Received 12 September 2023; revised 10 October 2023; accepted 26 October 2023. Date of publication 31 October 2023; date of current version 20 November 2023.

Digital Object Identifier 10.1109/OJCOMS.2023.3328822

Joint Transceiver and Reconfigurable Intelligent Surface Design for Multiuser mmWave MIMO Systems Relying on Non-Diagonal Phase Shift Matrices

JITENDRA SINGH [1] (Member, IEEE), SURAJ SRIVASTAVA [1] (Senior Member, IEEE), ADITYA K. JAGANNATHAM [2] (Senior Member, IEEE), AND LAJOS HANZO [2] (Life Fellow, IEEE)

[1]Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India
[2]School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K.

CORRESPONDING AUTHOR: L. HANZO (e-mail: lh@ecs.soton.ac.uk)

The work of Aditya K. Jagannatham was supported in part by the Qualcomm Innovation Fellowship; in part by the Qualcomm 6G UR Gift; and in part by the Arun Kumar Chair Professorship. The work of Lajos Hanzo was supported in part by the Engineering and Physical Sciences Research Council Projects under Grant EP/W016605/1, Grant EP/X01228X/1, and Grant EP/Y026721/1; and in part by the European Research Council's Advanced Fellow Grant QuantCom under Grant 789028.

ABSTRACT The downlink (DL) of a reconfigurable intelligent surface (RIS)-aided multi-user (MU) millimeter wave (mmWave) multiple-input multiple-output (MIMO) system relying on a non-diagonal RIS (NDRIS) phase shift matrix is considered. A max-min fairness (MMF) problem is formulated under the total transmit power constraint while employing joint active hybrid beamforming (HBF) both at the base station (BS) as well as at each user equipment (UE), and passive beamforming at the NDRIS. To solve this non-convex problem, a sequential optimization method is conceived, wherein the UE having the poorest channel is identified first, which is termed as the worst-case UE. Then the phase shifter coefficients of the NDRIS are optimized using the alternating direction method of multipliers (ADMM) followed by the hybrid transmit precoder (TPC) and receiver combiner (RC) design using the Karcher mean, the least squares and the regularized zero forcing (RZF) principles. Finally, the optimal power allocation is computed using the path-following algorithm. Simulation results show that the proposed NDRIS-HBF system yields an improved worst-case UE rate in comparison to its conventional diagonal RIS (DRIS)-HBF counterpart, while approaching the half-duplex relay (HDR)-HBF benchmark for large values of the number of reflecting elements (REs). Furthermore, the energy efficiency (EE) of the NDRIS structure is significantly higher than that of the DRIS, HDR systems, while being higher than that achieved by the full-duplex relay (FDR) system at high SNR.

INDEX TERMS Millimeter wave, reconfigurable intelligent surface (RIS), hybrid beamforming, multiple-input multiple-output, max-min fairness, energy efficiency.

I. INTRODUCTION

THE EVER-INCREASING demand for high data rates has given impetus to the development of 6G wireless technology. Millimeter wave (mmWave) communication, which exploits the frequency band of $30-300$GHz, can provide a high bandwidth in support of next-generation networks [1], [2], [3], [4], [5], [6]. However, mmWave communication is highly susceptible to blockages. As a remedy, reconfigurable intelligent surfaces (RIS) can play a crucial role in avoiding loss of signal quality due to blockages. RISs are intelligent metasurfaces comprised of electronically controllable elements that can create a passive beamforming effect by reflecting the incident electromagnetic waves in a specific direction. With the assistance of these controllable RIS elements, the desired signal power can be significantly enhanced at the receiver. As they do

not require dedicated RF chains or power supplies, RISs are cost- and energy-efficient. A brief review of the related literature in the area of RIS-aided mmWave communication is presented next.

A. LITERATURE REVIEW

A key challenge in RIS-aided systems is to jointly design the active beamformer at the BS and the passive beamformer at the RIS. The authors of [7], [8], [9], [10], [11], [12], [13], [14], [15] consider different system models and present algorithms for the joint design of these beamformers. In [7], the authors consider a setup with an access point (AP) that has multiple antennas communicating with multiple user equipment (UEs), each having a single antenna, with the help of an RIS. The authors therein minimize the total transmit power under a specific signal to interference plus noise ratio (SINR) based quality of service (QoS) constraint for each UE via joint active and passive beamforming at the AP and RIS, respectively. The resultant problem is solved using semi-definite relaxation (SDR). Wang et al. [8] jointly optimize the active and passive beamformers for a multiple-input and single-output (MISO) system, where a BS serves a single UE with the help of multiple RISs. Huang et al. [10] analyze the energy efficiency (EE) of an RIS-aided downlink (DL) multi-user (MU) MISO system under specific QoS and transmit power constraints for each user. The transmit power at the BS is optimized in their work to achieve the maximum EE of the system through the joint design of the active beamformer at the BS and the passive beamformer at the RIS. The authors have also shown that, by employing RISs, the EE may increase up to 300% in a practical outdoor environment with respect to amplify-and-forward (AF) relaying. The seminal research of Guo et al. [11] resulted in a joint active beamformer at the AP and passive beamformer at the RIS to maximize the weighted sum-rate of RIS-aided MU MISO systems considering both perfect and imperfect channel state information (CSI). The sum-rate of a system was maximized by using discrete phase shifts for passive beamformer at the RIS in a MU communication system by Di et al. [12]. The hybrid beamformer in their work employs digital beamforming at the BS and passive beamforming at the RIS. Kammoun et al. [13] consider a max-min fairness (MMF) problem to maximize the rate of the worst-case user in an RIS-assisted MISO system. They optimize the linear transmit precoder (TPC) and phase shift matrix considering both rank-1 and high-rank channels between the BS and RIS. As a further advance, Xie et al. [9], proposed a novel technique of jointly optimizing the transmit beamformers at the BSs and passive beamformer at the RIS to solve the MMF problem in an RIS-aided multi-cell MISO system. Fu et al. [15] jointly optimize the active and passive beamformers to minimize the total transmit power in RIS-aided MU MISO non-orthogonal multiple access (NOMA) networks. They adopt an alternating optimization framework, while a novel difference-of-convex (DC) programming algorithm was developed to solve the DC program via successive convex relaxation. The authors of [16], [17], [18] compare the performance of RIS with conventional relay aided systems. Specifically, Di Renzo et al. in [16], compare the performance of RIS with full-duplex relay (FDR) and half-duplex relay (HDR) systems at mmWave and sub-mmWave frequencies. The authors in [17] compare the performance of the RIS with decode-and-forward (DF) relaying. The performance of the RIS systems is contrasted with that of AF based FDR/HDR systems in the treatise [18] for single user MIMO systems. Their work presents cutting-edge schemes to design the transmit beamformers at the BS, FDR and HDR as well as the passive beamformer at the RIS. The papers [16], [17], [18] conclude that the RIS performs comparably to a HDR system in terms of SE, whereas it achieves an improved EE when compared to the FDR system.

All the above techniques conceived for joint active and passive beamformer design consider fully-digital beamforming (FDB). It is important to note that such FDB techniques are not applicable in mmWave MIMO systems, since they require an excessive number of radio frequency (RF) chains, which increase both the cost and power consumption. To reduce the number of RF chains in mmWave MIMO systems, hybrid analog and digital (A/D) beamforming (HBF) techniques have been proposed in [27], [28], [29], [30], [31], where the beamformer is divided into baseband (BB) and RF TPC. Therefore, it is necessary to employ the HBF philosophy in RIS-aided mmWave MIMO systems. However, there is limited literature [19], [20], [21], [22], [23], [24], [32], [33], [34], [35], [36], [37], [38], [39], [40] of HBF based RIS-assisted mmWave MIMO systems. Briefly, Tang et al. [32] present a free-space path loss model for RIS in the mmWave frequency range, which describes the key factors to be considered in the RIS based high-frequency bands. Wang et al. [19] maximize the SE of a single-user RIS-aided mmWave MIMO system by jointly optimizing the active hybrid beamformers at the BS as well as at the UE, and passive beamformer at the RIS. Therein, the phase shifters of the RIS are designed using the principle of manifold optimization (MO), while the HBF design is performed using the singular value decomposition (SVD) approach. On the other hand, Bahingayi and Lee [33] proposed a low-complexity algorithm to design the active HBF at the BS and the passive beamformer at the RIS to maximize the SE in DL RIS-aided single user mmWave massive MIMO system. In an interesting development, Wang et al. [34] proposed a solution for the joint design of an active HBF at the BS and passive beamformer in the DL of an RIS-aided mmWave MIMO MU system, relying on the popular subarray structure. As a further advance, Cheng et al. [35] proposed design techniques for low-complexity active beamforming at the BS and passive beamforming at the RIS using beamsteering codebooks in mmWave networks. Furthermore, Ning et al. [36] consider a multi-RIS aided mmWave MU MIMO system and propose a spatially-orthogonal scheme for active beamforming to suppress the interference at the BS. Moreover, for the sake

of nulling the interference power in RIS-aided MU mmWave systems, Ye et al. [37] proposed a Kronecker decomposition based RIS phase shift design, which maximizes the signal power of the desired user while nulling the interference power at the interfering user. Li et al. [20] optimize the transmit power at the BS in RIS-aided MU mmWave MIMO systems by taking the individual SINR constraints of all the users into consideration. A two-layer penalty-based algorithm is used for jointly designing the active and passive beamformers to solve the non-convex problem established therein. Additionally, to reduce the complexity, a sequential optimization method is used to obtain the RIS matrix's phase angles via the maximization of the worst-case user's channel gain, followed by the orthogonal matching pursuit (OMP) and second order cone programming (SOCP)-based procedures for designing the analog and BB beamformers, respectively. Chen et al. [38] design algorithms for RIS aided mmWave vehicular communication relying exclusively on statistical CSI. They proposed a solution for active beamforming at the BS and passive beamforming at the RIS that maximizes of the achievable average sum-rate under specific outage probability constraints. Hong et al. [21] take advantage of the sparse scattering nature and large antenna arrays of mmWave systems for jointly designing the active and passive beamformers in single-user RIS-aided scenarios. Their work considers both narrowband and wideband channels, which makes their study quite comprehensive. Ding et al. [22] perform dynamic resource allocation using a relaxed version of the alternating direction method of multipliers (ADMM) in RIS-aided mmWave MU MIMO systems, under individual QoS constraints of the users. As a further advance, Ying et al. [39] consider RIS-aided wideband mmWave MIMO systems and design the BB TPC using the geometric mean decomposition (GMD) and the analog TPC by the OMP algorithm. Furthermore, the RIS phase shifters are designed separately by exploiting the knowledge of the angles of arrivals (AoAs) and angles of departure (AoDs) of the multipath components at the RIS. By contrast, our recent work [41] jointly optimizes the transmit power and the number of active RF chains to maximize the EE of an RIS aided mmWave MIMO systems. Furthermore, Xiu et al. [23] develop a novel framework for weighted sum-rate maximization of RIS-aided MU mmWave MIMO NOMA networks, where they determine the optimal power allocation, the hybrid beamformers and the RIS phase shifters. Specifically, their algorithm initially performs power loading for a fixed hybrid beamformer as well as for the RIS phase shifters, and subsequently determines for the RIS phase shift angles using the principle of alternating manifold optimization (MO). The hybrid beamformers are finally designed using the successive convex approximation (SCA) technique. Jiao et al. [40] developed handover schemes for UEs by jointly optimizing the active and passive beamformers in RIS-aided mmWave cellular networks. Zhao et al. [24] consider a multi-RIS, multi-BS MU system to study the impact of RIS on user association. To study relay-aided mmWave MIMO systems, the authors of [42] proposed an HBF scheme for an AF based FDR/HDR aided mmWave MU MIMO downlink.

It must be noted that an important feature of [19], [20], [21], [22], [23], [24], [32], [33], [34], [35], [36], [37], [38], [39], [40] described above is the diagonal RIS phase shift matrix, which implies that an electromagnetic wave arriving at a specific RIS element is reflected back from the same element. Such an RIS structure is popularly known as DRIS, where there is no cooperation amongst the RIS elements. Thus, a DRIS does not fully exploit the potential advantages of RIS. However, Shen et al. [25] introduced a more sophisticated fully/group connected architecture for RIS, where the elements cooperate with each other to generate a favourable wave propogation environment. In such RISs, there are $\frac{N(N+1)}{2}$ and $\frac{N(G+1)}{2}$ non-zero entries in fully- and group-connected architectures, respectively, where N and G denote the number of RIS elements and group sizes. These have to be optimized and fed back to the RIS, thereby increasing the complexity and overhead. To avoid this disadvantage, Li et al. in [26] considered a non-diagonal RIS (NDRIS) phase shift matrix containing only N non-zero entries. In this design, the incident wave impinging on an RIS element can be reflected back from any other RIS element, which has the potential of significantly improving the system performance. In the same contribution, Li et al. proposed joint active and passive beamforming schemes using an NDRIS phase shift matrix for RIS-aided MU MIMO systems, which was shown to achieve a remarkably improved performance. However, implementing an NDRIS requires extra switches at the RIS that can route the information from one RIS element to another, as shown in [26]. The switches used therein rely on RF micro-electromechanical systems (MEMS) [43], which are low-cost and energy-efficient components suitable for practical communication systems. For practical implementation of the NDRIS structure, one can follow [44], where the authors proposed a transistor-based implementation of nonreciprocal non-gyrotropic phase gradient metasurface.

While the authors of previous treatises on RIS-aided MU MIMO systems have designed joint active and passive beamformers for the NDRIS structure, none of the contributions have explored the same for mmWave MIMO systems. Due to the high path loss of mmWave systems, the channel gain in conventional DRIS-aided mmWave MU MIMO systems is often insufficient. By contrast, the NDRIS structure can attain a higher channel gain than the conventional DRIS [26], which makes NDRIS attractive in practical RIS-aided mmWave MU MIMO systems. Hence, this work presents a scheme for the joint design of hybrid active beamformers at the BS and passive beamformers at the NDRIS in a NDRIS-aided mmWave MU MIMO system. As discussed in [26], NDRIS has the potential of attaining a higher channel gain for a moderate increase in hardware complexity and power consumption. However, the primary focus of this paper is to solve the max-min

TABLE 1. Summary of literature survey on RIS-aided mmWave MIMO systems.

	[4]	[13]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	Proposed
mmWave MIMO	✓		✓	✓	✓	✓	✓	✓			✓
RIS		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Multiple users	✓	✓		✓		✓	✓	✓			✓
MMF	✓	✓		✓							✓
RZF	✓										✓
Cooperation among RIS elements									✓	✓	✓
ADMM					✓						✓
Optimal power allocation	✓	✓					✓				✓
EE								✓			✓

fairness (MMF) problem in an NDRIS-aided MU mmWave MIMO system to jointly optimize the hybrid beamformer at the BS and the passive beamformer at the NDRIS, which has not been investigated in [26]. Furthermore, the authors of [26] consider continuous phase-shifts for the NDRIS phase shift matrix, which is challenging to realize in practice due to hardware limitations. In contrast, this paper considers a practical discrete phase-shift setup and presents a sequential approach to design the active and passive beamformers, thus easing practical implementation. Furthermore, since the EE is an important consideration in the deployment of an NDRIS, this work also analyzes the EE of the proposed system and compares the results to those of other state-of-the-art approaches, which is absent in [26].

A bold explicit summary of our contributions is given in Table 1. A more detailed list of our novel contributions is provided next.

B. CONTRIBUTIONS OF THIS WORK

- A DL RIS-aided MU mmWave MIMO system model is developed for multi-antenna UEs, each having a single RF chain. The end-to-end channel gains are determined for this system, considering both DRIS and NDRIS structures, employing the optimal passive beamformer at the RIS. Subsequently, an MMF problem is formulated for determining the joint active and passive beamformers for the NDRIS system with the objective of maximizing the transmission rate of the worst-case UE, which is seen to be non-convex.
- A sequential optimization technique is derived for solving the above MMF problem. The worst-case user is identified next, followed by determining the optimal phase-shifts of the NDRIS structure using the ADMM technique.
- Subsequently, the hybrid TPC and RC are designed for this system. The BB TPC is decomposed into two sub-matrices. The front-end BB, RF TPCs, and RCs are jointly determined for maximizing the SE of the system, in a single step, while the back-end BB TPC is obtained via the regularized zero forcing (RZF) technique. The optimal power allocation is found next by employing the path-following algorithm.
- To study the trade-off between the switching power of the NDRIS and the SE achieved by the system, we evaluate the EE of the NDRIS-HBF, DRIS-HBF and NDRIS-FDB modules.
- Simulation results illustrate that the worst-case UE rate for the proposed NDRIS-HBF is improved over that of the DRIS structure and approaches that of the HDR-HBF system, while its EE is higher than the HDR-HBF system for all SNR values, and also the FDR-HBF scheme in the high SNR regime.

C. ORGANIZATION OF THE PAPER

The remainder of this paper is organized as follows. Section II introduces the DL RIS-aided MU mmWave MIMO system, followed by the mmWave MIMO channel model and the MMF problem formulation. Section III introduces the design of the passive beamformer at the NDRIS, followed by the evaluation of the hybrid beamformer, and optimal power allocation. Section IV presents the simulation results and characterizes the EE, followed by our conclusions in Section V.

D. NOTATION

Unless otherwise specified, matrices and vectors are denoted by boldface uppercase letters ($\mathbf{A}$) and boldface lowercase letters ($\mathbf{a}$), respectively; $[\mathbf{A}]_{(i,j)}$ and $[\mathbf{A}]_{(:,i)}$ denote the (i,j)th element and ith column of matrix $\mathbf{A}$, respectively; $[\mathbf{a}]_i$ represents ith element of vector $\mathbf{a}$; The Hermitian of a matrix is denoted by $\mathbf{A}^H$; $||\mathbf{A}||_F$ and $||\mathbf{a}||_F$ represent the Frobenius norm of $\mathbf{A}$ and norm of $\mathbf{a}$, respectively; $|a|$ denotes the magnitude and $\angle a$ represents the phase of a. The notation $\mathcal{D}(\mathbf{a})$ denotes a diagonal matrix with vector $\mathbf{a}$ on its principal diagonal; row($\mathbf{A}$) and col($\mathbf{A}$) denote the number of rows and columns of the matrix $\mathbf{A}$; $\mathbf{I}_M$ denotes an $M \times M$ identity matrix; the symmetric complex Gaussian distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$ is represented as $\mathcal{CN}(\mathbf{a}, \mathbf{A})$.

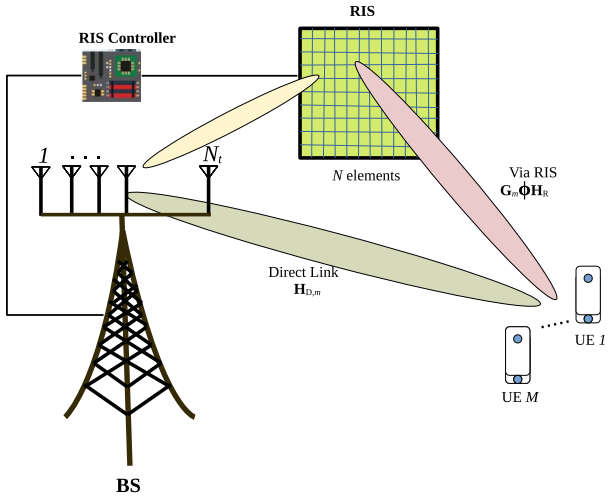

FIGURE 1. RIS-aided downlink MU mmWave MIMO system.

II. RIS-AIDED MMWAVE MU MIMO SYSTEMS

A. SYSTEM MODEL

Consider the MU mmWave DL, where a BS has N_t transmit antennas (TAs) and M_t RF chains for transmission to M UEs each having N_r receive antennas (RAs) using an RIS comprising N passive reflecting elements (REs), as shown in Figure 1. Each UE possesses a single RF chain, and the BS transmits $N_s = M \leq M_t$ streams, one stream for each UE. Let $\mathbf{s} \in \mathbb{C}^{M\times 1}$ denote the transmitted symbol before beamforming at the BS, where we have $\mathbf{s} = [s_1, \ldots, s_M]^T$ and s_m denotes the information signal of the mth UE, for $m \in \mathcal{M} \triangleq \{1, 2, \ldots, M\}$. The symbols s_m are assumed to be independent with average power p_m, i.e., $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathcal{D}(\mathbf{p})$, where $\mathbf{p} = [p_1, \ldots, p_M]^T \in \mathbb{R}^{M\times 1}$. The signal $\mathbf{s}$ is first precoded by the digital BB TPC $\mathbf{F}_{\text{BB}} = [\mathbf{f}_{\text{BB},1}, \ldots, \mathbf{f}_{\text{BB},M}] \in \mathbb{C}^{M_t\times M}$, and subsequently passed through the analog TPC $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t\times M_t}$. The RIS is assumed to have an NDRIS phase shift matrix $\Phi \in \mathbb{C}^{N\times N}$ having N non-zero quantities. The quantities $\mathbf{H}_{\text{D},m} = [\mathbf{h}_{\text{D},m,1}, \ldots, \mathbf{h}_{\text{D},m,N_r}]^H \in \mathbb{C}^{N_r\times N_t}$, $\mathbf{H}_{\text{R}} = [\mathbf{h}_{\text{R},1}, \ldots, \mathbf{h}_{\text{R},N_t}] \in \mathbb{C}^{N\times N_t}$ and $\mathbf{G}_m = [\mathbf{g}_{m,1}, \ldots, \mathbf{g}_{m,N_r}]^H \in \mathbb{C}^{N_r\times N}$ denote the channels spanning from the BS to the mth UE, BS to RIS, and RIS to mth UE, respectively. Thus, the effective channel emerging from the BS to the mth UE can be expressed as $\mathbf{H}_{\text{tot},m} = \mathbf{H}_{\text{D},m} + \mathbf{G}_m\Phi\mathbf{H}_{\text{R}}$. Therefore, the signal $\mathbf{y}_m \in \mathbb{C}^{N_r\times 1}$ received by UE m is given by

$$\begin{aligned}\mathbf{y}_m &= \mathbf{H}_{\text{tot},m}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathbf{s} + \mathbf{n}_m \\ &= \mathbf{H}_{\text{tot},m}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},m}s_m + \sum_{n=1,n\neq m}^{M} \mathbf{H}_{\text{tot},m}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},n}s_n + \mathbf{n}_m,\end{aligned} \tag{1}$$

where $\mathbf{n}_m \in \mathbb{C}^{N_r\times 1}$ represents the complex additive white Gaussian noise (AWGN) with distribution $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. The processed received signal $\widetilde{y}_m$ after employing the RC $\mathbf{w}_m \in \mathbb{C}^{N_r\times 1}$ at the mth UE is given by

$$\begin{aligned}\widetilde{y}_m &= \mathbf{w}_m^H\mathbf{H}_{\text{tot},m}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},m}s_m \\ &\quad + \sum_{n=1,n\neq m}^{M} \mathbf{w}_m^H\mathbf{H}_{\text{tot},m}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB},n}s_n + \widetilde{n}_m,\end{aligned} \tag{2}$$

where $\widetilde{n}_m = \mathbf{w}_m^H\mathbf{n}_m$. In this paper, we consider the hybrid beamforming architecture that constrains the magnitude of all elements of $\mathbf{F}_{\text{RF}}$ and $\mathbf{w}_m$ to be constant, which are set as $\frac{1}{\sqrt{N_t}}$ and $\frac{1}{\sqrt{N_r}}$, respectively. Additionally, we consider the total transmit power at the BS to be constrained as $||\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\mathcal{D}(\mathbf{p})||_F^2 \leq P_t$.

B. MMWAVE MIMO CHANNEL

We employ the widely used Saleh-Valenzuela model to represent the narrowband mmWave MIMO channel [21], which can be expressed as

$$\mathbf{H}_i = \sum_{l=1}^{N_i^{\text{p}}} \alpha_{i,l}\mathbf{a}_{\text{r}}\big(\phi_{i,l}^{\text{r}}, \theta_{i,l}^{\text{r}}\big)\mathbf{a}_{\text{t}}^H\big(\phi_{i,l}^{\text{t}}, \theta_{i,l}^{\text{t}}\big), \tag{3}$$

where $\mathbf{H}_i \in \{\mathbf{H}_{\text{D},m}, \mathbf{H}_{\text{R}}, \mathbf{G}_m\}$, N_i^p denotes the number of multipath components in $\mathbf{H}_i$. The quantity $\alpha_{i,l}$ represents the complex gain of the lth multipath component in $\mathbf{H}_i$. Furthermore, $\mathbf{a}_{\text{t}}(\phi_{i,l}^t\theta_{i,l}^t) \in \mathbb{C}^{\text{col}(\mathbf{H}_i)\times 1}$ denotes the transmit array response vector corresponding to the azimuth and elevation angles of departure (AoDs), namely $\phi_{i,l}^t$, $\theta_{i,l}^t$, respectively. Similarly, $\mathbf{a}_{\text{r}}(\phi_{i,l}^r\theta_{i,l}^r) \in \mathbb{C}^{\text{row}(\mathbf{H}_i)\times 1}$ denotes the receive array response vector corresponding to the azimuth and elevation angles of arrival (AoAs), namely $\phi_{i,l}^r$, $\theta_{i,l}^r$, respectively. We consider uniform planar arrays (UPAs) at the BS, the RIS, and at each UE. As a result, the array response vectors can be written as

$$\begin{aligned}\mathbf{a}_z(\phi, \theta) = \frac{1}{\sqrt{N_z}}\Big[1, \ldots, e^{j\frac{2\pi}{\lambda}d(o\sin\phi\sin\theta + p\cos\theta)}, \ldots, \\ e^{j\frac{2\pi}{\lambda}d\left(\left(N_z^h-1\right)\sin\phi\sin\theta + \left(N_z^v-1\right)\cos\theta\right)}\Big]^T,\end{aligned} \tag{4}$$

where $z \in \{\text{r}, \text{t}\}$, d is the antenna spacing or RIS element spacing, which is assumed to be half of the wavelength λ, $0 \leq o < N_z^h$ and $0 \leq p < N_z^v$, where N_z^h and N_z^v denote the number of horizontal and vertical elements of the UPA in the 2D plane, respectively. Furthermore, the receive array response matrix $\mathbf{A}_i^{\text{r}} \in \mathbb{C}^{N_z^hN_z^v\times N_i^{\text{p}}}$, the complex gain matrix $\Xi_i \in \mathbb{C}^{N_i^{\text{p}}\times N_i^{\text{p}}}$ and transmit array response matrix $\mathbf{A}_i^{\text{t}} \in \mathbb{C}^{N_i^{\text{p}}\times N_z^hN_z^v}$, for the channel $\mathbf{H}_i$, are defined as

$$\begin{aligned}\mathbf{A}_i^{\text{r}} &= \Big[\mathbf{a}_{\text{r}}\big(\phi_{i,0}^{\text{r}}, \theta_{i,0}^{\text{r}}\big) \ldots \mathbf{a}_{\text{r}}\Big(\phi_{i,N_i^{\text{p}}}^{\text{r}}, \theta_{i,N_i^{\text{p}}}^{\text{r}}\Big)\Big] \\ \Xi_i &= \mathcal{D}\Big(\Big[\alpha_{i,0}, \ldots, \alpha_{i,N_i^{\text{p}}}\Big]\Big) \\ \mathbf{A}_i^{\text{t}} &= \Big[\mathbf{a}_{\text{t}}\big(\phi_{i,0}^{\text{t}}, \theta_{i,0}^{\text{t}}\big) \ldots \mathbf{a}_{\text{t}}\Big(\phi_{i,N_i^{\text{p}}}^{\text{t}}, \theta_{i,N_i^{\text{p}}}^{\text{t}}\Big)\Big].\end{aligned} \tag{5}$$

Hence, the mmWave MIMO channel $\mathbf{H}_i$ can be compactly expressed as

$$\mathbf{H}_i = \mathbf{A}_i^{\text{r}}\Xi_i\big(\mathbf{A}_i^{\text{t}}\big)^H. \tag{6}$$

C. PROBLEM FORMULATION

We aim for jointly determining the optimal RF TPC $\mathbf{F}_{\mathrm{RF}}$ and the BB TPC $\mathbf{F}_{\mathrm{BB}}$ at the BS, the NDRIS phase shift matrix Φ at the RIS, and the RC $\mathbf{w}_m$ at each UE, while adhering to the total transmit power constraint. We initially determine the channel gain for a conventional DRIS, when there is no cooperation between the reflective units of the RIS. In the DRIS system, each element has a single connection, which implies that when a wave impinges on an element, a phase-shifted version of it is reflected from the same element. As a result, the output y_n corresponding to the incident signal x_n on the nth element of the RIS in a DRIS structure is given by

$$y_n = x_n e^{j\theta_{n,n}}, n \in \{1, \ldots, N\}, \tag{7}$$

where $\theta_{n,n} \in [0, 2\pi)$ denotes the phase shift occuring at the nth RIS element. As a result, the phase shift matrix $\widetilde{\Phi}$ of the DRIS system is diagonal, and it is given by

$$\widetilde{\Phi} = \mathcal{D}\left(\left[e^{j\theta_{1,1}} e^{j\theta_{2,2}} \ldots e^{j\theta_{N,N}}\right]\right). \tag{8}$$

However, the DRIS structure is unable to harness the full potential of the RIS for enhancing the channel gain of the worst-case UE. This can be seen as follows. Let the channels spanning from the rth BS TA to the RIS and RIS to the sth RA of the mth UE be denoted by $\mathbf{h}_{\mathrm{R},r} \in \mathbb{C}^{N\times 1}$ and $\mathbf{g}_{m,s}^H \in \mathbb{C}^{1\times N}$, respectively. The equivalent channel in the DRIS-based system emerging from the rth BS TA to the sth RA of the mth UE, via the RIS, is given by

$$\begin{aligned}\mathbf{g}_{m,s}^H \widetilde{\Phi} \mathbf{h}_{\mathrm{R},r} &= \sum_{n=1}^{N} \left[\mathbf{g}_{m,s}^H\right]_n e^{j\theta_{n,n}} \left[\mathbf{h}_{\mathrm{R},r}\right]_n, \\ &= \sum_{n=1}^{N} \left|\left[\mathbf{g}_{m,s}^H\right]_n\right| \left|\left[\mathbf{h}_{\mathrm{R},r}\right]_n\right| e^{j\left(\theta_{n,n} + \angle[\mathbf{g}_{m,s}^H]_n + \angle[\mathbf{h}_{\mathrm{R},r}]_n\right)}. \end{aligned} \tag{9}$$

Furthermore, the maximum channel gain, which can be attained by the DRIS design upon setting $\theta_{n,n} = -(\angle[\mathbf{g}_{m,s}^H]_n + \angle[\mathbf{h}_{\mathrm{R},r}]_n)$, is given by

$$\max \left|\mathbf{g}_{m,s}^H \widetilde{\Phi} \mathbf{h}_{\mathrm{R},r}\right|^2 = \left(\sum_{n=1}^{N} \left|\left[\mathbf{g}_{m,s}^H\right]_n\right| \left|\left[\mathbf{h}_{\mathrm{R},r}\right]_n\right|\right)^2. \tag{10}$$

Note that the channel gain in the DRIS is proportional to $(\sum_{n=1}^{N} |[\mathbf{g}_{m,s}^H]_n||[\mathbf{h}_{\mathrm{R},r}]_n|)^2$, in which the vectors $\mathbf{g}_{m,s}$ and $\mathbf{h}_{\mathrm{R},r}$ are combined based on the equal gain combining (EGC) criterion. However, this does not achieve its maximum, since the magnitudes of their elements are not in either ascending or descending order. This can be attributed to the lack of cooperation between the RIS elements in the DRIS structure. On the other hand, when there is cooperation between the reflective units of the RIS, which is possible only in an NDRIS, the incident wave can be rotated into a direction that is in coherence with the direction of the UE channel, thereby leading to the maximum channel gain. To further analyze such a system, consider an NDRIS structure in which the incident wave impinging on the nth RIS element is reflected from the n'th element after a phase-shift. The relationship between the incident and reflected wave can be represented as

$$y_{n'} = x_n e^{j\theta_{f(n),n}}, \tag{11}$$

where $n \in \mathcal{L}, n' \in \mathcal{L}'$ with $\mathcal{L} = \mathcal{L}' = \{1, \ldots N\}$. The bijective function $f{:}\mathcal{L} \rightarrow \mathcal{L}'$ determines the index of reflection for a given index of incidence, i.e., $n' = f(n)$. Note that there are only N non-zero phase entries in the phase shift matrix Φ of the NDRIS structure, which have to be optimized along with determining the optimal mapping function f. Thus, one has to optimize the N phase shift angles and their corresponding locations in the NDRIS. Consider now the corresponding channel gain between the rth BS TA and the sth RA of UE m via the RIS, which is given by

$$\left|\mathbf{g}_{m,s}^H \Phi \mathbf{h}_{\mathrm{R},r}\right| = \sum_{n=1}^{N} \left|\left[\mathbf{g}_{m,s}^H\right]_{f(n)}\right| \left|\left[\mathbf{h}_{\mathrm{R},r}\right]_n\right| e^{j\theta_{f(n),n}}, \tag{12}$$

where the optimal solution for $\theta_{f(n),n}$ can be obtained as $\theta_{f(n),n} = -(\angle[\mathbf{g}_{m,s}^H]_{f(n)} + \angle[\mathbf{h}_{\mathrm{R},r}]_n)$. Note that the channel gain generated by the NDRIS in (12) can be maximized by choosing an appropriate bijective mapping function f that determines the reflection of the incident wave. According to the Maximum Ratio Combining (MRC) criterion, the maximum channel gain in (12) is given by

$$\max \left|\mathbf{g}_{m,s}^H \Phi \mathbf{h}_{\mathrm{R},r}\right|^2 = \left(\sum_{n=1}^{N} \left|\left[\mathbf{g}_{m,s}^H\right]_{(n)}\right| \left|\left[\mathbf{h}_{\mathrm{R},r}\right]_{(n)}\right|\right)^2, \tag{13}$$

where $[\mathbf{g}_{m,s}^H]_{(n)}$ and $[\mathbf{h}_{\mathrm{R},r}]_{(n)}, \forall n$, denote the elements of $\mathbf{g}_{m,s}^H$ and $\mathbf{h}_{\mathrm{R},r}$ sorted in either ascending or descending order. Therefore, the design of the bijective mapping f aims for amalgamating the vectors $\mathbf{g}_{m,s}$ and $\mathbf{h}_{\mathrm{R},r}$ based on the MRC criterion to attain the best channel gain. For practical implementation, we assume REs having digitally controlled phase-shifts, wherein any $\theta_{i,j}$ can only take a finite number of values Q, that are equi-spaced in $[0, 2\pi)$. Therefore, the set $\mathcal{F}_D$ comprising of possible values for $\theta_{i,j}$ is given by

$$\mathcal{F}_D \triangleq \left\{\theta_{i,j} \in \left\{0, \frac{2\pi}{2^Q}, \ldots, \frac{2\pi\left(2^Q - 1\right)}{2^Q}\right\}\right\}, \tag{14}$$

where Q represents the number of control bits.

The maximum achievable SE of the mth UE using (2), is given by the well-known result

$$\mathcal{R}_m = \log_2(1 + \gamma_m), \tag{15}$$

where γ_m is the SINR of the mth UE, given by

$$\gamma_m = \frac{p_m \left|\mathbf{w}_m^H \mathbf{H}_{\mathrm{tot},m} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m}\right|^2}{\sum_{n=1,n\neq m}^{M} p_n \left|\mathbf{w}_m^H \mathbf{H}_{\mathrm{tot},m} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},n}\right|^2 + \sigma^2 ||\mathbf{w}_m||_F^2}. \tag{16}$$

The objective of this paper is to design the hybrid TPC $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$, RC $\mathbf{w}_m$ and phase shift matrix Φ to maximize the overall sum SE, while considering user fairness. In mmWave communication, the UEs can potentially have poor direct

channels due to blockages. Hence, direct maximization of the sum SE can lead to unfairness in resource allocation. As a result, we consider the MMF problem, where we optimize the joint active beamformers at the BS and UE, and passive beamformer at the RIS to maximize the rate of the worst-case UE, subject to both power constraints and mmWave transceiver hardware constraints at the BS along with the reflection constraints at the RIS. Therefore, the MMF problem can be formulated as

$$\mathcal{P}_0: \max_{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}},\Phi,\{\mathbf{w}_m\}_{m=1}^{m=M}} \min_{m\in\mathcal{M}} \mathcal{R}_m \quad \text{s.t.} \begin{cases} \left|[\mathbf{F}_{\mathrm{RF}}]_{(i,j)}\right| = \frac{1}{\sqrt{N_{\mathrm{t}}}}, \forall i,j, \\ |[\mathbf{w}_m]_i| = \frac{1}{\sqrt{N_{\mathrm{r}}}}, \forall i,m, \\ \theta_{f(n),n} \in \mathcal{F}_D, \forall n, \\ \left|\left|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathcal{D}\left(\sqrt{\mathbf{p}}\right)\right|\right|_F^2 \le P_t. \end{cases} \tag{17}$$

It can be readily observed that the constant modulus constraint imposed on the elements of $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{w}_m$, and Φ and the non-convex objective function (OF) render the optimization problem $\mathcal{P}_0$ intractable. The design of the bijective function f, which determines the RE indices of the incident and reflected signals, involves combinatorial optimization, which renders the problem even more challenging. In order to solve this, we develop a low-complexity sequential optimization approach, where the NDRIS phase shift matrix Φ is designed first, followed by the optimization of the hybrid TPC components $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$ and the RC $\mathbf{w}_m$.

However, the NDRIS architecture takes advantage of the higher passive beamforming gain as a result of the MRC criterion by employing a suitable array of switches for connecting the different REs [26]. But, it increases the total power consumption in comparison to the conventional DRIS systems. Thus, EE becomes an important metric since there are various parameters with which the rate of a system can increase but might not be energy efficient due to their high power requirement. Moreover, this metric becomes even more important in mmWave systems due to their massive antenna arrays and power hungry RF chains. Therefore, we evaluate the EE of the system, which is defined as the ratio of sum-rate to total power consumption within a coherence time. Thus, the EE is given by

$$EE = \frac{\sum_{m=1}^{M} \mathcal{R}_m}{P}, \tag{18}$$

where P denotes the power consumption, which comprises of the BS transmit power and hardware static power consumed at the BS and the NDRIS. It should be noted that the NDRIS does not consume any transmit power because its REs are passive devices, which is significantly different from the HDR and FDR architectures. The power consumption in an NDRIS systems depends on the type and the resolution of its individual REs that effectively perform phase shifting on the impinging signal and the power consumed per switch in the reconfiguration of the switches to perform the mapping function f.

III. SEQUENTIAL OPTIMIZATION FOR JOINT RIS PHASE SHIFTER AND HYBRID BEAMFORMER DESIGN

A. RIS PHASE SHIFTER DESIGN

To solve the optimization problem, which becomes non-convex due to the MUI, we decompose the BB TPC matrix $\mathbf{F}_{\mathrm{BB}}$ into the two sub-matrices $\mathbf{F}_{\mathrm{BB}}^{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{BB}}^{2}$. In the first stage, the matrices $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}^{\mathrm{I}}$ are designed jointly for a fixed phase shift matrix Φ, while ignoring the MUI. Subsequently, in the second stage, the matrix $\mathbf{F}_{\mathrm{BB}}^{2}$ is designed to cancel the MUI. The NDRIS phase shift matrix is designed for fixed matrices $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}^{\mathrm{I}}$ and RC $\mathbf{w}_{m^*}$ for the worst-case UE.

Furthermore, one can write the mutual information $\mathcal{I}_m$ of the mth UE, while ignoring the MUI at both the transmitter and receiver, as

$$\begin{aligned}
\mathcal{I}_m &= \log_2\Bigg(1 + \frac{p_m}{||\mathbf{w}_m||^2\sigma^2}\mathbf{w}_m^H\mathbf{H}_{\mathrm{tot},m}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\left(\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right)^H \\
&\quad \times \mathbf{F}_{\mathrm{RF}}^H\mathbf{H}_{\mathrm{tot},m}^H\mathbf{w}_m\Bigg), \\
&= \log_2\Bigg(1 + \frac{p_m}{||\mathbf{w}_m||^2\sigma^2}\mathbf{w}_m^H(\mathbf{H}_{\mathrm{D},m} + \mathbf{G}_m\Phi\mathbf{H}_{\mathrm{R}})\mathbf{F}_{\mathrm{RF}} \\
&\quad \times \mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\left(\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right)^H\mathbf{F}_{\mathrm{RF}}^H(\mathbf{H}_{\mathrm{D},m} + \mathbf{G}_m\Phi\mathbf{H}_{\mathrm{R}})^H\mathbf{w}_m\Bigg), \\
&= \log_2\Bigg(1 + \frac{p_m}{||\mathbf{w}_m||^2\sigma^2}\Bigg(\left|\mathbf{w}_m^H\left(\mathbf{H}_{\mathrm{D},m}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right)\right|^2 \\
&\quad + \left|\mathbf{w}_m^H(\mathbf{G}_m\Phi\mathbf{H}_{\mathrm{R}})\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right|^2 + 2\mathbf{w}_m^H\mathbf{H}_{\mathrm{D},m}\mathbf{F}_{\mathrm{RF}} \\
&\quad \times \mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\left(\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right)^H\mathbf{H}_{\mathrm{R}}^H\Phi^H\mathbf{G}_m^H\mathbf{w}_m\Bigg)\Bigg), \\
&= \log_2\Bigg(1 + \frac{p_m}{||\mathbf{w}_m||^2\sigma^2}\Bigg(\left|\mathbf{w}_m^H\left(\mathbf{H}_{\mathrm{D},m}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right)\right|^2 \\
&\quad + \left|\mathbf{w}_m^H(\mathbf{G}_m\Phi\mathbf{H}_{\mathrm{R}})\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right|^2 + 2\mathbf{w}_m^H\mathbf{H}_{\mathrm{D},m}\mathbf{H}_{\mathrm{R}}^H \\
&\quad \times \Phi^H\mathbf{G}_m^H\mathbf{w}_m\Bigg)\Bigg), \\
&\approx \log_2\Bigg(1 + \frac{p_m}{||\mathbf{w}_m||^2\sigma^2}\Bigg(\left|\mathbf{w}_m^H\mathbf{H}_{\mathrm{D},m}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right|^2 \\
&\quad + \left|\mathbf{w}_m^H\mathbf{G}_m\Phi\mathbf{H}_{\mathrm{R}}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},m}^{\mathrm{I}}\right|^2\Bigg)\Bigg).
\end{aligned} \tag{19}$$

The approximation in the last step of (19) is due to the fact that the term $\mathbf{H}_{\mathrm{D},m}\mathbf{H}_{\mathrm{R}}^H$ converges to $\mathbf{0}$ for a large number of TAs. This can be seen as follows upon using (6), one obtains

$$\mathbf{H}_{\mathrm{D},m}\mathbf{H}_{\mathrm{R}}^H = \mathbf{A}_{\mathrm{D},m}^{\mathrm{r}}\Xi_{\mathrm{D},m}\left(\mathbf{A}_{\mathrm{D},m}^{\mathrm{t}}\right)^H\mathbf{A}_{\mathrm{R}}^{\mathrm{t}}\Xi_{\mathrm{R}}\mathbf{A}_{\mathrm{R}}^{\mathrm{t}}. \tag{20}$$

Since the AoDs of the different paths at the BS can be modeled as independent random variables, it follows that the event $E = \{\phi_{\mathrm{D},m,a_1}^t \neq \phi_{\mathrm{R},a_2}^t, \theta_{\mathrm{D},m,a_1}^t \neq \theta_{\mathrm{R},a_2}^t, \forall a_1, a_2 \in \{1,\ldots N_i^p\}\}$ occurs with probability 1 [21]. Furthermore, for UPA with large N_{t}, $|\mathbf{a}_t^H(\phi_{\mathrm{D},m,a_1},\theta_{\mathrm{D},m,a_1})\mathbf{a}_t(\phi_{\mathrm{R},a_2},\theta_{\mathrm{R},a_2})| \to 0$, for any $(\phi_{\mathrm{D},m,a_1},\theta_{\mathrm{D},m,a_1}) \neq (\phi_{\mathrm{R},a_2},\theta_{\mathrm{R},a_2})$ [45]. Therefore, the columns of $\mathbf{A}_{\mathrm{D},m}^{\mathrm{t}}$ and $\mathbf{A}_{\mathrm{R}}^{\mathrm{t}}$ form orthonormal set for large values of N_{t}, which results in

$$\left(\mathbf{A}_{\mathrm{D},m}^{\mathrm{t}}\right)^H\mathbf{A}_{\mathrm{R}}^{\mathrm{t}} \to \mathbf{0}^{N_{\mathrm{D},m}^P \times N_{\mathrm{R}}^P}, \text{as} N_{\mathrm{t}} \to \infty. \tag{21}$$

Note that the approximation in (19) comprises the gains corresponding to two links, viz., the gain $|\mathbf{w}_m^H \mathbf{H}_{\mathrm{D},m} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m}^{\mathrm{l}}|^2$ of the direct BS-UE link, and the gain $|\mathbf{w}_m^H \mathbf{G}_m \Phi \mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m}^{\mathrm{l}}|^2$ of the BS-RIS-UE link. Since, the signal power received at any UE via reflection by the RIS is typically much lower in comparison to the direct signal arriving from the BS, the UE having the highest path-loss for the BS-UE link is considered to be the worst-case UE. Since mmWave communication is adversely affected due to its severe path loss, this paper optimizes the worst-case UE based on the highest path loss, while ignoring the MUI in the first stage. This can be mathematically determined as follows. Let the SVD of $\mathbf{H}_{\mathrm{D},m}$ be given as $\mathbf{H}_{\mathrm{D},m} = \mathbf{U}_{\mathrm{D},m} \Sigma_{\mathrm{D},m} \mathbf{V}_{\mathrm{D},m}$. The ideal fully digital TPC and RC can now be determined for user m as $\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m}^{\mathrm{l}} = [\mathbf{V}_{\mathrm{D},m}]_{(:,1)}$ and $\mathbf{w}_m = [\mathbf{U}_{\mathrm{D},m}]_{(:,1)}$. The worst-case UE is now identified as

$$\begin{aligned} m^* = \arg \min_{m \in \mathcal{M}} \quad & \left| \mathbf{w}_m^H \mathbf{H}_{\mathrm{D},m} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m}^{\mathrm{l}} \right|^2 \\ \text{s.t.} \quad & \mathbf{w}_m = \left[\mathbf{U}_{\mathrm{D},m} \right]_{(:,1)}, \\ & \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m}^{\mathrm{l}} = \left[\mathbf{V}_{\mathrm{D},m} \right]_{(:,1)}. \end{aligned} \tag{22}$$

From the above equation, one can obtain the worst-case UE m^* by computing the maximum eigenvalue of each matrix $\mathbf{H}_{\mathrm{D},m} \mathbf{H}_{\mathrm{D},m}^H$, $\forall m$, and choosing the one that is the minimum. Following this, we design the hybrid beamformer at the BS, each UE, and also the passive beamformer at the RIS, by maximizing the rate of the worst-case UE. The resultant optimization problem is conceived as

$$\begin{aligned} \mathcal{P}_1 : \max_{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \Phi, \{\mathbf{w}_m\}_{m=1}^{m=M}} \quad & \mathcal{R}_{m^*} \\ \text{s.t.} \quad & \begin{cases} \left| [\mathbf{F}_{\mathrm{RF}}]_{(i,j)} \right| = \frac{1}{\sqrt{N_{\mathrm{t}}}}, \forall i, j, \\ |[\mathbf{w}_m]_i| = \frac{1}{\sqrt{N_{\mathrm{r}}}}, \forall i, m, \\ \theta_{f(n),n} \in \mathcal{F}_D, \forall n, \\ \left|\left| \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \mathcal{D}\left(\sqrt{\mathbf{p}}\right) \right|\right|_F^2 \leq P_t. \end{cases} \end{aligned} \tag{23}$$

Note that the above optimization problem is non-convex due to the non-convex constant magnitude constraints imposed on the elements of the RF TPC, RC and RIS coefficients. Since the variables $\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \mathbf{w}_m$ and Φ are coupled in the OF of (23), we use the principle of alternating optimization (AO). As part of this procedure, the passive beamforming matrix Φ is initially determined for a fixed TPC $\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m}^{\mathrm{l}}$ and RC $\mathbf{w}_m$. Subsequently, the hybrid TPC and RC are computed for a fixed Φ. To begin with, set the TPC and RC for the worst-case UE as the ideal fully-digital equivalents obtained via the SVD of the channel matrix $\mathbf{H}_{\mathrm{R}}$ and $\mathbf{G}_{m^*}$, respectively. Toward this, let us define the SVD of $\mathbf{G}_{m^*}$ as $\mathbf{U}_{\mathrm{G},m^*} \Sigma_{\mathrm{G},m^*} \mathbf{V}_{\mathrm{G},m^*}^H$, the SVD of $\mathbf{H}_{\mathrm{R}}$ as $\mathbf{U}_{\mathrm{R}} \Sigma_{\mathrm{R}} \mathbf{V}_{\mathrm{R}}^H$, and set $\mathbf{w}_{m^*} = [\mathbf{U}_{\mathrm{G},m^*}]_{(:,1)}$, $\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}} = [\mathbf{V}_{\mathrm{R}}]_{(:,1)}$. Note that the TPC and RC are used here to extract the phase shifts of Φ. The actual hybrid TPC and RC are designed later. The passive beamformer at the RIS in now designed by maximizing the rate of the worst-case UE m^*. Therefore, the optimization problem to design Φ can be framed as

$$\begin{aligned} \mathcal{P}_2 : \max_{\Phi} \quad & \left| \mathbf{w}_{m^*}^H \mathbf{G}_{m^*} \Phi \mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}} \right|^2 \\ \text{s.t.} \quad & \begin{cases} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}} = [\mathbf{V}_{\mathrm{R}}]_{(:,1)}, \\ \mathbf{w}_{m^*} = \left[\mathbf{U}_{\mathrm{G},m^*} \right]_{(:,1)}, \\ \theta_{f(n),n} \in \mathcal{F}_D, \forall n. \end{cases} \end{aligned} \tag{24}$$

For a fixed TPC $\mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}}$ and RC $\mathbf{w}_{m^*}$, the optimization problem $\mathcal{P}_2$ is similar to (13). Hence, the beamforming gain $|\mathbf{w}_{m^*}^H \mathbf{G}_{m^*} \Phi \mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}}|^2$ becomes maximum, when the phases in Φ are matched to the phases of the ordered elements of the row vector $\mathbf{w}_{m^*}^H \mathbf{G}_{m^*}$ and column vector $\mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}}$. The ordering of the elements is achieved via the mapping function f. Toward ordering the elements, one can define the permutation matrices $\mathbf{J}_{\mathrm{t}}$ and $\mathbf{J}_{\mathrm{r}}$, that permute the elements of $\mathbf{w}_{m^*}^H \mathbf{G}_{m^*}$ and $\mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}}$, to arrange the amplitudes of the vectors in ascending order, when multiplied on the right and left of the corresponding vectors, respectively. The equivalent optimization problem can be derived as

$$\begin{aligned} \mathcal{P}_3 : \max_{\widetilde{\Phi}} \quad & \left| \mathbf{w}_{m^*}^H \mathbf{G}_{m^*} \mathbf{J}_{\mathrm{t}} \widetilde{\Phi} \mathbf{J}_{\mathrm{r}} \mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}} \right|^2 \\ \text{s.t.} \quad & \begin{cases} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}} = [\mathbf{V}_{\mathrm{R}}]_{(:,1)} \\ \mathbf{w}_{m^*} = \left[\mathbf{U}_{\mathrm{G},m^*} \right]_{(:,1)} \\ \theta_{n,n} \in \mathcal{F}_D, \forall n. \end{cases} \end{aligned} \tag{25}$$

Once the matrices $\mathbf{J}_{\mathrm{t}}$ and $\mathbf{J}_{\mathrm{r}}$ are obtained, the row vector $\mathbf{w}_{m^*}^H \mathbf{G}_{m^*} \mathbf{J}_{\mathrm{t}}$ and column vector $\mathbf{J}_{\mathrm{r}} \mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}}$ are fixed and their elements are arranged in ascending order of magnitudes. Note that, the problem (25) is equivalent to (9), which designs the phase shifts of the DRIS matrix. Assume now that $\mathbf{r}_{m^*} = \mathcal{D}(\mathbf{w}_{m^*}^H \mathbf{G}_{m^*} \mathbf{J}_{\mathrm{t}}) \mathbf{J}_{\mathrm{r}} \mathbf{H}_{\mathrm{R}} \mathbf{F}_{\mathrm{RF}} \mathbf{f}_{\mathrm{BB},m^*}^{\mathrm{l}}$. As a result, the problem concerning the design of the DRIS phase shift matrix $\widetilde{\Phi} = \mathcal{D}(\boldsymbol{\theta})$, $\boldsymbol{\theta} = [\theta_{1,1}, \ldots, \theta_{N,N}]$, can be formulated as

$$\begin{aligned} \mathcal{P}_4 : \max_{\boldsymbol{\theta}} \quad & \left| \boldsymbol{\theta}^H \mathbf{r}_{m^*} \right|^2 \\ \text{s.t.} \quad & \theta_{n,n} \in \mathcal{F}_D, \forall n. \end{aligned} \tag{26}$$

The above problem is non-convex due to constraints pertaining to $\mathcal{F}_D$ in equation (26). A novel algorithm based on the ADMM principle is now devised to solve the above problem. Toward this, we introduce an auxiliary variable $\boldsymbol{\vartheta}$ for $\boldsymbol{\theta}$ and the penalty term μ for $\boldsymbol{\vartheta} \neq \boldsymbol{\theta}$. Upon employing these quantities, $\mathcal{P}_4$ can be recast as

$$\begin{aligned} \mathcal{P}_{4a} : \max_{\boldsymbol{\theta}, \boldsymbol{\vartheta}} \quad & \left| \boldsymbol{\vartheta}^H \mathbf{r}_{m^*} \right|^2 - \frac{\mu}{2} \| \boldsymbol{\vartheta} - \boldsymbol{\theta} \|^2 \\ \text{s.t.} \quad & \boldsymbol{\vartheta} = \boldsymbol{\theta}, \\ & \theta_{n,n} \in \mathcal{F}_D, \forall n. \end{aligned} \tag{27}$$

The Lagrangian function associated with the above equation (27) is derived as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\vartheta}, \boldsymbol{\theta}, \lambda) = & \left| \boldsymbol{\vartheta}^H \mathbf{r}_{m^*} \right|^2 - \sum_{n=1}^{N} 1\!\!1_{\mathcal{F}_D}(\theta_n) - \frac{\mu}{2} \| \boldsymbol{\vartheta} - \boldsymbol{\theta} \|^2 \\ & + \lambda^H (\boldsymbol{\vartheta} - \boldsymbol{\theta}), \end{aligned} \tag{28}$$

Algorithm 1 NDRIS Phase Shift Matrix Design

Input: $\mathbf{H}_{\mathrm{D},m}, \mathbf{G}_m, \mathbf{H}_{\mathrm{R}}, \forall m$

1: Calculate SVD$(\mathbf{H}_{\mathrm{D},m}) = \mathbf{U}_{\mathrm{D},m}\Sigma_{\mathrm{D},m}\mathbf{V}^H_{\mathrm{D},m}$
2: Find the worst-case user index m^* using Eq. (22)
3: $\mathbf{J}_{\mathrm{t}}$=asc$(\mathbf{w}^H_{m^*}\mathbf{G}_{m^*})$ and $\mathbf{J}_{\mathrm{g}}$ =asc$(\mathbf{H}_{\mathrm{R}}\mathbf{F}_{\mathrm{RF}}\mathbf{f}^{\mathrm{l}}_{\mathrm{BB},m^*})$
4: Design $\boldsymbol{\theta}^*$ by solving (26)
5: NDRIS phase shift matrix is given by $\Phi^{\mathrm{opt}} = \mathbf{J}_{\mathrm{t}}\widetilde{\Phi}\mathbf{J}_{\mathrm{g}}$

Output: $\Phi, \mathbf{w}_m, \forall m, \mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}$

where $1\!\mathrm{l}_{\mathcal{F}_D}\{\psi\}$ is the indicator function of the set $\mathcal{F}_D$, which is defined as

$$1\!\mathrm{l}_{\mathcal{F}_D}\{\psi\} = \begin{cases} 0, & \psi \in \mathcal{F}_D, \\ \infty, & \text{otherwise.} \end{cases} \tag{29}$$

The dual problem for $\mathcal{P}_{4a}$ can be formulated as

$$\mathcal{P}_{4b}: \min_{\lambda} \mathcal{D}(\lambda) = \max_{\boldsymbol{\theta},\boldsymbol{\vartheta}} \{\mathcal{L}(\boldsymbol{\vartheta}, \boldsymbol{\theta}, \lambda)\}. \tag{30}$$

The solution to the above dual problem (30) can be found using the ADMM technique, that has the following iterative steps:

$$\boldsymbol{\theta}^{t+1} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\vartheta}^t, \boldsymbol{\theta}, \bar{\boldsymbol{\lambda}}^t\right), \tag{31}$$

$$\boldsymbol{\vartheta}^{t+1} = \arg\max_{\boldsymbol{\vartheta}} \mathcal{L}\left(\boldsymbol{\vartheta}, \boldsymbol{\theta}^{t+1}, \bar{\boldsymbol{\lambda}}^t\right), \tag{32}$$

$$\bar{\boldsymbol{\lambda}}^{t+1} = \bar{\boldsymbol{\lambda}}^t - \mu\left(\boldsymbol{\vartheta}^{t+1} - \bar{\boldsymbol{\lambda}}^{t+1}\right), \tag{33}$$

where t is the iteration index.

1) *Optimizing $\boldsymbol{\theta}$:* One can obtain the optimal value of $\boldsymbol{\theta}$ from Eq. (31) for fixed $\boldsymbol{\vartheta}^t$ and λ^t, which can be evaluated as

$$\boldsymbol{\theta}^{t+1} = \mathrm{Pj}_{\mathcal{F}}\left(\boldsymbol{\vartheta}^t - \frac{1}{\mu}\bar{\boldsymbol{\lambda}}^t\right), \tag{34}$$

where $\mathrm{Pj}_{\mathcal{F}}$ represents the projection function. Considering $\boldsymbol{\theta} = \boldsymbol{\vartheta}^t - \frac{1}{\mu}\bar{\boldsymbol{\lambda}}^t$, one can set

$$\angle\theta_n^{t+1} = \arg \min_{\varphi_n \in \{0, \frac{2\pi}{Q}, \ldots, \frac{2\pi(Q-1)}{Q}\}} |\varphi_n - \angle\bar{\theta}_n|. \tag{35}$$

2) *Optimizing $\boldsymbol{\vartheta}$:* For a given $\boldsymbol{\theta}^{t+1}$ and λ^t, $\boldsymbol{\vartheta}$ can be determined using Eq. (32) as

$$\boldsymbol{\vartheta}^{t+1} = \left(2\mathbf{r}_{m^*}\mathbf{r}^H_{m^*} + \mu\mathbf{I}_N\right)^{-1}\left(\mu\boldsymbol{\theta}^{t+1} + \bar{\boldsymbol{\lambda}}^t\right). \tag{36}$$

The optimal NDRIS matrix Φ^{opt} is computed as

$$\Phi^{\mathrm{opt}} = \mathbf{J}_{\mathrm{t}}\widetilde{\Phi}\mathbf{J}_{\mathrm{g}}. \tag{37}$$

B. HYBRID BEAMFORMER DESIGN

This subsection designs the hybrid TPC and RC by solving the MMF problem based on the net channel $\widetilde{\mathbf{H}}_{\mathrm{tot},m} = \mathbf{H}_{\mathrm{D},m} + \mathbf{G}_m\Phi^{\mathrm{opt}}\mathbf{H}_{\mathrm{R}}$. The pertinent optimization problem can be formulated as

$$\mathcal{P}_5: \max_{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \{\mathbf{w}_m\}_{m=1}^{m=M}} \min_{m\in\mathcal{M}} \mathcal{R}_m,$$

$$\text{s.t.} \begin{cases} \left|[\mathbf{F}_{\mathrm{RF}}]_{(i,j)}\right| = \frac{1}{\sqrt{N_{\mathrm{t}}}}, \forall i, j, \\ |[\mathbf{w}_m]_i| = \frac{1}{\sqrt{N_{\mathrm{r}}}}, \forall i, m, \\ \left|\left|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathcal{D}\left(\sqrt{\mathbf{p}}\right)\right|\right|_F^2 \le P_t. \end{cases} \tag{38}$$

As mentioned previously, the BB TPC $\mathbf{F}_{\mathrm{BB}}$ is decomposed into $\mathbf{F}^{\mathrm{l}}_{\mathrm{BB}}$ and $\mathbf{F}^{2}_{\mathrm{BB}}$. In Stage-1, the RF TPC and the front end BB TPC $\mathbf{F}^{\mathrm{l}}_{\mathrm{BB}}$ are designed jointly. Toward this, let us decompose the matrix $\widetilde{\mathbf{H}}_{\mathrm{tot},m}$ using the SVD as

$$\widetilde{\mathbf{H}}_{\mathrm{tot},m} = \widetilde{\mathbf{U}}_{\mathrm{tot},m}\widetilde{\Sigma}_{\mathrm{tot},m}\widetilde{\mathbf{V}}^H_{\mathrm{tot},m}, \tag{39}$$

which yields the optimal TPC $\mathbf{f}^{\mathrm{opt}}_m$ and RC $\mathbf{w}^{\mathrm{opt}}_m$ for the mth UE as

$$\begin{aligned} \mathbf{f}^{\mathrm{opt}}_m &= \left[\widetilde{\mathbf{V}}_{\mathrm{tot},m}\right]_{(:,1)}, \\ \mathbf{w}^{\mathrm{opt}}_m &= \left[\widetilde{\mathbf{U}}_{\mathrm{tot},m}\right]_{(:,1)}. \end{aligned} \tag{40}$$

According to [27], designing a hybrid TPC $\mathbf{F}_{\mathrm{RF}}\mathbf{f}^{\mathrm{l}}_{\mathrm{BB},m}$ that maximizes the rate is equivalent to minimizing the distance between the optimal unconstrained fully-digital TPC $\mathbf{f}^{\mathrm{opt}}_m$ and the hybrid TPC $\mathbf{F}_{\mathrm{RF}}\mathbf{f}^{\mathrm{l}}_{\mathrm{BB},m}$, which can formulated as the problem

$$\begin{aligned} \left(\mathbf{F}^{\mathrm{opt}}_{\mathrm{RF}}, \mathbf{f}^{\mathrm{l,opt}}_{\mathrm{BB}}\right) = \arg\min \quad & \left|\left|\mathbf{f}^{\mathrm{opt}}_m - \mathbf{F}_{\mathrm{RF}}\mathbf{f}^{\mathrm{l}}_{\mathrm{BB},m}\right|\right|_F^2, \forall m \\ \text{s.t.} \quad & \left|[\mathbf{F}_{\mathrm{RF}}]_{(i,j)}\right| = \frac{1}{\sqrt{N_{\mathrm{t}}}}, \forall i, j. \end{aligned} \tag{41}$$

Note that the BS and each UE design their TPC and RC independently, while ignoring the MUI. Concatenate now $\mathbf{f}^{\mathrm{opt}}_m$, to obtain the matrix $\mathbf{F}^{\mathrm{opt}} \in \mathbb{C}^{N_{\mathrm{t}}\times M}$ and $\mathbf{f}^{\mathrm{l}}_{\mathrm{BB},m}$, to obtain $\mathbf{F}^{\mathrm{l}}_{\mathrm{BB}} \in \mathbb{C}^{M_{\mathrm{t}}\times M}$ as formulated

$$\begin{aligned} \mathbf{F}^{\mathrm{opt}} &= \left[\mathbf{f}^{\mathrm{opt}}_1, \ldots, \mathbf{f}^{\mathrm{opt}}_M\right], \\ \mathbf{F}^{\mathrm{l}}_{\mathrm{BB}} &= \left[\mathbf{f}^{\mathrm{l}}_{\mathrm{BB},1}, \ldots, \mathbf{f}^{\mathrm{l}}_{\mathrm{BB},M}\right]. \end{aligned} \tag{42}$$

Using the above quantities, the problem of jointly designing of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}^{\mathrm{l}}_{\mathrm{BB}}$ is formulated as

$$\begin{aligned} \left(\mathbf{F}^{\mathrm{opt}}_{\mathrm{RF}}, \mathbf{F}^{\mathrm{l,opt}}_{\mathrm{BB}}\right) = \arg \min_{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}^{\mathrm{l}}_{\mathrm{BB}}} \quad & \left|\left|\mathbf{F}^{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}^{\mathrm{l}}_{\mathrm{BB}}\right|\right|_F^2, \\ \text{s.t.} \quad & \left|[\mathbf{F}_{\mathrm{RF}}]_{(i,j)}\right| = \frac{1}{\sqrt{N_{\mathrm{t}}}}, \forall i, j. \end{aligned} \tag{43}$$

It can be seen from (43), that the hybrid TPC $\mathbf{F}_{\mathrm{RF}}\mathbf{F}^{\mathrm{l}}_{\mathrm{BB}}$ can be designed by projecting $\mathbf{F}_{\mathrm{RF}}\mathbf{F}^{\mathrm{l}}_{\mathrm{BB}}$ onto $\mathbf{F}^{\mathrm{opt}}$, while taking also the constraints imposed on $\mathbf{F}_{\mathrm{RF}}$ into account. However, determining such a projection under the non-convex constraints on the elements of $\mathbf{F}_{\mathrm{RF}}$ is intractable. Toward this, the authors of [4], [27], [28], [29] proposed various techniques with and without codebooks to solve the hybrid TPC and RC design problem. We design a low-complexity hybrid TPC and RC without relying on any codebook, as discussed next.

It can also be observed that the RF TPC $\mathbf{F}_{\mathrm{RF}}$ is common to all the UEs. To minimize the average distance between

$\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}^{\mathrm{opt}}$, the analog TPC $\mathbf{F}_{\mathrm{RF}}$ design problem can be formulated as

$$\left(\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}\right) = \arg\min_{\mathbf{F}_{\mathrm{RF}}} \quad \mathbb{E}\left[d_{\mathrm{chord}}\left(\mathbf{F}^{\mathrm{opt}}, \mathbf{F}_{\mathrm{RF}}\right)\right], \tag{44}$$

$$\text{s.t.} \quad \left|[\mathbf{F}_{\mathrm{RF}}]_{(i,j)}\right| = \frac{1}{\sqrt{N_{\mathrm{t}}}}, \forall i, j, \tag{45}$$

where d_{chord} represents the chordal distance function. The Karcher mean [29] can be used to find the solution of the above problem, which is given by

$$\mathbf{F}_{\mathrm{RF}} = \frac{1}{\sqrt{N_{\mathrm{t}}}} \overline{\mathbf{U}}_m^1 \oslash |\overline{\mathbf{U}}_m^1|, \tag{46}$$

where the operator $\oslash$ divides each element of the matrix by its magnitude, and $\overline{\mathbf{U}}_m^1$ represents the first M_{t} eigenvectors of $\mathbf{F}^{\mathrm{opt}}(\mathbf{F}^{\mathrm{opt}})^H$. Along similar lines, the optimal RF RC for each UE is given by

$$\mathbf{w}_m = \frac{1}{\sqrt{N_{\mathrm{r}}}} \mathbf{w}_m^{\mathrm{opt}} \oslash |\mathbf{w}_m^{\mathrm{opt}}|, \forall m. \tag{47}$$

Subsequently, the BB TPC $\mathbf{F}_{\mathrm{BB}}^1$ is designed for the fixed RF TPC $\mathbf{F}_{\mathrm{RF}}$ obtained from (46) as

$$\mathbf{F}_{\mathrm{BB}}^{1,\mathrm{opt}} = \arg\min_{\mathbf{F}_{\mathrm{BB}}^1} \left|\left|\mathbf{F}^{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}^1\right|\right|_F^2. \tag{48}$$

The solution to this problem is given by the least squares estimator

$$\mathbf{F}_{\mathrm{BB}}^1 = \left(\mathbf{F}_{\mathrm{RF}}^H \mathbf{F}_{\mathrm{RF}}\right)^{-1} \mathbf{F}_{\mathrm{RF}}^H \mathbf{F}^{\mathrm{opt}}. \tag{49}$$

Furthermore, MUI mitigation is performed by the back-end BB TPC $\mathbf{F}_{\mathrm{BB}}^2$. This procedure is discussed next.

To suppress the MUI, one can design $\mathbf{F}_{\mathrm{BB}}^2$ based on the RZF principle [4], which strikes a trade-off between the MUI rejection and noise power amplification using a regularization parameter. According to this principle, the BS obtains the effective channel $\mathbf{h}_m^{\mathrm{eff}} \in \mathbb{C}^{1\times M}, \forall m$, as

$$\mathbf{h}_m^{\mathrm{eff}} = \mathbf{w}_m^H \mathbf{H}_{\mathrm{tot},m} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}^1. \tag{50}$$

Let us concatenate now $\mathbf{h}_m^{\mathrm{eff}}, \forall m$, to obtain $\mathbf{H}^{\mathrm{eff}} \in \mathbb{C}^{M\times M}$ as

$$\mathbf{H}^{\mathrm{eff}} = \left[\left(\mathbf{h}_1^{\mathrm{eff}}\right)^T, \ldots, \left(\mathbf{h}_M^{\mathrm{eff}}\right)^T\right]^T. \tag{51}$$

Furthermore, let us define the matrix $\bar{\mathbf{H}} \in \mathbb{C}^{M\times M}$ as

$$\bar{\mathbf{H}} \triangleq \left[\bar{\mathbf{h}}_1^T, \ldots, \bar{\mathbf{h}}_M^T\right]^T = \left[\left(\mathbf{h}_1^{\mathrm{eff}}\right)^T/\sqrt{\beta_1}, \ldots, \left(\mathbf{h}_1^{\mathrm{eff}}\right)^T/\sqrt{\beta_M}\right]^T, \tag{52}$$

where β_i is defined as

$$\sqrt{\beta_i} \triangleq \|\mathbf{h}_i^{\mathrm{eff}}\|_F, i = 1, \ldots, M. \tag{53}$$

Therefore, $\mathbf{H}^{\mathrm{eff}}$ can be expressed as

$$\mathbf{H}^{\mathrm{eff}} = \mathcal{D}\left(\sqrt{\boldsymbol{\beta}}\right)\bar{\mathbf{H}}, \tag{54}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_M]^T \in \mathbb{R}^{M\times 1}$. Following [4], $\mathbf{F}_{\mathrm{BB}}^2$ is formulated as

$$\mathbf{F}_{\mathrm{BB}}^2 = \bar{\mathbf{H}}^H \left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H + \eta I\right)^{-1}, \tag{55}$$

where η is the regularization parameter that is given by $\eta = M\sigma^2/P_t$. Finally, the matrix $\mathbf{F}_{\mathrm{BB}}$ is constructed as $\mathbf{F}_{\mathrm{BB}} = \mathbf{F}_{\mathrm{BB}}^1\mathbf{F}_{\mathrm{BB}}^2$.

C. OPTIMAL POWER ALLOCATION

It now remains for us to obtain the power allocation vector $\mathbf{p}$, which solves the MMF problem $\mathcal{P}_5$. Therefore, to design $\mathbf{p}$, equation (1) can be expressed as

$$\begin{aligned}\mathbf{y} &= \mathcal{D}\left(\sqrt{\boldsymbol{\beta}}\right)\bar{\mathbf{H}}\bar{\mathbf{H}}^H\left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H + \eta I_M\right)^{-1}\mathcal{D}\left(\sqrt{\mathbf{p}}\right)\mathbf{s} + \mathbf{n} \\ &= \mathcal{D}\left(\sqrt{\boldsymbol{\beta}}\right)\bar{\mathbf{H}}\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}} + \eta I_M\right)^{-1}\bar{\mathbf{H}}^H\mathcal{D}\left(\sqrt{\mathbf{p}}\right)\mathbf{s} + \mathbf{n} \\ &= \mathcal{D}\left(\sqrt{\boldsymbol{\beta}}\right)\hat{\mathbf{H}}\mathcal{D}\left(\sqrt{\mathbf{p}}\right)\mathbf{s} + \mathbf{n},\end{aligned} \tag{56}$$

where $\hat{\mathbf{H}} = \bar{\mathbf{H}}(\bar{\mathbf{H}}^H\bar{\mathbf{H}} + \eta I_M)^{-1}\bar{\mathbf{H}}^H \in \mathbb{C}^{M\times M}$. Therefore, the rate of the mth UE is given by

$$R_m(\mathbf{p}) = \ln\left[1 + \frac{p_m\left|\left[\hat{\mathbf{H}}\right]_{m,m}\right|^2}{\lambda_m(\mathbf{p})}\right], \tag{57}$$

where

$$\lambda_m(\mathbf{p}) \triangleq \sum_{n=1,n\neq m}^{M}\left|\left[\hat{\mathbf{H}}\right]_{m,n}\right|^2 p_n + \frac{\sigma^2}{\beta_m}. \tag{58}$$

The MMF design problem $\mathcal{P}_5$ can be reformulated as

$$\mathcal{P}_{5a}: \max_{\mathbf{p}} \min_{m\in\mathcal{M}} R_m(\mathbf{p}), \quad \text{s.t.} \ \left|\left|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathcal{D}\left(\sqrt{\mathbf{p}}\right)\right|\right|_F^2 \leq P_t. \tag{59}$$

Since (59) is once again non-convex, which arises due to the non-convexity of the OF, we employ the path-following approach. This yields feasible points having an improved OF value in each iteration, thus eventually converging to the optimal solution. Toward this end, consider $\mathbf{p}^{(\kappa)} = [p_1{}^{(\kappa)}, \ldots, p_M{}^{(\kappa)}]^T$ to be a feasible point for (59), which is computed in the $(\kappa-1)$st iteration. Following the inequality given in the Appendix of [4], $R_m(\mathbf{p})$ can be lower-bounded as

$$\begin{aligned}R_m(\mathbf{p}) &\geq R_m^{(\kappa)}(\mathbf{p}) \\ &\triangleq R_m\left(\mathbf{p}^{(\kappa)}\right) + \frac{p_m^{(\kappa)}}{p_m^{(\kappa)} + \lambda_m\left(\mathbf{p}^{(\kappa)}\right)}\left(2 - \frac{p_m^{(\kappa)}}{p_m} - \frac{\lambda_m(\mathbf{p})}{\lambda_m\left(p^{(\kappa)}\right)}\right),\end{aligned} \tag{60}$$

where the function $R_m^{(\kappa)}(\mathbf{p})$ is concave in nature. Therefore, to obtain the next feasible point $\mathbf{p}^{(\kappa+1)} = [p_1{}^{(\kappa+1)}, \ldots, p_M{}^{(\kappa+1)}]^T$ of $\mathcal{P}_{5a}$, we solve the following convex optimization problem

$$\mathcal{P}_{5b}: \max_{\mathbf{p}} \min_{m\in\mathcal{M}} R_m^{(\kappa)}(\mathbf{p}), \quad \text{s.t.} \ \left|\left|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathcal{D}\left(\sqrt{\mathbf{p}}\right)\right|\right|_F^2 \leq P_t. \tag{61}$$

Algorithm 2 Hybrid Transceiver Design at BS

Input: $\mathbf{H}_{\mathrm{D},m}$, $\mathbf{G}_m$, $\mathbf{H}_{\mathrm{R}}$, Φ^{opt}, $\forall m$

1: Obtain $\widetilde{\mathbf{H}}_{\mathrm{tot},m} = \mathbf{H}_{\mathrm{D},m} + \mathbf{G}_m \Phi^{\mathrm{opt}} \mathbf{H}_{\mathrm{R}}$
2: $[\widetilde{\mathbf{U}}_{\mathrm{tot},m} \widetilde{\mathbf{\Sigma}}_{\mathrm{tot},m} \widetilde{\mathbf{V}}^H_{\mathrm{tot},m}] = \mathrm{SVD}(\widetilde{\mathbf{H}}_{\mathrm{tot},m})$
3: $\mathbf{f}_m^{\mathrm{opt}} = [\widetilde{\mathbf{V}}_m]_{(:,1)}$ and $\mathbf{w}_m^{\mathrm{opt}} = [\widetilde{\mathbf{U}}_m]_{(:,1)}$
4: Design the RF precoder $\mathbf{F}_{\mathrm{RF}}$ using (46)
5: Obtain the combiner $\mathbf{w}_m, \forall m$, from (47)
6: Evaluate the front end of BB precoder $\mathbf{F}_{\mathrm{BB}}^1$ using (49)
7: Obtain the back end of BB precoder $\mathbf{F}_{\mathrm{BB}}^2$ using (55)
8: Set $\mathbf{F}_{\mathrm{BB}} = \mathbf{F}_{\mathrm{BB}}^1 \mathbf{F}_{\mathrm{BB}}^2$, $\kappa = 0$ and initialize $\mathbf{p}^{(0)}$ to the feasible value.
9: **repeat**
10: Solve the convex problem (61) to generate $\mathbf{p}^{\kappa+1}$
11: $\kappa = \kappa + 1$
12: **until** the objective function in (59) converges

Output: $\mathbf{w}_m, \forall m, \mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \mathbf{p}$

Note that we have $\min_{m\in\mathcal{M}} R_m^{\kappa}(\mathbf{p}^{(\kappa)}) \geq \min_{m\in\mathcal{M}} R_m^{(\kappa)}(\mathbf{p}^{(\kappa+1)})$ since $\mathbf{p}^{\kappa}, \mathbf{p}^{\kappa+1}$ represent a feasible point and the optimal solutions of (61), respectively. It follows that

$$\min_{m\in\mathcal{M}} R_m\left(\mathbf{p}^{(\kappa+1)}\right) \geq \min_{m\in\mathcal{M}} R_m^{(\kappa)}\left(\mathbf{p}^{(\kappa+1)}\right) > \min_{m\in\mathcal{M}} R_m^{(\kappa)}\left(\mathbf{p}^{(\kappa)}\right) = \min_{m\in\mathcal{M}} R_m\left(\mathbf{p}^{(\kappa)}\right). \quad (62)$$

The quantity $\mathbf{p}^{(\kappa+1)}$ is an improved feasible point for (59) when compared to $\mathbf{p}^{(\kappa)}$. The various steps of the proposed hybrid transceiver design relying on the MMF problem are succinctly summarized in Algorithm 2.

IV. SIMULATION RESULTS

This section presents the simulation results of our proposed scheme and compares its performance to that of the DRIS-HBF scheme presented in [20], as well as to the AF-based FDR-HBF, HDR-HBF schemes proposed in [42]. Finally, it also benchmarks them against the NDRIS-FDB, FDB random RIS where the phase shift of each RE is randomly selected from $[0, 2\pi)$ and FDB without RIS. We consider an RIS-aided downlink MU mmWave MIMO system operating at 28 GHz. A UPA of dimension $N_{\mathrm{t_x}} \times N_{\mathrm{t_y}} = N_{\mathrm{t}}$ is considered at the BS located at (0m, 0m) and serving M UEs each having a single RF chain. At the BS, we fix the number of RF chains to $M_{\mathrm{t}} = M$. The RIS has N elements with a structure of $N_{\mathrm{x}} \times N_{\mathrm{y}}$ units, where $N = N_{\mathrm{x}}N_{\mathrm{y}}$. The RIS is situated at (d_{RIS}m, 10m) and the UEs are randomly and uniformly distributed within a circle centred at (100m, 0m) having a radius of 10m, as shown in Figure 2. To compare the performance of the proposed NDRIS-HBF scheme to that of the FDR- and HDR-HBF schemes, both the FDR and HDR are considered to be equipped with N_{t} antennas and M_{t} RF chains, and are deployed similar to the RIS in Figure 2. HBF is performed at the FDR/HDR to design the RF RC, and RF and BB TPCs as discussed in [42], followed by optimal power allocation based on the path-following algorithm to maximize the rate of the worst-case UE. For the mmWave MIMO channel, the coefficients $\alpha_{i,l}$ are distributed independently obeying the distribution as $\mathcal{CN}(0, \gamma_i^2 10^{-0.1PL(d_i)}), \forall l = \{1, \ldots, N_i^p\}$, where $\gamma_i = \sqrt{\mathrm{row}(\mathbf{H}_i)\mathrm{col}(\mathbf{H}_i)/N_i^p}$ denotes the normalization factor. The quantity $PL(d_i)$ is the path-loss that depends on the distance d_i associated with the corresponding link, which is modeled as [46]

$$PL(d_i)[\mathrm{dB}] = \alpha + 10\beta \log_{10}(d_i) + \zeta, \quad (63)$$

where $\zeta \in \mathcal{CN}(0, \sigma_\zeta^2)$. At the carrier frequency of 28 GHz, the parameters of (63) are: $\alpha = 61.4$, $\beta = 2$, $\sigma_\zeta = 5.8$dB for line of sight (LoS), and $\alpha = 72.0$, $\beta = 2.92$, $\sigma_\zeta = 8.7$dB for non line of sight (NLoS) paths [46]. To characterize the RIS, we assume that each path in the direct link spanning from the BS to the UEs is of NLoS nature, passing through tinted glass walls to experience an additional penetration loss of 40.1dB [47]. Moreover, we set the number of propagation paths to $N_i^{\mathrm{p}} = 10, \forall i$, with an angular spread of 10 degrees, similar to [20], [21]. The azimuth and elevation angles of departure and arrival follow the Laplacian distribution around the mean angle. The antenna spacing at the BS and each UE is set to half-wavelength, i.e., $d_{\mathrm{t}} = d_{\mathrm{r}} = \frac{\lambda}{2}$. The noise variance σ^2 at each UE is set to -91dBm. The simulation results are averaged over 500 independent channel realizations. The SNR is defined as $\mathrm{SNR} = \frac{P_{\mathrm{t}}}{M\sigma^2}$, and its range is varied from -10 dB to 20 dB to study the performance in both the low- as well as high-SNR region. The performance of the proposed and existing techniques are compared both in terms of the worst-case UE rate and EE, as discussed next.

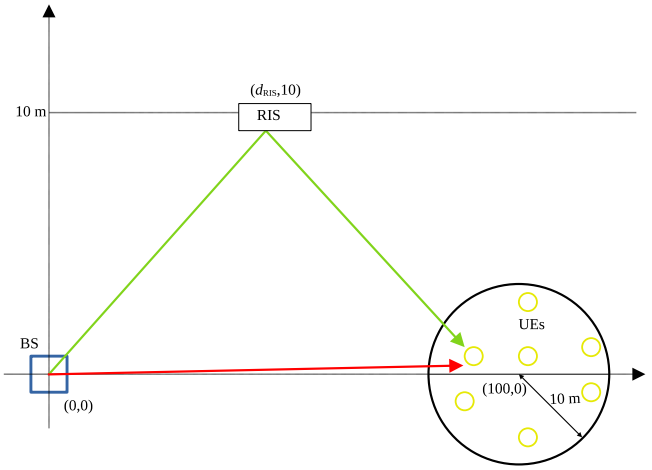

FIGURE 2. Simulation setup for the RIS-aided communication.

A. WORST-CASE UE RATE

Figure 3 shows the rate of the worst-case UE achieved by an 8×128 system, where a BS having $N_{\mathrm{t}} = 8 \times 16 = 128$ antennas is communicating with $M = 4$ UEs

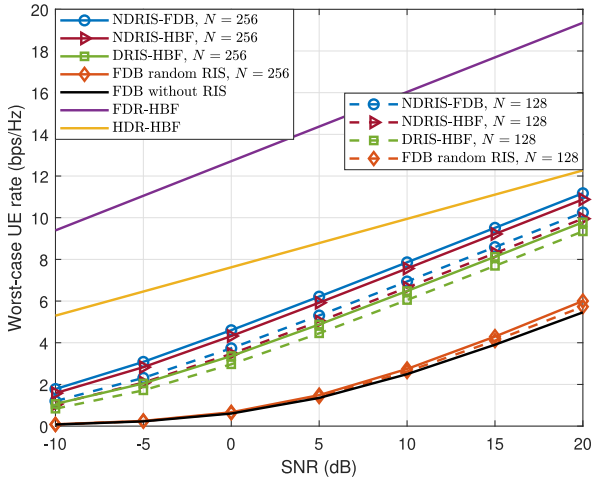

FIGURE 3. Worst-case UE rate versus SNR in an 8 × 128 RIS-aided MU mmWave MIMO system with $M = 4$.

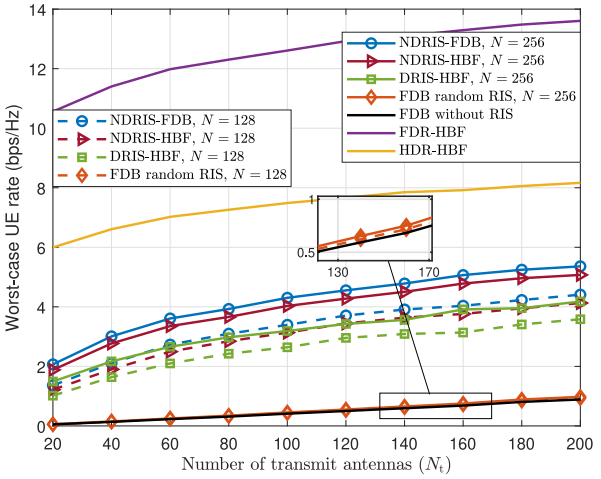

FIGURE 4. Worst-case UE rate versus number of transmit antennas N_t at the BS in an RIS-aided MU mmWave MIMO system with $M = 4$ at SNR = 0 dB.

each having $N_r = 2 \times 4 = 8$ antennas. We compare the results for the $8 \times 16 = 128$ and $16 \times 16 = 256$ structures at the RIS. It can be readily observed from the figure that the worst-case UE rate increases with the SNR. The performance of the proposed NDRIS-HBF scheme approaches that of the optimal NDRIS-FDB, and it is much improved in comparison to the DRIS-HBF. This is due to the high beamforming gain of NDRIS in comparison to DRIS, which was supported by our mathematical analysis in Section II-C. Furthermore, the NDRIS-HBF scheme outperforms its counterparts, viz., FDB random RIS and FDB without an RIS, which shows that in order to obtain performance gains, it is important to jointly design the TPC and RIS phase shift matrix carefully. Furthermore, the RIS-aided system outperforms its counterpart operating without an RIS, which shows the benefits of passive beamforming. This result clearly demonstrates that the proposed joint active HBF and passive beamformer designs are eminently suitable for the RIS-aided mmWave MIMO system, since the NDRIS-HBF has a much lower hardware cost than the NDRIS-FDB. Moreover, since FDR/HDR systems employ active terminals, they outperform the RIS-based system - regardless of the specific RIS structure. However, the proposed NDRIS-HBF scheme associate with $N = 256$ achieves a performance close to that of the HDR-HBF scheme at high SNRs, which is a benefit of the high passive beamforming gain.

Figure 4 shows the worst-case UE rate versus the number of TAs N_t at the BS (FDR/HDR) for $M = 4$ and SNR $= 0$ dB for $N = 128$ and 256. Observe that as expected, the rate increases upon increasing N_t due to the increased multiplexing gain. A RIS with $N = 256$ elements performs better than $N = 128$ due to the higher passive beamforming gain of the former. Furthermore, the proposed NDRIS-HBF preforms better than the DRIS-HBF. Note that upon increasing N_t, the rate of the worst-case UE with RIS becomes much higher than without.

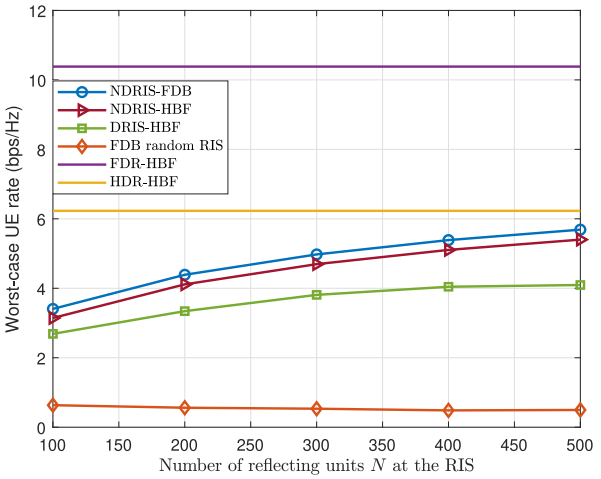

FIGURE 5. Worst-case UE rate versus number of REs N at RIS in an RIS-aided MU mmWave MIMO system with $M = 4$ at SNR = 0 dB.

Figure 5 shows the rate of the worst-case UE versus the number of reflective units N at the RIS for $N_t = 128$, $M = 4$ and SNR $= 0$ dB. Observe that the worst-case UE rate of the proposed HBF scheme for an NDRIS structure approaches that of the optimal FDB that also employs an NDRIS. The performance gap of both with respect to the DRIS structure widens upon increasing N. This is attributed to the high passive beamforming gain achieved by the NDRIS in comparison to the DRIS at high N. Furthermore, at large N, the rate of the DRIS-HBF scheme saturates, while that of the NDRIS-HBF continues to increase, which shows the effectiveness of our scheme. Note that the NDRIS-HBF rate approaches the HDR-HBF rate for large N even for low SNR values, which can be attributed to the large passive beamforming gain of the former. This demonstrates that the NDRIS approaches the performance of an HDR system via passive beamforming, albeit at much lower cost and power. On the other hand, the performance of FDB random RIS does not improve monotonically with N,

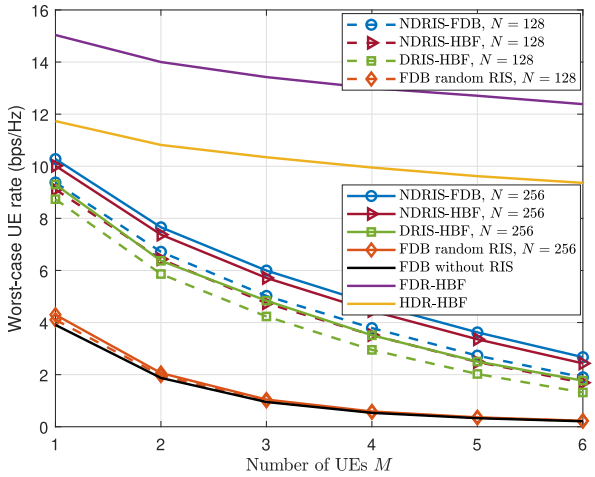

FIGURE 6. Worst-case UE rate versus number of UEs M in an RIS-aided MU mmWave MIMO system at SNR = 0 dB.

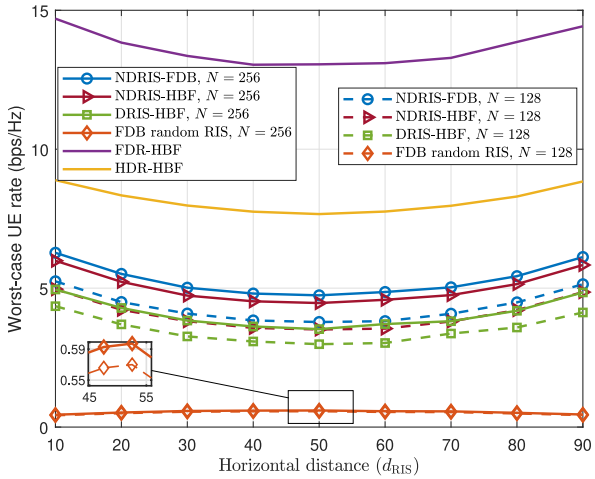

FIGURE 7. Worst-case UE rate versus d_{RIS} in an RIS-aided MU mmWave MIMO system with $M = 4$ at SNR = 0 dB.

indicating that irrespective of the number of REs used at the NDRIS, judicious design of the NDRIS phase shift matrix is necessary to attain a higher worst-case UE rate.

Figure 6 shows the rate of the worst-case UE versus the number of UEs M, which are assumed to be distributed uniformly in the circle of radius 10 m. A fixed SNR of 0 dB is set for this analysis. Observe that as M increases, the worst-case UE rate decreases, which is due to the increased MUI and reduced transmit power per UE. Also, the proposed scheme performs better than a DRIS-HBF for both $N = 256$ and $N = 128$. Note that the rate of the worst-case UE without RIS is poor, which justifies the importance of RIS systems in MU communication. Furthermore, the FDR/HDR-HBF schemes perform better than the proposed NDRIS-HBF scheme due to their improved ability to suppress MUI, jointly by the BS and the relay.

Figure 7 shows the worst-case UE rate versus the horizontal distance d_{RIS}. It can be seen that as d_{RIS} increases, the rate decreases initially and achieves its minimum value at 50 m, following which it increases again. This is due to the fact that the power received via RIS in the far field region is proportional to $d_{\text{BS-RIS}}^{-2} d_{\text{RIS-UE}}^{-2}$, where $d_{\text{BS-RIS}}$ and $d_{\text{RIS-UE}}$ denote the distances from the BS to RIS and RIS to UE, respectively. Hence, it can be concluded that the RIS has to be placed near the BS or UEs for a higher rate.

B. ENERGY EFFICIENCY (EE)

In this subsection, we evaluate the EE of the system. Let us determine the power consumed by the HBF-DRIS, HBF-NDRIS and FDB-NDRIS modules in our system. The total power consumption $P_{\text{DRIS}}^{\text{HBF}}$ of the proposed system model using an HBF at the BS and DRIS phase shift matrix at the RIS is given by [10], [48], [49]

$$P_{\text{DRIS}}^{\text{HBF}} = \eta P_{\text{t}} + M_{\text{t}} P_{\text{RF}} + P_{\text{cir}} + N P_n(b), \tag{64}$$

where P_{t} is the total power radiated by the BS for downlink communication with $\eta \triangleq \lambda^{-1}$ and λ is related to the amplifier efficiency, P_{RF} is the power consumed by each RF chain, P_{cir} is the total power consumption in the circuit, and $P_n(b)$ is the power consumption of each RIS element for b bits of phase resolution. The typical power consumed by each phase shifter is $1.5, 4.5, 6$ and 7.8mW for $3-, 4-, 5-$ and $6-$bit resolution [41], [50].

Furthermore, the power $P_{\text{NDRIS}}^{\text{HBF}}$ consumed by an HBF at the BS and NDRIS structure at the RIS is given by

$$P_{\text{NDRIS}}^{\text{HBF}} = \eta P_{\text{t}} + M_{\text{t}} P_{\text{RF}} + P_{\text{cir}} + N(P_n(b) + P_{\text{SW}}), \tag{65}$$

where the additional term P_{SW} represents the power consumed per switch in reconfiguration of the switches to perform the mapping function f. It is worth noting that the switches used at the RIS are MEMS switches that are widely used in wireless communication, with the typical power required for a single switch being 5mW [51].

In the proposed NDRIS scheme with FDB for the given system model, where each antenna at the BS is connected to a single RF chain, the power requirement can be expressed as

$$P_{\text{NDRIS}}^{\text{FDB}} = \eta P_{\text{t}} + N_{\text{t}} P_{\text{RF}} + P_{\text{cir}} + N(P_n(b) + P_{\text{SW}}). \tag{66}$$

Figure 8 depicts the EE versus SNR of an 8×128 system for $M = 4$, $P_n(b) = 6$ mW and $b = 5$ bits of phase resolution. As seen from the figure, for large N, the EE of the system increases, which implies that one can deploy an NDRIS with a large number of REs to increase the EE. It can also be seen that our proposed NDRIS system has a higher EE than the DRIS system. This is due to the fact that the higher sum-rate of the NDRIS-HBF system dominates the additional power required for reconfiguration of the phase shifts. Furthermore, the EE of the NDRIS-FDB and FDB random RIS schemes is much lower than the proposed NDRIS-HBF scheme due to the requirement of a large number of power-hungry RF chains in the FDB and the lack of judicious phase shift matrix design in the random RIS scheme. This demonstrates the importance of both TPC and RIS phase shift matrix design in RIS-aided mmWave

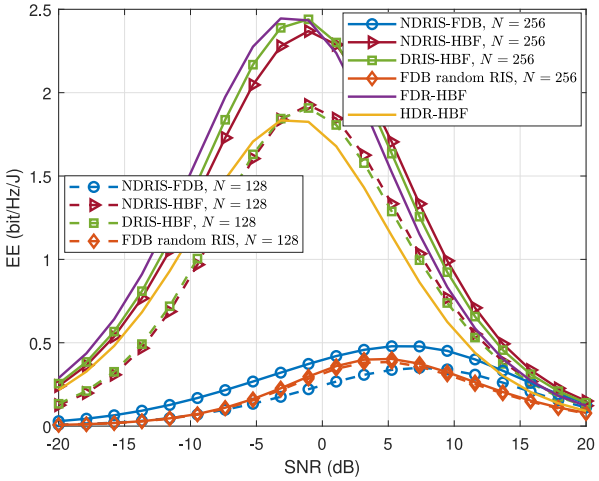

FIGURE 8. EE versus SNR in an RIS-aided MU mmWave MIMO system with $M = 4$.

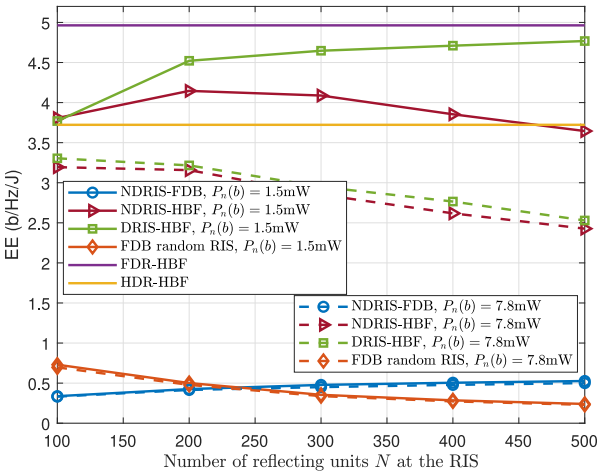

FIGURE 9. EE versus number of REs N in an RIS-aided MU mmWave MIMO system with $M = 4$ at SNR = 0 dB.

MIMO systems. One can also observe that the EE of the proposed scheme with $N = 256$ is significantly higher than that of an HDR system. Furthermore, it approaches that of the FDR-HBF scheme at low SNR, and in fact performs better at high SNR. This superior performance is due to the high passive gain of the NDRIS coupled with the lower power consumption of its semi-passive components. Note that the EE of the system first increases with the SNR, achieves its maximum, and subsequently decreases due to the reduced amplifier-efficiency at high SNR. Hence, there is a specific value of transmit power for each scheme at which the EE of the system is maximum.

We finally plot the EE versus N of an 8×128 system with $M = 4$ and SNR $= 0$ dB, $P_n(b) = 1.5$ mW and $P_n(b) = 7.8$ mW in Figure 9. It can be observed that for $N < 200$, the EE increases with N for all the schemes. However, for $N > 200$, the EE starts decreasing for all the schemes except for the DRIS-HBF with $P_n(b) = 1.5$mW and optimal NDRIS-FDB. This is due to the fact that $P_n(b) = 1.5$ mW per RIS element is extremely low, which necessitates a large N to observe a noticeable increment in the power consumption. Note that for a given $P_n(b)$ and SNR, there is an optimal value of the number of reflective elements N for which the EE is maximum. Hence, a large N may increase the sum-rate of the system, but can also lead to a reduction in the EE of the system.

V. CONCLUSION

A framework was presented for the design of the hybrid beamformer/combiner and passive beamformer at the NDRIS in a MU mmWave MIMO system. An MMF problem was formulated for optimizing the above design, which is solved using the principle of alternating optimization. The NDRIS is determined using the ADMM algorithm, while the RF TPC, the front end BB TPC, back-end TPC are designed using the Karcher mean, least squares and RZF techniques, respectively. Furthermore, the path-following algorithm was employed for optimal power allocation to maximize the rate of the worst-case user. Our results demonstrate that the rate of the worst-case UE approaches that of its the optimal fully-digital counterpart, while performing close to the HDR-HBF scheme for a large number of RIS elements with the aid of the design procedure proposed for our NDRIS-based mmWave MIMO system. Furthermore, the EE of the NDRIS-HBF scheme is observed to be higher than that of the conventional DRIS-HBF, HDR-HBF and the FDR-HBF systems. This demonstrates that the former scheme can be beneficially employed in practical RIS-aided mmWave systems.

REFERENCES

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[2] A. Ghosh et al., "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[3] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, 2015.

[4] A. A. Nasir, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Hybrid beamforming for multi-user millimeter-wave networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2943–2956, Mar. 2020.

[5] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 870–913, 2nd Quart., 2018.

[6] J. Singh, I. Chatterjee, S. Srivastava, and A. K. Jagannatham, "Hybrid transceiver design and optimal power allocation in downlink mmWave hybrid MIMO cognitive radio systems," in *Proc. Nat. Conf. Commun. (NCC)*, 2022, pp. 178–183.

[7] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[8] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Dec. 2020.

[9] H. Xie, J. Xu, and Y.-F. Liu, "Max-min fairness in IRS-aided multi-cell MISO systems with joint transmit and reflective beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1379–1393, Feb. 2021.

[10] 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.

[11] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, May 2020.

[12] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.

[13] A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, Dec. 2020.

[14] Q. Wang, C. Xing, C. Du, L. Zhao, and L. Hanzo, "Hybrid nonlinear transceiver optimization for the RIS-aided MIMO downlink," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6441–6455, Oct. 2022.

[15] M. Fu, Y. Zhou, Y. Shi, and K. B. Letaief, "Reconfigurable intelligent surface empowered downlink non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3802–3817, Jun. 2021.

[16] M. Di Renzo et al., "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 798–807, 2020.

[17] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.

[18] Q. Gu, D. Wu, X. Su, J. Jin, Y. Yuan, and J. Wang, "Performance comparisons between reconfigurable intelligent surface and full/half-duplex relays," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, 2021, pp. 1–6.

[19] P. Wang, J. Fang, L. Dai, and H. Li, "Joint transceiver and large intelligent surface design for massive MIMO mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1052–1064, Feb. 2021.

[20] R. Li, B. Guo, M. Tao, Y.-F. Liu, and W. Yu, "Joint design of hybrid beamforming and reflection coefficients in RIS-aided mmWave MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2404–2416, Apr. 2022.

[21] S. H. Hong, J. Park, S.-J. Kim, and J. Choi, "Hybrid beamforming for intelligent reflecting surface aided millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7343–7357, Sep. 2022.

[22] Q. Ding, X. Gao, and Z. Wu, "Joint resource optimization for IRS-assisted mmWave MIMO under QoS constraints," *IEEE Trans. Veh. Technol.*, vol. 70, no. 11, pp. 12243–12247, Nov. 2021.

[23] Y. Xiu et al., "Reconfigurable intelligent surfaces aided mmWave NOMA: Joint power allocation, phase shifts, and hybrid beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8393–8409, Dec. 2021.

[24] D. Zhao, H. Lu, Y. Wang, H. Sun, and Y. Gui, "Joint power allocation and user association optimization for IRS-assisted mmWave systems," *IEEE Trans. Wireless Commun*, vol. 21, no. 1, pp. 577–590, Jan. 2022.

[25] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1229–1243, Feb. 2022.

[26] Q. Li et al., "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6367–6383, Jun. 2022.

[27] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[28] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng, and L. Hanzo, "Hybrid precoding for wideband millimeter wave MIMO systems in the face of beam squint," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1847–1860, Mar. 2021.

[29] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[30] I. Chatterjee, J. Singh, S. Srivastava, and A. K. Jagannatham, "Frequency selective hybrid beamforming and optimal power loading for multiuser millimeter wave cognitive radio networks," *IEEE Access*, vol. 11, pp. 96052–96067, 2023.

[31] J. Singh, I. Chatterjee, S. Srivastava, A. Agrahari, A. K. Jagannatham, and L. Hanzo, "Hybrid transceiver design and optimal power allocation for the cognitive mmWave multiuser MIMO downlink relying on limited feedback," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 241–256, Jan. 2023.

[32] W. Tang et al., "Path loss modeling and measurements for reconfigurable intelligent surfaces in the millimeter-wave frequency band," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 6259–6276, Sep. 2022.

[33] E. E. Bahingayi and K. Lee, "Low-complexity beamforming algorithms for IRS-aided single-user massive MIMO mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9200–9211, Nov. 2022.

[34] J.-B. Wang, X. Wang, F. Yang, H. Zhang, M. Lin, and J. Wang, "Intelligent reflecting surface aided millimeter wave communication using subarray-connected structure," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5581–5586, May 2022.

[35] M. Cheng, J.-B. Wang, H. Zhang, J.-Y. Wang, M. Lin, and J. Cheng, "Impact of finite-resolution precoding and limited feedback on rates of IRS based mmWave networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5172–5186, May 2022.

[36] B. Ning, P. Wang, L. Li, Z. Chen, and J. Fang, "Multi-IRS-aided multi-user MIMO in mmWave/THz communications: A space-orthogonal scheme," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8138–8152, Dec. 2022.

[37] J. Ye, A. Kammoun, and M.-S. Alouini, "Reconfigurable intelligent surface enabled interference nulling and signal power maximization in mmWave bands," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9096–9113, Nov. 2022.

[38] Y. Chen, Y. Wang, and L. Jiao, "Robust transmission for reconfigurable intelligent surface aided millimeter wave vehicular communications with statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 928–944, Feb. 2022.

[39] K. Ying, Z. Gao, S. Lyu, Y. Wu, H. Wang, and M.-S. Alouini, "GMD-based hybrid beamforming for large reconfigurable intelligent surface assisted millimeter-wave massive MIMO," *IEEE Access*, vol. 8, pp. 19530–19539, 2020.

[40] L. Jiao, P. Wang, A. Alipour-Fanid, H. Zeng, and K. Zeng, "Enabling efficient blockage-aware handover in RIS-assisted mmWave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2243–2257, Apr. 2022.

[41] J. Singh, S. Srivastava, S. P. Yadav, A. K. Jagannatham, and L. Hanzo, "Energy efficiency optimization in reconfigurable intelligent surface aided hybrid multiuser mmWave MIMO systems," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 581–589, Aug. 2023.

[42] N. Song, T. Yang, and H. Sun, "Efficient hybrid beamforming for relay assisted millimeter-wave multi-user massive MIMO," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2019, pp. 1–6.

[43] G. Rebeiz and J. Muldavin, "RF MEMS switches and switch circuits," *IEEE Microw. Mag.*, vol. 2, no. 4, pp. 59–71, Dec. 2001.

[44] G. Lavigne and C. Caloz, "Nonreciprocal phase gradient metasurface: Principle and transistor implementation," in *Proc. 13th Int. Congr. Artif. Mater. Novel Wave Phenomena (Metamaterials)*, 2019, pp. X–212–X–214.

[45] J. Chen, "When does asymptotic orthogonality exist for very large arrays?" in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2013, pp. 4146–4150.

[46] 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, Jun. 2014.

[47] H. Zhao et al., "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York City," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2013, pp. 5163–5167.

[48] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy-efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2370–2382, Oct. 2017.

[49] M. Sefunç, A. Zappone, and E. A. Jorswieck, "Energy efficiency of mmWave MIMO systems with spatial modulation and hybrid beamforming," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 1, pp. 95–108, Mar. 2020.
[50] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen, "Energy efficient multi-user MISO communication using low resolution large intelligent surfaces," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–6.
[51] R. Méndez-Rial, C. Rusu, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, "Channel estimation and hybrid combining for mmWave: Phase shifters or switches?" in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2015, pp. 90–97.

JITENDRA SINGH (Member, IEEE) received the integrated dual B.Tech. and M.Tech. degrees in electronics and communication engineering with specialization in wireless communication and networks from Gautam Buddha University, Greater Noida, India, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur, India. His research interests include cognitive radio networks, mmWave communication, intelligent reflecting surface, integrated sensing, and communication systems. He is a member of team who won Qualcomm 6G University Research India Program in 2023.

SURAJ SRIVASTAVA (Senior Member, IEEE) received the M.Tech. degree in electronics and communication engineering from the Indian Institute of Technology Roorkee, India, in 2012, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology Kanpur, Kanpur, India, in 2022. From July 2012 to November 2013, he was employed as a Staff-I Systems Design Engineer with Broadcom Research India Pvt. Ltd., Bengaluru, and from November 2013 to December 2015, he was employed as a Lead Engineer with Samsung Research India, Bengaluru, where he worked on developing layer-2 of the 3G UMTS/WCDMA/HSDPA modem. His research interests include applications of sparse signal processing in 5G wireless systems, mmWave and terahertz communication, orthogonal time-frequency space, joint radar and communication, and optimization and machine learning. He was awarded the Outstanding Ph.D. Thesis and Outstanding Teaching Assistant Awards from IIT Kanpur. He was awarded Qualcomm Innovation Fellowship from Qualcomm in 2018 and 2022.

ADITYA K. JAGANNATHAM (Senior Member, IEEE) received the bachelor's degree from the Indian Institute of Technology Bombay, Mumbai, and the M.S. and Ph.D. degrees from the University of California at San Diego, La Jolla, CA, USA. From April 2007 to May 2009, he was employed as a Senior Wireless Systems Engineer with Qualcomm Inc., San Diego, where he was a part of the Qualcomm CDMA Technologies Division. He is currently a Professor with the Department of Electrical Engineering, IIT Kanpur, where he also holds the Arun Kumar Chair Professorship. His research interests include next generation wireless cellular and WiFi networks, with a special emphasis on various 5G technologies, such as massive MIMO, mmWave MIMO, FBMC, NOMA, as well as emerging 6G technologies, such as OTFS, IRS, THz systems, and VLC. He has been twice awarded the P. K. Kelkar Young Faculty Research Fellowship for excellence in research, received multiple Qualcomm Innovation Fellowships in 2018 and 2022, the IIT Kanpur Excellence in Teaching Award, the CAL(IT)2 Fellowship at the University of California at San Diego, the Upendra Patel Achievement Award at Qualcomm San Diego, and the Qualcomm 6G UR India Gift.

LAJOS HANZO (Life Fellow, IEEE) received the first Honorary Doctorate degree from the Technical University of Budapest in 2009 and the second Honorary Doctorate degree Edinburgh University in 2015. He holds the IEEE Eric Sumner Technical Field Award. He is a Foreign Member of the Hungarian Science-Academy and a Fellow of the Royal Academy of Engineering, IET, and EURASIP. For more information, see https://en.wikipedia.org/wiki/Lajos_Hanzo.