

Joint Channel Estimation and Beamforming for Reconfigurable Intelligent Surface Aided MIMO Systems: Sparsity-Based Approach

Sung Hyuck Hong¹ and Junil Choi²

¹Algorithm Team, FuriosaAI, Inc., Seoul, Republic of Korea, 06036

²School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea, 34141
Email: sunghyuck.hong@furiosa.ai, junil@kaist.ac.kr

Abstract—Continuous efforts have been devoted to integrate millimeter wave (mmWave) and terahertz (THz) bands into future communication standards in order to overcome the bandwidth shortage problem and achieve high data rates, primarily through developing accompanying technologies that can overcome the severe propagation loss and blockage associated with increased carrier frequency. One of the most notable accompanying technologies is reconfigurable intelligent surface (RIS), which uses a large number of low-cost passive reflecting elements to reconfigure the propagation environments for improved communication performance and coverage. Despite its numerous benefits, RIS can make channel estimation more difficult due to its lack of radio frequency (RF) chains that can perform baseband signal processing. In addition, the cascaded channel structure of RIS-aided communication systems, which differs from that in conventional systems, brings about significant challenges in both channel estimation and beamforming. In this paper, we propose the joint channel estimation and beamforming optimization algorithm for RIS-aided multiple-input multiple-output (MIMO) communication systems. By carefully exploiting the angular sparsity of mmWave/THz channels, our proposed algorithm successfully designs the RIS matrices that not only facilitate the channel estimation process but also achieve the passive beamforming gain through increased channel capacity. Simulation results demonstrate that our proposed algorithm provides the systems of interest with significant improvement in spectral efficiency.

Index Terms—Millimeter wave (mmWave) communications, terahertz (THz) communications, reconfigurable intelligent surface (RIS), multiple-input multiple-output (MIMO), channel sparsity.

I. INTRODUCTION

Recent wireless communication systems employ relatively high frequency bands in order to meet the ever-increasing mobile traffic demand and resolve the issue of bandwidth shortage [?]. Thanks to its potential of achieving extremely

This work was done when Sung Hyuck Hong was affiliated with Korea Advanced Institute of Science and Technology. Current affiliation: FuriosaAI.

This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) - Information Technology Research Center (ITRC) grant funded by the Ministry of Science and ICT (MSIT) of the Korea government (IITP-2025-RS-2020-II201787, contribution rate 50%) and in part by IITP under 6G Cloud Research and Education Open Hub grant funded by MSIT. (IITP-2025-RS-2024-00428780, contribution rate 50%)

high data rates and enabling efficient packing of large antenna elements into small physical dimension, millimeter wave (mmWave) communication has been adopted as a pivotal component of the fifth generation (5G) wireless network [?]. The trend of exploiting high frequency bands is expected to become more apparent with the arise of terahertz (THz) communication, which offers up to hundreds of gigahertz (GHz) of contiguous bandwidth and thus is expected to bring significant improvement in the performance of future wireless communication systems [?]. Both mmWave and THz channels exhibit high sparsity in the angular domain as a result of the reduced diffraction and increased specular propagation, thereby inspiring numerous research efforts on communication systems with inherent channel sparsity [?], [?].

In order to successfully realize the beamforming gains in communication systems with multiple antennas, channel matrices must be accurately estimated prior to precoder and combiner design [?]. However, channel state information (CSI) acquisition would become even more challenging in future communication systems that are expected to employ a large number of antennas to overcome the severe propagation loss of mmWave and THz carrier signals [?]. While reconfigurable intelligent surface (RIS) has also been proposed as an effective technology to harness the benefits of mmWave/THz communications, it can further complicate the channel estimation process not only because the surface lacks sensing capabilities to explicitly obtain the channel matrices, but also because the channel structure in RIS-aided systems differs from that in conventional systems [?]. These challenges in channel estimation inevitably degrades the performance of beamforming optimization, which requires accurately estimated channels to successfully achieve its objective of maximizing the spectral efficiency.

Motivated by these facts, we propose in this paper the joint channel estimation and beamforming algorithm for RIS-aided multiple-input multiple-output (MIMO) systems. Our proposed algorithm achieves significant improvement in spectral efficiency through the judicious use of angular sparsity present in mmWave/THz channels. Even in the absence of any iterative optimization procedure and perfect CSI, our proposed algorithm successfully constructs the pilot beamformers that not only simplify the cascaded channel structure of RIS-aided

MIMO systems, but also realize the substantial performance gains that RIS is expected to bring about. By overcoming the limitations of prior works that either (i) require the knowledge of perfect CSI to design high-performing beamformers [?], [?], [?] or (ii) fail to exploit the beamforming capabilities of RIS during the channel estimation process [?], [?], [?], our proposed algorithm is expected to play a pivotal role in enhancing the performance of RIS-aided mmWave/THz MIMO systems.

The remainder of this paper is organized as follows. The system model is described in Section II. The joint approach to channel estimation and beamforming optimization in RIS-aided systems is described in Section III. The proposed sparsity-based algorithm is explained in Section IV. Simulation results and conclusions are presented in Section V and Section VI, respectively.

Notation: Vectors and matrices are represented by lower and upper boldface letters. The transpose, conjugate transpose, inverse, determinant, and rank of the matrix \mathbf{A} are represented by \mathbf{A}^T , \mathbf{A}^H , \mathbf{A}^{-1} , $\det(\mathbf{A})$, and $\text{rank}(\mathbf{A})$, respectively. The Frobenius norm of \mathbf{A} is denoted by $\|\mathbf{A}\|_F$, while $\text{row}(\mathbf{A})$ and $\text{col}(\mathbf{A})$ indicate the number of rows and columns of \mathbf{A} , respectively. The element in the i -th row and j -th column of \mathbf{A} is denoted by $[\mathbf{A}]_{i,j}$, while $\text{diag}(\mathbf{x})$ represents the diagonal matrix that contains the elements of \mathbf{x} on its main diagonal. The $N \times N$ identity matrix is represented by \mathbf{I}_N , and the set of all $m \times n$ complex matrices is denoted by $\mathbb{C}^{m \times n}$. The expectation operator is written as $\mathbb{E}[\cdot]$, while $|\cdot|$ represents the absolute value of a scalar. The function $\max(a, 0)$ defined for a real number a is expressed as $(a)^+$. The function that returns the arguments corresponding to the top- k greatest values is denoted by $\text{argmax}_k(\cdot)$, where k is a positive integer. Finally, the complex and real normal distributions with mean m and variance σ^2 are respectively denoted by $\mathcal{CN}(m, \sigma^2)$ and $\mathcal{N}(m, \sigma^2)$.

II. SYSTEM MODEL

This paper studies an RIS-aided MIMO system in which a transmitter (TX) with N_t antennas communicates with a receiver (RX) with N_r antennas, as shown in Fig. 1. We assume that the direct link between the TX and RX is blocked, a phenomenon that is expected to commonly arise in future communication systems that utilize higher frequency bands [?]. The RIS that consists of M passive reflecting elements facilitates the communication between the TX and RX. The channel from the TX to RIS and that from the RIS to RX are each denoted by $\mathbf{H}_{\text{TS}} \in \mathbb{C}^{M \times N_t}$ and $\mathbf{H}_{\text{SR}} \in \mathbb{C}^{N_r \times M}$. The total cascaded channel from the TX to RX can then be expressed as

$$\mathbf{H}_{\text{tot}} = \mathbf{H}_{\text{SR}} \mathbf{\Phi} \mathbf{H}_{\text{TS}}, \quad (1)$$

where $\mathbf{\Phi} = \text{diag}([e^{j\theta_1}, \dots, e^{j\theta_M}]) \in \mathbb{C}^{M \times M}$ and $\theta_m \in [0, 2\pi)$ each denote the RIS matrix and the phase shift of the m -th RIS element, $m \in \{1, \dots, M\}$. We consider the system of interest to operate in time-division duplexing (TDD) mode, even though the formulation in the paper can be readily

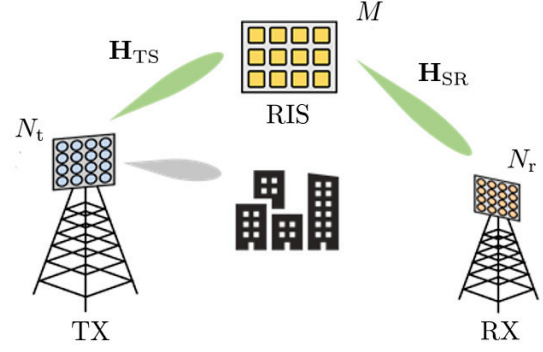


Fig. 1. Block diagram of an RIS-aided MIMO system. The RIS enables the communication between the TX and RX despite the blockage of the direct link (represented by the gray signal) between them.

extended to frequency-division duplexing (FDD) mode by assuming the use of feedback channel [?].

We primarily consider in this paper mmWave and THz frequency bands, both of which are projected to play a pivotal role in future communication standards [?], [?]. To accurately describe mmWave/THz channels, we choose the commonly adopted Saleh-Valenzuela model [?], [?], [?], [?] that expresses each of the channel matrices as

$$\mathbf{H}_i = \sum_{q=0}^{N_{\text{path}}^i - 1} \alpha_{i,q} \mathbf{a}_r(\phi_{i,q}^r, \theta_{i,q}^r) \mathbf{a}_t(\phi_{i,q}^t, \theta_{i,q}^t)^H, \quad (2)$$

where $i \in \{\text{TS}, \text{SR}\}$ and N_{path}^i respectively denote the channel subscript and the number of propagation paths in \mathbf{H}_i . The complex gain $\alpha_{i,q}$ of the q -th path in \mathbf{H}_i is independently distributed with $\mathcal{CN}(0, \gamma_i^2 10^{-0.1 PL(d_i)})$, $\forall q \in \{0, \dots, N_{\text{path}}^i - 1\}$, where $\gamma_i = \sqrt{\text{row}(\mathbf{H}_i) \text{col}(\mathbf{H}_i) / N_{\text{path}}^i}$ represents the normalization factor associated with \mathbf{H}_i . Without loss of generality, we assume that $|\alpha_{i,m}| \geq |\alpha_{i,n}|, \forall m, n \in \{0, \dots, N_{\text{path}}^i - 1\}$ such that $m < n$. As will be explained in Section V, the distance-dependent path loss model $PL(d_i)$ is appropriately adjusted according to the frequency band of interest. The distance between the TX and RIS and that between the RIS and RX are respectively given by d_{TS} and d_{SR} . The azimuth (elevation) angles of arrival and departure (AoAs and AoDs) of the q -th path in \mathbf{H}_i are each given by $\phi_{i,q}^r$ ($\theta_{i,q}^r$) and $\phi_{i,q}^t$ ($\theta_{i,q}^t$), with the corresponding receive and transmit array response vectors denoted as $\mathbf{a}_r(\phi_{i,q}^r, \theta_{i,q}^r) \in \mathbb{C}^{\text{row}(\mathbf{H}_i) \times 1}$ and $\mathbf{a}_t(\phi_{i,q}^t, \theta_{i,q}^t) \in \mathbb{C}^{\text{col}(\mathbf{H}_i) \times 1}$. For specificity, we assume that uniform planar arrays (UPAs) are installed at the TX, RX, and RIS. The transmit array response vector $\mathbf{a}_t(\phi_{\text{SR},q}^t, \theta_{\text{SR},q}^t) \in \mathbb{C}^{M \times 1}$ of the q -th path in \mathbf{H}_{SR} is then given by

$$\begin{aligned} \mathbf{a}_t(\phi_{\text{SR},q}^t, \theta_{\text{SR},q}^t) &= \frac{1}{\sqrt{M}} \left[1, \dots, e^{j \frac{2\pi d}{\lambda} (i_h \sin(\phi_{\text{SR},q}^t) \sin(\theta_{\text{SR},q}^t) + i_v \cos(\theta_{\text{SR},q}^t))}, \right. \\ &\quad \left. \dots, e^{j \frac{2\pi d}{\lambda} ((M_t^h - 1) \sin(\phi_{\text{SR},q}^t) \sin(\theta_{\text{SR},q}^t) + (M_t^v - 1) \cos(\theta_{\text{SR},q}^t))} \right]^T, \end{aligned} \quad (3)$$

where λ and d each denote the signal wavelength and the spacing between the antennas or RIS elements. The horizontal and vertical indexes for the RIS reflecting elements are given by $0 \leq i_h < M^h$ and $0 \leq i_v < M^v$, where $M = M^h M^v$. We omit the explicit definitions of the other array response vectors for the sake of brevity.

The TX uses the precoder $\mathbf{F} \in \mathbb{C}^{N_t \times N_s}$ to send N_s data streams to the RX, which processes the received signal with the combiner $\mathbf{W} \in \mathbb{C}^{N_r \times N_s}$. The processed received signal is given by

$$\mathbf{y} = \mathbf{W}^H \mathbf{H}_{\text{tot}} \mathbf{F} \mathbf{s} + \mathbf{W}^H \mathbf{n}, \quad (4)$$

where the symbol vector $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is defined to satisfy $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{N_s}$, and the precoder \mathbf{F} meets the total power constraint, i.e., $\|\mathbf{F}\|_F^2 \leq P_{\text{TX}}$. The entries of an additive white Gaussian noise (AWGN) vector $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ are independently and identically distributed (i.i.d) with $\mathcal{CN}(0, \sigma_n^2)$. While we do not explicitly include hybrid beamforming architecture in this paper, the system model of interest can be extended to such scenario thanks to the various prior works on hybrid beamformer design [?]. The achievable spectral efficiency with the normally distributed transmitted symbols is written as

$$R = \log_2 \det(\mathbf{I}_{N_s} + \mathbf{R}_{\bar{\mathbf{n}}}^{-1} \mathbf{W}^H \mathbf{H}_{\text{tot}} \mathbf{F} \mathbf{F}^H \mathbf{H}_{\text{tot}}^H \mathbf{W}), \quad (5)$$

where $\mathbf{R}_{\bar{\mathbf{n}}} = \sigma_n^2 \mathbf{W}^H \mathbf{W}$ denotes the covariance matrix of the noise $\bar{\mathbf{n}} = \mathbf{W}^H \mathbf{n}$ after combining.

III. JOINT APPROACH TO CHANNEL ESTIMATION AND BEAMFORMING OPTIMIZATION

In this section, we first formulate the problem of channel estimation and beamforming optimization in RIS-aided MIMO systems. We then use the Bayesian framework to explain the motivation and contribution of our proposed algorithm.

A. Problem Formulation

The problem of beamforming optimization in RIS-aided MIMO systems can be written as

$$(P1) \quad \max_{\mathbf{W}, \Phi, \mathbf{F}} R \quad (6a)$$

$$\text{subject to } \Phi = \text{diag}([e^{j\theta_1}, \dots, e^{j\theta_M}]), \quad (6b)$$

$$\mathbf{R}_{\bar{\mathbf{n}}} = \sigma_n^2 \mathbf{W}^H \mathbf{W}, \quad (6c)$$

$$\mathbf{H}_{\text{tot}} = \mathbf{H}_{\text{SR}} \Phi \mathbf{H}_{\text{TS}}, \quad (6d)$$

$$\|\mathbf{F}\|_F^2 \leq P_{\text{TX}}. \quad (6e)$$

In other words, it aims to maximize the spectral efficiency by designing the beamformer $\mathbf{A}^* = \{\mathbf{W}^*, \Phi^*, \mathbf{F}^*\}$ suited to the given channel realizations \mathbf{H}_{TS} and \mathbf{H}_{SR} . However, \mathbf{H}_{TS} and \mathbf{H}_{SR} are usually not known and thus must be estimated prior to constructing beamformers.

To estimate the channel matrices in practical communication systems, the RX processes the pilot signals that are transmitted by the TX. The processed received pilot signal at the time index $t \in \{0, \dots, T-1\}$ is expressed as

$$\mathbf{y}_p[t] = \mathbf{W}_p^H[t] \mathbf{H}_{\text{SR}} \Phi_p[t] \mathbf{H}_{\text{TS}} \mathbf{F}_p[t] \mathbf{s}[t] + \mathbf{W}_p^H[t] \mathbf{n}[t], \quad (7)$$

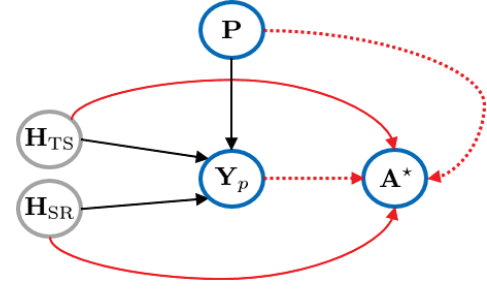


Fig. 2. Bayesian network of the joint channel estimation and beamforming optimization problem in RIS-aided MIMO system. Gray (blue) circles represent the latent (observable) variables, while black (red) arrows indicate the conditional dependencies associated with channel estimation (beamforming optimization). Unlike previous works, our algorithm directly exploits the red dotted arrows by incorporating \mathbf{P} and \mathbf{Y}_p into the design of \mathbf{A}^* .

where $\mathbf{P}[t] = \{\mathbf{W}_p[t], \Phi_p[t], \mathbf{F}_p[t]\}$ is the pilot beamformer, $\mathbf{s}[t]$ is the symbol vector, and each element of $\mathbf{n}[t]$ is i.i.d with $\mathcal{CN}(0, \sigma_n^2)$. The pilot beamformers $\mathbf{P} = \{\mathbf{P}[0], \dots, \mathbf{P}[T-1]\}$ must be constructed so that \mathbf{H}_{TS} and \mathbf{H}_{SR} can be accurately estimated from the received pilot signal matrix $\mathbf{Y}_p = [\mathbf{y}_p[0], \dots, \mathbf{y}_p[T-1]]$. Note that the cascaded channel structure in (1) sets RIS-aided systems apart from conventional counterparts and necessitates novel estimation methods [?].

B. Bayesian Perspective on Proposed Algorithm

Practical communication systems without CSI usually perform channel estimation and beamforming optimization in a sequential manner. Joint channel estimation and beamforming optimization is thus associated with the posterior distribution

$$\begin{aligned} p(\mathbf{A}^*, \mathbf{H}_{\text{TS}}, \mathbf{H}_{\text{SR}} | \mathbf{Y}_p, \mathbf{P}) \\ = p(\mathbf{H}_{\text{TS}}, \mathbf{H}_{\text{SR}} | \mathbf{Y}_p, \mathbf{P}) p(\mathbf{A}^* | \mathbf{Y}_p, \mathbf{P}, \mathbf{H}_{\text{TS}}, \mathbf{H}_{\text{SR}}), \end{aligned} \quad (8)$$

where $p(\mathbf{H}_{\text{TS}}, \mathbf{H}_{\text{SR}} | \mathbf{Y}_p, \mathbf{P})$ and $p(\mathbf{A}^* | \mathbf{Y}_p, \mathbf{P}, \mathbf{H}_{\text{TS}}, \mathbf{H}_{\text{SR}})$ respectively describe the posterior distributions of interest during the pilot-aided channel estimation and beamforming optimization. Fig. 2 shows the Bayesian network of the joint channel estimation and beamforming optimization problem in RIS-aided MIMO system. The latent variables \mathbf{H}_{TS} and \mathbf{H}_{SR} are inferred from the observable variables \mathbf{Y}_p and \mathbf{P} during the channel estimation, while the beamforming optimization involves constructing the beamformer \mathbf{A}^* that maximizes the spectral efficiency R in (5). The red dotted arrows from \mathbf{P} to \mathbf{A}^* and from \mathbf{Y}_p to \mathbf{A}^* show that the pilot beamformers \mathbf{P} , if carefully designed, can also be effectively utilized during the beamforming optimization process. However, this opportunity was overlooked in prior works that limited the use of \mathbf{P} and \mathbf{Y}_p to estimate channels, i.e., they adopted the approximation

$$p(\mathbf{A}^* | \mathbf{Y}_p, \mathbf{P}, \mathbf{H}_{\text{TS}}, \mathbf{H}_{\text{SR}}) \approx p(\mathbf{A}^* | \mathbf{H}_{\text{TS}}, \mathbf{H}_{\text{SR}}) \quad (9)$$

for the sake of simplicity but at the cost of performance degradation [?], [?], [?].

Motivated by these facts, we propose the joint channel estimation and beamforming optimization algorithm for RIS-aided MIMO systems. In addition to accurately estimating

the channel matrices, the proposed algorithm successfully reaps the considerable capacity gains from RIS through the judicious processing of received signals. Unlike prior works, our proposed algorithm directly leverages the pilot signals for passive beamforming optimization and thus is highly suitable for future communication systems that aim to support extremely high spectral efficiency and low latency at the same time [?].

IV. PROPOSED JOINT CHANNEL ESTIMATION AND BEAMFORMING OPTIMIZATION ALGORITHM

As explained in Section III, it can be highly beneficial to exploit the pilot beamformers and received pilot signals to not only estimate the channel matrices but also optimize the beamformers in RIS-aided MIMO systems. Motivated by this, we propose in this section the joint channel estimation and beamforming optimization algorithm that achieves significant improvement in spectral efficiency. In contrast to the majority of the beamforming methods that assume the availability of perfect CSI, the proposed algorithm exploits the channel sparsity present in the systems of interest so as to efficiently obtain the beamformer $\mathbf{A}^* = \{\mathbf{W}^*, \mathbf{\Phi}^*, \mathbf{F}^*\}$ from the received pilot signal matrix \mathbf{Y}_p and AoA/AoD parameters, the latter of which remain relatively constant during the coherence time and thus can be acquired with the aid of various proposed estimation techniques [?].

Let us respectively denote the path indexes for \mathbf{H}_{TS} and \mathbf{H}_{SR} as $i \in \{0, \dots, N_{\text{path}}^{\text{TS}} - 1\}$ and $j \in \{0, \dots, N_{\text{path}}^{\text{SR}} - 1\}$. Then, consider setting the RIS pilot matrix as

$$\mathbf{\Phi}_p[t] = \mathbf{\Phi}_p^{i,j} = \text{diag}(\mathbf{v}_p^{i,j}), \quad (10)$$

where $\mathbf{v}_p^{i,j} = M \text{diag}(\mathbf{a}_r(\phi_{\text{TS},i}^r, \theta_{\text{TS},i}^r))^H \mathbf{a}_t(\phi_{\text{SR},j}^t, \theta_{\text{SR},j}^t)$ is the scaled UPA array response vector constructed with the i -th receive path of \mathbf{H}_{TS} and j -th transmit path of \mathbf{H}_{SR} . Without the loss of generality, we omit the time index t throughout the rest of this section for the sake of notational convenience. The asymptotic orthogonality of UPA array response vectors ensures that

$$\mathbf{H}_{\text{SR}} \mathbf{\Phi}_p^{i,j} \mathbf{H}_{\text{TS}} \rightarrow \alpha_{\text{TS},i} \alpha_{\text{SR},j} \mathbf{a}_r(\phi_{\text{SR},j}^r, \theta_{\text{SR},j}^r) \mathbf{a}_t(\phi_{\text{TS},i}^t, \theta_{\text{TS},i}^t)^H, \quad (11)$$

as $M \rightarrow \infty$ [?]. Since the systems of interest are expected to deploy a large number of passive reflecting elements, setting the RIS matrix as $\mathbf{\Phi}_p^{i,j}$ in (10) has the effect of simplifying the cascaded structure of $\mathbf{H}_{\text{tot}} = \mathbf{H}_{\text{SR}} \mathbf{\Phi}_p \mathbf{H}_{\text{TS}}$ and thus enables the accurate channel estimation despite the absence of baseband signal processing capabilities at the RIS [?]. As will be shown later, $\mathbf{\Phi}_p^{i,j}$ in (10) offers the additional benefit of increasing the capacity of \mathbf{H}_{tot} , thereby successfully reconfiguring the propagation environments for improved communication.

By defining the symbol matrices as $\mathbf{S}_p = [s[0], \dots, s[N_s - 1]] \in \mathbb{C}^{N_s \times N_s}$, we can express the received pilot signal matrix

$\mathbf{Y}_p^{i,j}$ corresponding to the adjusted total channel $\mathbf{H}_{\text{tot},(i,j)} = \mathbf{H}_{\text{SR}} \mathbf{\Phi}_p^{i,j} \mathbf{H}_{\text{TS}}$ as

$$\begin{aligned} \mathbf{Y}_p^{i,j} &= \mathbf{W}_p^H \mathbf{H}_{\text{SR}} \mathbf{\Phi}_p^{i,j} \mathbf{H}_{\text{TS}} \mathbf{F}_p \mathbf{S}_p + \mathbf{W}_p^H \mathbf{N}^{i,j} \\ &= \mathbf{W}_p^H \mathbf{H}_{\text{tot},(i,j)} \mathbf{F}_p \mathbf{S}_p + \mathbf{W}_p^H \mathbf{N}^{i,j}, \end{aligned} \quad (12)$$

where each column vector of $\mathbf{N}^{i,j}$ is identically distributed with $\mathbf{n}[t]$ in (7). In order to accurately estimate $\mathbf{H}_{\text{tot},(i,j)}$ by exploiting its asymptotic property described in (11), we set the pilot combiner $\mathbf{W}_p \in \mathbb{C}^{N_r \times N_s}$ and precoder $\mathbf{F}_p \in \mathbb{C}^{N_t \times N_s}$ as

$$\mathbf{W}_p = [\mathbf{a}_r(\phi_{\text{SR},0}^r, \theta_{\text{SR},0}^r) \quad \dots \quad \mathbf{a}_r(\phi_{\text{SR},N_s-1}^r, \theta_{\text{SR},N_s-1}^r)], \quad (13)$$

$$\mathbf{F}_p = [\mathbf{a}_t(\phi_{\text{TS},0}^t, \theta_{\text{TS},0}^t) \quad \dots \quad \mathbf{a}_t(\phi_{\text{TS},N_s-1}^t, \theta_{\text{TS},N_s-1}^t)]. \quad (14)$$

It then holds that the element $\max(\mathbf{W}_p^H \mathbf{H}_{\text{tot},(i,j)} \mathbf{F}_p)$ of $\mathbf{W}_p^H \mathbf{H}_{\text{tot},(i,j)} \mathbf{F}_p$ with the maximum absolute value satisfies

$$\max(\mathbf{W}_p^H \mathbf{H}_{\text{tot},(i,j)} \mathbf{F}_p) \rightarrow \alpha_{\text{TS},i} \alpha_{\text{SR},j}, \quad (15)$$

as $N_r, N_t, M \rightarrow \infty$. As a result, the combined complex gain $\alpha_{\text{TS},i} \alpha_{\text{SR},j}$ can be estimated by $[\tilde{\mathbf{Y}}_p^{i,j}]_{r_0^{i,j}, c_0^{i,j}}$, where $\tilde{\mathbf{Y}}_p^{i,j} = \mathbf{Y}_p^{i,j} \mathbf{S}_p^{-1}$ is the equalized received signal matrix and

$$\{r_z^{i,j}, c_z^{i,j}\}_{z=0, \dots, k-1} = \underset{r,c}{\text{argmax}} k \left(\left| [\tilde{\mathbf{Y}}_p^{i,j}]_{r,c} \right| \right). \quad (16)$$

The indexes $r_z^{i,j}$ and $c_z^{i,j}$ correspond to the row and column indexes of the element of $\tilde{\mathbf{Y}}_p^{i,j}$ with the top- $(z+1)$ absolute value, and the hyperparameter $k \in \{1, \dots, N_s\}$ defines the number of elements to sample from $\tilde{\mathbf{Y}}_p^{i,j}$. The result in (15) ensures that $[\tilde{\mathbf{Y}}_p^{i,j}]_{r_0^{i,j}, c_0^{i,j}}$ approaches the unbiased estimator of $\alpha_{\text{TS},i} \alpha_{\text{SR},j}$ in the limit of N_r, N_t, M , demonstrating the suitability of the proposed estimator for RIS-aided MIMO systems with channel sparsity. The rank- N_{rank} approximation of $\mathbf{H}_{\text{tot},(i,j)}$ can then be written as

$$\begin{aligned} \hat{\mathbf{H}}_{\text{tot},(i,j)}^{N_{\text{rank}}} &= \\ & \sum_{z=0}^{N_{\text{rank}}-1} [\tilde{\mathbf{Y}}_p^{i,j}]_{r_z^{i,j}, c_z^{i,j}} \mathbf{a}_r(\phi_{\text{SR},r_z^{i,j}}^r, \theta_{\text{SR},r_z^{i,j}}^r) \mathbf{a}_t(\phi_{\text{TS},c_z^{i,j}}^t, \theta_{\text{TS},c_z^{i,j}}^t)^H, \end{aligned} \quad (17)$$

where $N_{\text{rank}} \in \{1, \dots, k\}$. As each of \mathbf{W}_p and \mathbf{F}_p becomes semi-unitary in the limit of its row dimension, $\hat{\mathbf{H}}_{\text{tot},(i,j)}^{N_{\text{rank}}}$ in (17) can be considered as the approximation of the truncated singular value decomposition (SVD) of $\mathbf{H}_{\text{SR}} \mathbf{\Phi}_p^{i,j} \mathbf{H}_{\text{TS}}$, whose angular sparsity allows our proposed algorithm to efficiently acquire CSI without resorting to any iterative method [?].

We now leverage the proposed complex gain estimator to realize the spectral efficiency gain that RIS is expected to bring about. As mmWave and THz channels typically exhibit the presence of one dominant path pair, we set the RIS matrix as $\mathbf{\Phi}^* = \mathbf{\Phi}_p^{i^*, j^*}$, where

$$i^*, j^* = \underset{i,j}{\text{argmax}} 1 \left(\left| [\tilde{\mathbf{Y}}_p^{i,j}]_{r_0^{i,j}, c_0^{i,j}} \right| \right). \quad (18)$$

In other words, we use the complex gain estimator $\{[\tilde{\mathbf{Y}}_p^{i,j}]_{r_0^{i,j}, c_0^{i,j}}\}_{\forall i,j}$ to determine $\mathbf{H}_{\text{tot},(i^*, j^*)}$ that is likely to

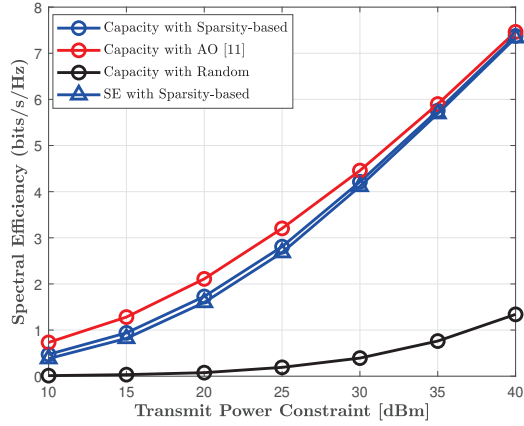


Fig. 3. Capacity and spectral efficiency achieved in RIS-aided mmWave MIMO systems versus the transmit power constraint P_{TX} .

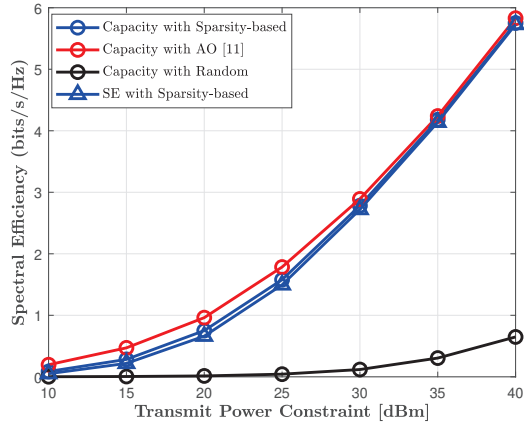


Fig. 4. Capacity and spectral efficiency achieved in RIS-aided THz MIMO systems versus the transmit power constraint P_{TX} .

achieve the highest capacity among $\{\mathbf{H}_{\text{tot},(i,j)}\}_{\forall i,j}$. In fact, the chosen path pair in (18) is guaranteed to be asymptotically optimal in the limit of large arrays and vanishing noise, i.e.,

$$C(\mathbf{H}_{\text{tot},(i^*,j^*)}) \rightarrow \max_{i,j} C(\mathbf{H}_{\text{tot},(i,j)}), \quad (19)$$

as $N_r, N_t, M \rightarrow \infty$ and $\sigma_n \rightarrow 0$. In (19), $C(\mathbf{H})$ denotes the capacity of the channel matrix \mathbf{H} of appropriate dimension. Finally, the corresponding combiner and precoder are obtained as

$$\mathbf{W}^* = \mathbf{W}_{\text{opt}}(\hat{\mathbf{H}}_{\text{tot},(i^*,j^*)}^{N_{\text{rank}}}, N_s), \quad (20)$$

$$\mathbf{F}^* = \mathbf{F}_{\text{opt}}(\hat{\mathbf{H}}_{\text{tot},(i^*,j^*)}^{N_{\text{rank}}}, N_s, P_{\text{TX}}), \quad (21)$$

where the optimal combiner $\mathbf{W}_{\text{opt}}(\mathbf{H}, N_s)$ and precoder $\mathbf{F}_{\text{opt}}(\mathbf{H}, N_s, P_{\text{TX}})$ for \mathbf{H} under the transmit power constraint P_{TX} can be obtained by SVD and waterfilling. We finally note that the effect of imperfect orthogonality of array response vectors in the finite antenna regime can be compensated by setting $N_{\text{rank}} > 1$ for $\hat{\mathbf{H}}_{\text{tot},(i,j)}^{N_{\text{rank}}}$ in (17).

V. SIMULATION RESULTS

In this section, we present simulation results to demonstrate the effectiveness of our proposed channel estimation and beamforming optimization algorithm. We assume that both the TX and RX are equipped with $N_t = N_r = 8 \times 8 = 64$ antennas, while the RIS is equipped with $M = 16 \times 16 = 256$ passive reflecting elements. The number of paths is set to be $N_{\text{path}}^{\text{TS}} = N_{\text{path}}^{\text{SR}} = 4$, while AoAs/AoDs are sampled from the uniform distribution. The distance-dependent path loss model $PL(d_i)$ is given by

$$PL(d_i) [\text{dB}] = \alpha + 10\beta \log_{10}(d_i) + \xi, \quad (22)$$

where $\xi \sim \mathcal{N}(0, \sigma^2)$. We consider both mmWave and THz channels by choosing the appropriate values for the parameters in (22). For 28 GHz mmWave channel, the parameters are set to be $\alpha = 61.4, \beta = 2, \sigma = 5.8$ dB for a line-of-sight (LOS) path of \mathbf{H}_i , and $\alpha = 72.0, \beta = 2.92, \sigma = 8.7$ dB for its non-line-of-sight (NLOS) paths [?]. For THz channel, we consider the usage of 142 GHz carrier frequency by respectively choosing $\alpha = 75.44, \beta = 2.1, \sigma = 2.8$ dB and $\alpha = 75.44, \beta = 3.1, \sigma = 8.3$ dB for LOS and NLOS paths [?]. Other simulation parameters are set as follows: $N_{\text{rank}} = 1, N_s = 4, d_{\text{TS}} = 35$ m, $d_{\text{SR}} = 15$ m, $d = \lambda/2$, and $\sigma_n^2 = -91$ dBm. All the simulation results are averaged over 10,000 independent channel realizations.

Fig. 3 shows the capacity and spectral efficiency achieved by the proposed sparsity-based algorithm and other benchmarks in RIS-aided mmWave MIMO systems. The capacity $C(\mathbf{H}_{\text{tot},(i^*,j^*)})$ of the sparsity-based algorithm is defined as the maximum spectral efficiency achievable under the total cascaded channel $\mathbf{H}_{\text{tot},(i^*,j^*)} = \mathbf{H}_{\text{SR}} \Phi_p^{i^*,j^*} \mathbf{H}_{\text{TS}}$ and the transmit power constraint P_{TX} , i.e.,

$$C(\mathbf{H}_{\text{tot},(i^*,j^*)}) = \sum_{l=1}^{N_s} \log_2 \left(1 + \frac{P_l}{\sigma_n^2} \|\Sigma_{\text{tot}}^* \|_{l,l} \right). \quad (23)$$

In (23), $\Sigma_{\text{tot}}^* \in \mathbb{C}^{N_r \times N_t}$ contains the singular values of $\mathbf{H}_{\text{tot},(i^*,j^*)}$ on its main diagonal, $P_l = \left(\frac{1}{\hat{\eta}} - \frac{\sigma_n^2}{\|\Sigma_{\text{tot}}^* \|_{l,l}^2} \right)^+$ represents the power allocated to the l -th eigenchannel of $\mathbf{H}_{\text{tot},(i^*,j^*)}$, and $\hat{\eta}$ is determined such that $\sum_{l=1}^{N_s} P_l = P_{\text{TX}}$ holds. The capacity of the alternating optimization (AO)-based algorithm can be similarly defined as $C(\mathbf{H}_{\text{tot,AO}})$, where $\mathbf{H}_{\text{tot,AO}}$ is the total channel adjusted according to the method described in [?]. For comparison, we also included the capacity attained when the phase shift of each RIS element is randomly selected from $[0, 2\pi)$. The result shows that the RIS matrix constructed by the proposed algorithm performs close to that constructed by the AO-based algorithm, which assumes the availability of perfect CSI and requires the computational complexity of $\mathcal{O}(3MN_r^3 + 2MN_r^2N_t)$. In contrast, by carefully exploiting the channel sparsity in RIS matrix design, our proposed algorithm achieves substantial capacity gains while requiring only the AoA/AoD parameters and significantly reduced complexity of $\mathcal{O}(MN_rN_t)$. It should also be noted that the relatively poor performance of random RIS design

shows that the performance gain from a large number of RIS elements can only be harnessed through sophisticated passive beamforming optimization, further underscoring the importance of our proposed algorithm. Furthermore, the small difference between the capacity and spectral efficiency of the sparsity-based algorithm demonstrates its capability of accurately estimating the total cascaded channels of RIS-aided MIMO systems.

In Fig. 4, the capacity and spectral efficiency achieved by the proposed sparsity-based algorithm and other benchmarks in RIS-aided THz MIMO systems are plotted as a function of P_{TX} . The result shows that the proposed algorithm achieves spectral efficiency close to the capacity attained by the AO-based algorithm, thereby demonstrating its suitability for THz channels as well. It can thus be concluded that the proposed joint channel estimation and beamforming optimization algorithm can attain significant spectral efficiency gains in future communication systems, which are expected to deploy RIS in order to fully reap the performance benefits that mmWave and THz communications are envisioned to offer.

VI. CONCLUSIONS

In this paper, we proposed the joint channel estimation and beamforming optimization algorithm for RIS-aided MIMO systems. The proposed sparsity-based algorithm successfully constructs the pilot matrices that not only facilitate CSI acquisition process but also provide the systems of interest with considerable capacity gains, thereby resolving the inefficiency involved with the conventional method of limiting the usage of pilot signals to channel estimation. Effectively utilizing the channel sparsity present in future communication systems, our proposed algorithm is capable of achieving significant spectral efficiency improvement while requiring limited computational complexity.