left(N_t - N_t^B \cdot \mathrm{mod}\left(t-1, \frac{N_t}{N_t^B}\right)\right) \times N_t^B}\right]^{\mathsf{T}}. \end{cases} \tag{24}$$

chosen satisfying $\mathrm{mod}(M_x, M_x^G) = 0$, and it is similar for M_y^G. With (17), when $M_x^G = M_x, M_y^G = M_y$ and $K = M$, the estimated channel in (16) can be simplified to

$$\mathbf{H} = \mathrm{vec}_{N_r,N_t}^{-1}\left((\mathbf{V}_{N_t}^* \otimes \mathbf{V}_{N_r})\boldsymbol{\Lambda}\widetilde{\mathbf{D}}\boldsymbol{\Psi}\right), \tag{18}$$

which can also be easily adapted for $K < M$. Other reflection pattern designs such as discrete Fourier transform (DFT) matrices and Hadamard matrices [10] also contribute more or less to the sparsity of $\boldsymbol{\Lambda}$. The performance comparison between the reflection pattern designs will be shown in section IV.

D. Beam Pattern Design

Based on the CS theory, it is known that smaller coherence of the sensing matrix $\mathbf{Q}_k$ improves the performance of the sparse signal recovery [11], [12]. However, due to the high dimension of RIS-assisted MIMO systems, methods such as [13] that considers the RIS-assisted MISO systems can not be applied directly.

With the improved channel estimation formulation in (14), the beam pattern design is based on the minimization of total coherence which is defined as

$$\mu^t(\mathbf{Q}_k) = \sum_{\substack{1 \le q_1, q_2 \le Q, \\ q_1 \ne q_2}} \langle \mathbf{Q}_k(:, q_1), \mathbf{Q}_k(:, q_2) \rangle, \tag{19}$$

which is the summation of inner product of the q_1-th column and the q_2-th column. It is easy to prove $\|\mathbf{Z}_{k,t}(:, n)\|_2 = 1$ by assuming equal power of each beam for $n = 1, 2, \ldots, Q$, where $\mathbf{Z}_{k,t} \triangleq \mathbf{F}_{k,t} \otimes \mathbf{W}_{k,t}^*$ is the Kronecker product of the transmit and receive beamforming. Therefore, (19) can be written as

$$\begin{aligned} \mu^t(\mathbf{Q}_k) &= \|\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k - \mathbf{I}_G\|_F^2 \\ &= \mathrm{tr}\left(\left(\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k - \mathbf{I}_G\right)^2\right) \\ &= \mathrm{tr}\left(\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k - 2\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k + \mathbf{I}_G\right) \\ &= \mathrm{tr}\left(\mathbf{Q}_k\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k\mathbf{Q}_k^{\mathsf{H}} - 2\mathbf{Q}_k\mathbf{Q}_k^{\mathsf{H}} + \mathbf{I}_Q\right) + (G - Q) \\ &= \|\mathbf{Q}_k\mathbf{Q}_k^{\mathsf{H}} - \mathbf{I}_Q\|_F^2 + (G - Q) \\ &= \|\mathbf{Z}_k^{\mathsf{T}}(\mathbf{V}_N\mathbf{V}_N^{\mathsf{H}})\mathbf{Z}_k^* - \mathbf{I}_Q\|_F^2 + (G - Q), \end{aligned} \tag{20}$$

where $\mathbf{Z}_k \triangleq [\mathbf{Z}_{k,1}, \mathbf{Z}_{k,2}, \ldots, \mathbf{Z}_{k,T}]$, $\mathbf{Q}_k = \mathbf{Z}_k^{\mathsf{T}}\mathbf{V}_N$, $G \triangleq N_t^G N_r^G$. It can be found that $\mu^t(\mathbf{Q}_k) \ge (G - Q)$, and when $\mathbf{Z}_k^{\mathsf{T}}(\mathbf{V}_N\mathbf{V}_N^{\mathsf{H}})\mathbf{Z}_k^* - \mathbf{I}_Q = \mathbf{O}_{Q\times Q}$, the total coherence in $\mathbf{Z}_k^{\mathsf{T}}\mathbf{V}_N$ is minimized. $\mathbf{V}_N \triangleq \mathbf{V}_{N_t}^* \otimes \mathbf{V}_{N_r}$, where $\mathbf{V}_{N_t}$ and $\mathbf{V}_{N_r}$ are defined in (4) and

$$\mathbf{V}_N\mathbf{V}_N^{\mathsf{H}} = \frac{N_t^G N_r^G}{N_t N_r}\mathbf{I}_{N_t N_r}. \tag{21}$$

Substituting (21), (20) can be converted to

$$\begin{aligned} &\min_{\mathbf{Z}_k} \left\| \frac{N_t^G N_r^G}{N_t N_r}\mathbf{Z}_k^{\mathsf{T}}\mathbf{Z}_k^* - \mathbf{I}_Q \right\|_F^2, \\ &\text{s.t. } \|\mathbf{Z}_k(:, n)\|_2^2 = 1, \quad n = 1, 2, \ldots, Q. \end{aligned} \tag{22}$$

So far, the singular value decomposition (SVD) can be employed to find the optimal $\mathbf{Z}_k$ as the following proposition.

Proposition 1. *When $Q \le N_t N_r$, (22) is optimized when*

$$\mathbf{Z}_k = \mathbf{U}_1\left[\mathbf{I}_Q \quad \mathbf{O}_{Q\times(N_tN_r-Q)}\right]^{\mathsf{T}}\mathbf{U}_2^{\mathsf{H}}, \tag{23}$$

where $\mathbf{U}_1 \in \mathbb{C}^{N_tN_r\times N_tN_r}$, $\mathbf{U}_2 \in \mathbb{C}^{Q\times Q}$ are both unitary matrices.

Proposition 1 can be easily proved. With the design of $\mathbf{Z}_k$, beam pattern matrices $\mathbf{W}_{k,t}$ and $\mathbf{F}_{k,t}$ can be obtained based on approximation from $\mathbf{Z}_{k,t} = \mathbf{F}_{k,t} \otimes \mathbf{W}_{k,t}^*$ [14]. As one possible exact solution, suppose $\mathrm{mod}(N_t, N_t^B) = \mathrm{mod}(N_r, N_r^B) = 0$, $\mathbf{W}_{k,t}$ and $\mathbf{F}_{k,t}$ can be designed as (24), where $\lfloor x \rfloor$ denotes the floor function returning the greatest integer less than or equal to x. Notably, hybrid beamfoming technology also can be employed to reduce radio frequency (RF) chains as well as the power consumption. With the design in (24), it is easy to verify that

$$\mathbf{Z}_k(p, q) = \begin{cases} 1, & p = a_q N_r N_t^B + b_q(N_r + N_r^B) + c_q, \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

where $a_q = \lfloor (q-1)/N_tN_r \rfloor$, $b_q = \lfloor (q - 1 - a_qN_tN_r)/N_r^B \rfloor$, $c_q = 1 + \mathrm{mod}(q - 1 - a_qN_tN_r, N_r^B)$ satisfies Proposition 1.

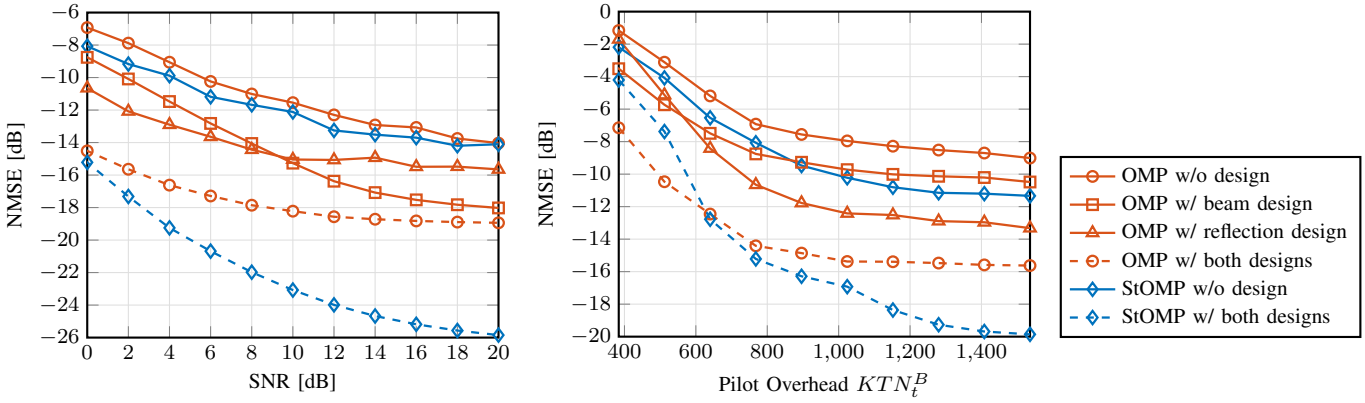
IV. Simulation Results

In this section, we evaluate the performance of CE with the proposed designs in (17) and (24) via computer simulations. Simulation uses C++ with Armadillo linear algebra library. The system parameters are listed in Table II.

TABLE II
SYSTEM PARAMETERS IN SIMULATIONS.

Parameters	Values
(N_t, N_t^G, N_t^B), (N_r, N_r^G, N_r^B)	$(8, 8, 2)$, $(16, 16, 2)$
$M_x \times M_y = M$, $M_x^G \times M_y^G = M^G$	$8 \times 8 = 64$, $8 \times 8 = 64$
(L_1, L_2), K	$(5, 5)$, 64

In Fig. 3, the proposed beam pattern and reflection pattern design are compared with the conventional methods (i.e. random beam patterns and truncated DFT reflection pattern design [10]). Greedy CS algorithms such as OMP and stagewise OMP (StOMP) [15] are used to evaluate the performance. (In StOMP, the maximum number of support selections in each iteration is constrained to enhance stability.) The normalized mean square error (NMSE) performance is defined as $\mathbb{E}[\|\widehat{\mathbf{H}} - \mathbf{H}\|_F^2 / \|\mathbf{H}\|_F^2]$, where $\mathbf{H}$ and $\widehat{\mathbf{H}}$ represent the real channel and the estimated channel respectively. It is worth noting that CS algorithms with random reflection patterns only achieve $0\,\mathrm{dB}$ NMSE performance and are therefore omitted in the comparisons.



(a) NMSE v.s. SNR, with $KTN_t^B = 768$. (b) NMSE v.s. pilot overhead, with $\mathrm{SNR} = 0\,\mathrm{dB}$.

Fig. 3. Comparison of proposed pattern designs and the traditional method with OMP and StOMP.

Fig. 3(a) depicts the NMSE performance versus the SNR with $KTN_t^B = 768$ pilots (LS needs at least $KN_tN_r/N_r^B = 4096$). For OMP algorithm, the proposed RIS reflection pattern design outperforms the widely adopted DFT reflection pattern by around $2\,\mathrm{dB}$. When the beam pattern design is applied with the reflection pattern design, the performance can be further enhanced by more than $3\,\mathrm{dB}$. In other words, compared with existing pattern designs, more than $5\,\mathrm{dB}$ improvement can be achieved with our proposed patterns.

Fig. 3(b) shows the NMSE performance versus the varying pilot overhead KTN_t^B from 384 to 1536. It can be observed that, for OMP, in order to achieve the same NMSE performance such as $-8\,\mathrm{dB}$, compared with the DFT reflection pattern, the proposed reflection pattern design without and with proposed beam pattern design can save about 35% and 55% of pilot overhead respectively.

StOMP algorithm also achieves desirable improvement in both Fig. 3(a) and Fig. 3(b). The proposed beam pattern and reflection design are based on the CS theory instead of a specific CS algorithm. Thus, the works in this paper can be employed in other CS algorithms, especially for the greedy CS algorithms. Notably, our proposed patterns still work for the non-greedy algorithms including ℓ_1-minimization based algorithms and sparse Bayesian learning (SBL) [16]. Since the NMSE performance of these algorithms has a smaller gap to the lower bound, the improvement is not as large as greedy CS algorithms. Considering their unaffordable high computational complexity, these algorithms are not practical in real systems and are therefore not detailed in simulation.

V. Conclusion

In this paper, we propose a novel CE formulation with the beam pattern and reflection pattern design for the CE in RIS-assisted mmWave MIMO systems based on the CS theory. Simulation demonstrates that the proposed pattern designs can reduce more than 50% pilot overhead to achieve the same CE accuracy and improve more than $5\,\mathrm{dB}$ CE performance with the same pilot overhead. The benefit of the proposed scheme is specifically evident for CS-based algorithms.

References

[1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[2] 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)*, May 2019, pp. 4659–4663.

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

[4] Y. You, L. Zhang, M. Yang *et al.*, "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.

[5] 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.

[6] Z. Wan, Z. Gao, and M.-S. Alouini, "Broadband channel estimation for intelligent reflecting surface aided mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[7] W. Zhang, J. Xu, W. Xu *et al.*, "Cascaded channel estimation for IRS-assisted mmWave multi-antenna with quantized beamforming," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 593–597, Feb. 2021.

[8] 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, Sep. 2021.

[9] 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.

[10] Z. Zhou, N. Ge, Z. Wang *et al.*, "Joint transmit precoding and reconfigurable intelligent surface phase adjustment: A decomposition-aided channel estimation approach," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1228–1243, Feb. 2021.

[11] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.

[12] 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.

[13] J. Chen, Y.-C. Liang, H. V. Cheng *et al.*, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems," 2019, *arXiv:1912.03619*. [Online]. Available: https://arxiv.org/abs/1912.03619

[14] C. F. Van Loan and N. Pitsianis, "Approximation with Kronecker products," in *Linear Algebra for Large Scale and Real-Time Applications*, M. S. Moonen, G. H. Golub, and B. L. R. Moor, Eds. Leuven, Belgium: Springer, 1993, ch. 17, pp. 293–314.

[15] D. L. Donoho, Y. Tsaig, I. Drori *et al.*, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[16] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Jun. 2001.