The Reconfigurable Intelligent Surface-Aided Multi-Node IoT Downlink: Beamforming Design and Performance Analysis

Qingchao Li, Mohammed El-Hajjar, *Senior Member, IEEE*, Ibrahim Hemadeh, *Member, IEEE*, Deepa Jagyasi, Arman Shojaeifard, *Senior Member, IEEE*, Ertugrul Basar, *Fellow, IEEE*, Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Reconfigurable intelligent surfaces (RIS) are capable of enhancing the wireless propagation environment of the future Internet of Things (IoT). Recently, they have also been configured as a transmitter to realize information modulation at low hardware complexity. In this paper, we conceive a transmitter relying on a single radio frequency (RF) chain for low-complexity RIS-aided multi-user downlink communication. More explicitly, in the proposed architecture, the multi-user information is transmit precoded and modulated at the RIS by appropriately configuring the phase shift and amplitude of each RIS element. We assume that the distribution of multiple users obeys on a Poisson point process (PPP), where we jointly optimize the total power reflected from the RIS and the power allocation fraction assigned to each user, under the practical constraint of a realistic amplitude limitation of each RIS element. Additionally, we theoretically analyse the ergodic rate, symbol error probability, outage probability and coverage range of the proposed RIS-aided single-RF downlink and confirm the accuracy of our analysis by simulations. Finally, we compare its performance to that of the conventional multiple-input-multiple-output (MIMO) systems employing multiple RF-chains.**

***Index Terms*—Reconfigurable intelligent surfaces (RIS), Internet of Things (IoT) networks, passive modulation, power allocation, alternating optimization (AO).**

I. Introduction

THE fifth-generation (5G) systems are being rolled out across the globe and research is well under way on massive multiple-input-multiple-output (MIMO) techniques [1], millimeter wave (mmWave) communications [2] and ultra-dense networking (UDN) [3]. However, driven by the ever-increasing demands of the Internet of Things (IoT), research has turned to the exploration of next generation concepts [4].

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). E. Basar was supported by TUBITAK under Grant 120E401. *(Corresponding author: Lajos Hanzo.)*

Qingchao Li, Mohammed El-Hajjar and Lajos Hanzo are with the Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: Qingchao.Li@soton.ac.uk; meh@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Ibrahim Hemadeh, Deepa Jagyasi, Arman Shojaeifard are with InterDigital, London EC2A 3QR, U.K. (e-mail: Ibrahim.Hemadeh@InterDigital.com; Deepagurmukhdas.Jagyasi@InterDigital.com; Arman.Shojaeifard@InterDigital.com).

Ertugrul Basar is with the Communications Research and Innovation Laboratory (CoreLab), Department of Electrical and Electronics Engineering, Koç University, Sariyer, 34450 Istanbul, Turkey (e-mail: ebasar@ku.edu.tr).

To intensify the exploitation of the spatial dimension, ultra-massive MIMO (UM-MIMO) techniques have been proposed, where thousands of antennas are employed at transceivers to support rates in excess of 1 Terabit per second (Tbps) in support of the IoT [5]. However, a large number of radio frequency (RF) chains is required by UM-MIMO systems, which would impose excessive hardware complexity and cost. To tackle this challenge, reconfigurable intelligent surfaces (RIS) have been advocated as a benefit of their energy-efficiency and cost-efficiency [6]–[11].

Briefly, the RIS is comprised of a large number of passive reflecting elements, which can intelligently control the phase shift and even the amplitude of the impinging signals [12]. As a benefit, RISs have found a wide range of applications in IoT networks, such as joint beamforming for improving the systems' reception reliability [13], constructing physical layer security networks [14], assisting the communication of unmanned aerial vehicles (UAVs) [15]–[17], and localization, positioning, sensing as well as mobile edge computing [18]. Furthermore, it was also shown that RISs are capable of acting as a transmitter to modulate signals at a lower hardware complexity than massive MIMO systems [19]–[27]. Specifically, in massive MIMO systems, a large number of RF-chains are required to modulate signals, while in the RIS-aided transmitter scheme of [19]–[27] information is modulated on the passive RIS by appropriately configuring the reflection coefficient of each element, resulting in an energy- and cost-efficient hardware architecture.

In [19], Basar *et al.* proposed a RIS-aided single-user transmitter scheme, where the RIS is empowered by an unmodulated carrier generated from a low-cost RF, which contains a RF digital-to-analog converter with an internal memory and a power amplifier, where the information is conveyed via controlling the phase shift of each passive RIS element. Given the channel state information (CSI), the phase of all RIS elements can be adjusted for maximizing the received signal-to-noise ratio (SNR), where M-level phase shifts can be imposed on the signals reflected from all RIS elements to create a M-ary phase shift keying (PSK) signal constellation. As a further advance [20], Basar *et al.* employed a similar system setup, in which a so-called blind access point-RIS modulation scheme is employed for RIS-aided single-user transmitter without any need for CSI. A binary phase shift, i.e. 0 and π, is imposed on all RIS elements to create a binary phase shift keying (BPSK) constellation. Compared to

[19], the blind access point-RIS modulation scheme cuts down the channel estimation overhead, albeit at the cost of some performance erosion. In [21], Khaleel and Basar employed RIS to realize Alamouti's scheme, where an unmodulated carrier generated by a single RF empowers the RIS. Specifically, the RIS is divided into two blocks, where Alamouti's scheme is used based on configuring the phase shift of the RIS elements. It was demonstrated that in line with Alamouti, a transmit diversity order of two is achievable by the proposed RIS-aided modulation scheme. In [22], Basar constructed an RIS-based index modulation scheme, where the RIS is employed between a single RF-chain and a multi-antenna receiver. A pair of transmitter methods was proposed, namely RIS-aided space shift keying (SSK) and RIS-aided spatial modulation (SM). In the RIS-aided SSK, the signals radiated from the RF-chain are unmodulated and information is only conveyed on the specific receiver antenna. The phase shift of each RIS element is configured for realizing passive beamforming from the RIS to the selected receiver antenna. In contrast to RIS-aided SM, the signal in the single RF-chain is modulated and information is conveyed via both the modulated signals on the one and only RF-chain and by the specific index of the selected receiver antenna. At the receiver, a greedy detector and a maximum likelihood (ML) detector are separately employed for recovering the original information. To increase the throughput, Yuan *et al.* [23] proposed a so-called RIS-aided receive quadrature reflecting modulation (RIS-RQRM) scheme, where the whole RIS is divided into two halves for processing the in-phase and quadrature components, respectively. The information is conveyed via each half of the RIS to form a directional beam to a specific receiver antenna.

Tang *et al.* [24]–[26] reported on a range of practical experiments for confirming the beneficial application prospects of RIS-aided modulations, given its energy and cost-effective hardware architecture. In [24], a single-user RIS-aided 8-PSK architecture is proposed for configuring the phase shift of 8×32 RIS elements, achieving a 6.144 Mbps data rate at 4.25 GHz carrier frequency. In [25], the authors firstly designed quadrature amplitude modulation (QAM) based on independently controlling the amplitude and phase shift of each RIS element by introducing a non-linear modulation technique under the constraint of a constant envelope. In [26], a real-time RIS-based Alamouti space-time scheme is realized.

However, the above RIS-aided modulation schemes are designed for single-user communication scenarios. By contrast, Liu *et al.* [27] employed RIS as a transmitter to deliver information to multiple users, where a single-RF signal generator provides energy for the passive RIS, and the information is mapped to the phase shift of the RIS elements. When the number of users is K and M-ary information is transmitted in each time slot, there are M^K possible symbol vectors. A symbol-level transmit precoding (TPC) method is employed, which finds the optimal M^K RIS phase shift patterns to deliver information to K users. The Riemannian conjugate gradient algorithm and branch-and-bound algorithm are employed for optimizing the RIS phase shift patterns for minimizing the symbol error probability (SEP) of multi-bit RIS phase shift resolution and that of single-bit RIS phase shift resolution, respectively. However, since the number of RIS phase shift patterns increases exponentially with the number of users, the symbol-level TPC method only remains practically feasible for a small number of users. Furthermore, no power allocation algorithms may be readily employed for this system model to ensure a particular quality of service (QoS) target and/or fairness.

Against this backdrop, we propose an RIS-aided single-RF downlink transmitter architecture for supporting multiple users, where we apply the zero-forcing (ZF) optimization criterion for our TPC scheme. Additionally, we design the power allocation for the different users by optimizing the sum-rate, the min-rate and the geometric-mean-rate. In our proposed design, a single RF-chain is used at the RIS, and the baseband signal is modulated at the RIS by configuring the phase shifts and amplitudes of the RIS reflecting elements. Additionally, we compare our theoretical analysis and simulation results. Explicitly, our novel contributions can be summarized as follows:

- We propose a RIS-aided single-RF downlink transmitter design for supporting multiple users, where the multi-user information is modulated at the RIS by configuring the reflection phase shift and amplitude of each element. The proposed design deploys a single RF chain to communicate with multiple users, where the information is modulated based on the configuration of the low-complexity RIS elements. Hence, our proposed scheme has reduced the hardware cost at high energy efficiency.
- We consider a multiple user scenario with the users distributed according to a Poisson point process, and design the power allocation algorithms by jointly optimizing the total power reflected from the RIS and the power allocation ratio of each user, under the constraint that the amplitude of each passive RIS element must be smaller than or equal to Unity. The system throughput and the user rate fairness can be realized by adopting different power allocation techniques.
- We provide both the theoretical analysis and simulation results for characterizing the performance of our proposed RIS-aided single-RF transmitter architecture, including its ergodic rate, symbol error probability, outage probability and its coverage range. Furthermore, we analytically determine the number of RIS elements required by our single-RF scheme for achieving the same channel capacity as that of the conventional MIMO systems employing multiple RF-chains. Explicitly, our results show that the proposed single-RF chain RIS-based transmitter achieves a comparable system performance to the conventional fully-digital MIMO using multiple RF chains, while employing low-cost passive RIS elements.

Finally, Table I explicitly contrasts our contributions to the literature.

The rest of this paper is organized as follows. In Section II, we present the system model, while our power allocation methods are described in Section III. Section IV presents the theoretical analysis of our proposed RIS-aided transmitter scheme and contrast it to conventional MIMO systems.

TABLE I
NOVELTY COMPARISON WITH THE LITERATURE.

	Our paper	[12]	[13]	[14]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]
RIS-based modulation	✔					✔	✔	✔	✔	✔	✔	✔	✔	✔
Single RF-chain	✔					✔	✔	✔	✔	✔	✔	✔	✔	✔
Multi-user	✔	✔	✔	✔	✔									✔
Precoding	✔	✔	✔	✔	✔									
Correlated channel model	✔													
Power allocation	✔	✔	✔											
Sum-rate maximization	✔	✔	✔											
Min-rate maximization	✔													
Geometric-mean-rate maximization	✔	✔												

Our simulation results are presented in Section V, while we conclude in Section VI.

Notations: Vectors and matrices are denoted by boldface lower and upper case letters, respectively, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ represent the operation of transpose and Hermitian transpose, respectively, $\mathbb{C}^{m\times n}$ denotes the space of $m\times n$ complex-valued matrices, $\mathrm{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal elements being the elements of $\mathbf{a}$ in order, $\mathbf{I}_n$ represents the $n\times n$ identity matrix, $\mathbf{0}_n$ represents the $n\times 1$ zero vector, $|a|$ (or $|\mathbf{a}|$) represent the amplitude of the complex scalar a (or complex vector $\mathbf{a}$), $\|\mathbf{A}\|_\infty$ denotes the infinity norm of matrix $\mathbf{A}$, $f_X(x)$ and $F_X(x)$ are the probability density function (PDF) and cumulative distribution function (CDF) of a random variable X, respectively, a circularly symmetric complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted as $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\mathbb{E}[X]$ represents the mean of the random variable X.

II. SYSTEM MODEL

The proposed RIS-assisted downlink model is illustrated in Fig. 1, where a single RF-chain generates an unmodulated sine carrier wave of power E_s and frequency f_c, for illuminating the N-element passive RIS. The RIS controller configures the reflection coefficients, including the amplitude and phase shift, of each RIS reflecting element, based on the multi-user information[1]. At the receiver, K single-antenna users are randomly distributed among a circ, obeying the PPP.

Compared to conventional massive MIMO systems, our proposed RIS-aided transmitter system has a significantly reduced hardware cost and complexity. In the conventional massive MIMO systems, the baseband signals emanating from the digital TPC are converted to RF signals by multiple active RF-chains, and then radiated from multiple transmit antennas. By contrast, in our proposed architecture, only a single RF-chain is required for illuminating the RIS, and the multi-user information is passively precoded based on the configuration of RIS elements, yielding a low-complexity MIMO system.

As seen in [19], [22], [23], [27], and in Fig. 1, the RF generator is close to the RIS, and its power can be focused on the RIS by a horn antenna without path-loss. Therefore, the signal $\mathbf{y}\in\mathbb{C}^{K\times 1}$ received by the K users is formulated as:

$$\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{q}=\mathbf{H}\mathbf{P}\boldsymbol{\Lambda}\mathbf{s}+\mathbf{q}, \tag{1}$$

where $\mathbf{H}=[\mathbf{h}_1,\mathbf{h}_2,\cdots,\mathbf{h}_K]^{\mathrm{H}}\in\mathbb{C}^{K\times N}$ with $\mathbf{h}_k^{\mathrm{H}}\in\mathbb{C}^{1\times N}$ being the channel vector of the links spanning from the N RIS elements to the kth user, $\mathbf{q}\sim\mathcal{N}_{\mathbb{C}}(\mathbf{0}_K,\sigma_q^2\mathbf{I}_K)\in\mathbb{C}^{K\times 1}$ is the circularly symmetric complex Gaussian noise, and $\mathbf{x}=\mathbf{P}\boldsymbol{\Lambda}\mathbf{s}\in\mathbb{C}^{N\times 1}$ is the baseband signal mapped to the RIS elements, in which $\mathbf{s}\in\mathbb{C}^{K\times 1}$ is the multi-user information symbol vector, $\mathbf{P}\in\mathbb{C}^{N\times K}$ is the TPC matrix and $\boldsymbol{\Lambda}\in\mathbb{C}^{K\times K}$ is the power allocation matrix, which will be detailed in Section III.

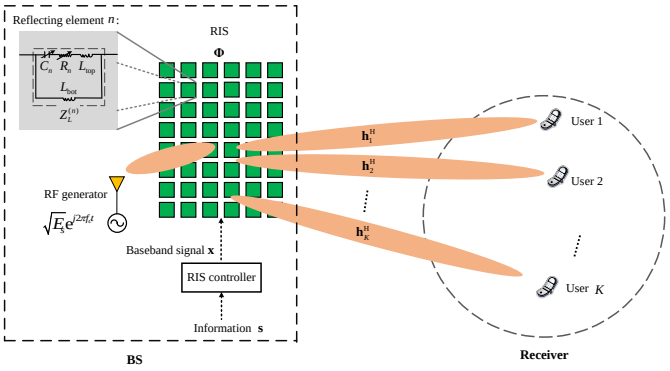

Fig. 1. Illustration of the RIS-aided single-RF downlink transmitter system model, including a base station (BS) with N-element RIS empowered by a single-RF generator, transmitting information to K single-antenna users.

A. Channel Model

In most treatises [19]–[27], the signals reflected by all RIS elements to the user are assumed to undergo independent fading, i.e. all N entries of the channel vector $\mathbf{h}_k^{\mathrm{H}}$ are considered independent. This assumption is only valid when the distance between the adjacent RIS elements is large enough [28]. However, in practical RIS scenarios, the entries in the channel vector $\mathbf{h}_k^{\mathrm{H}}$ are correlated due to the limited physical size of the RIS. As in [28], we employed the classic exponential correlation channel model [29] for representing the channel fading between the RIS and the users. Additionally, given the

[1]For M-ary signals and K users we have M^K possible multi-user signal combinations.

mobility of users, we assume that the signals between the RIS and users experience Rician fading [30], [31]. The channel vector $\mathbf{h}_k$ is given by

$$\mathbf{h}_k = \sqrt{\upsilon_k}\mathbf{w}_k, \quad k = 1, 2, \cdots, K, \tag{2}$$

where the path-loss from the RIS to the kth user is $\upsilon_k = \varrho_0 d_k^{-\alpha}$, with ϱ_0 representing the path-loss at the reference distance of 1 meter, α representing the path-loss exponent from the RIS to the users, and d_k representing the distance between the RIS and the kth user, and $\mathbf{w}_k$ is the small-scale fading of the links spanning from the RIS to the kth user, given by

$$\mathbf{w}_k^{\mathrm{H}} \sim \mathcal{N}_{\mathbb{C}}\Big(\sqrt{\frac{\kappa}{1+\kappa}}\overline{\mathbf{w}}_k^{\mathrm{H}}, \frac{1}{1+\kappa}\mathbf{R}\Big), \tag{3}$$

with $\mathbf{R}$ being the covariance matrix of the non-line-of-sight (NLoS) component of the channel vector $\mathbf{w}_k^{\mathrm{H}}$, and κ denoting the Rician factor. Furthermore, $\overline{\mathbf{w}}_k^{\mathrm{H}}$ represents the line-of-sight (LoS) component given by

$$\begin{aligned}\overline{\mathbf{w}}_k^{\mathrm{H}} = \big[1, \cdots, \mathrm{e}^{-j2\pi f_c d_0 (n_x \sin\psi^{(k)} \cos\varphi^{(k)} + n_y \cos\psi^{(k)})}, \cdots, \\ \mathrm{e}^{-j2\pi f_c d_0 ((N_x-1)\sin\psi^{(k)} \cos\varphi^{(k)} + (N_y-1)\cos\psi^{(k)})}\big]^{\mathrm{H}},\end{aligned} \tag{4}$$

where N_x and N_y represent the number of RIS elements at the vertical and horizontal direction, respectively, $0 \le n_x \le N_x - 1$, $0 \le n_y \le N_y - 1$. d_0 is the distance between adjacent RIS elements, and $\psi^{(k)}$ and $\varphi^{(k)}$ are the elevation and azimuth angle of departure (AoD) of the signals from the RIS to the kth user, respectively. According to the exponential correlation channel model [29], $\mathbf{R}$ is determined by the distance between the adjacent RIS elements. The (n_1, n_2)th entry in $\mathbf{R}$ is $r_{n_1,n_2} = \exp(-\frac{\delta_{n_1,n_2}}{\delta_0})$, where δ_{n_1,n_2} is the distance between the n_1th and n_2th RIS element, while δ_0 is a constant that controls the level of correlation. For a user equipment, δ_0 would be around half a wavelength, whereas for a base station it could be as high as tens of wavelengths [32]. We represent the path-loss from the RIS to all K users as $\boldsymbol{\Upsilon} = \mathrm{diag}\{\upsilon_1, \upsilon_2, \cdots, \upsilon_K\}$, and the small-scale fading from the RIS to all K users as $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K]^{\mathrm{H}}$. Therefore, the equivalent channel matrix between the RIS and all K users is given by

$$\mathbf{H} = \sqrt{\boldsymbol{\Upsilon}}\mathbf{W}. \tag{5}$$

We assume the instantaneous CSI, i.e. $\mathbf{W}$, can be attained at the BS. In practice this is acquired by channel estimation, as detailed in [33].

B. Configuration of RIS Elements

As shown in Fig. 1, the RIS elements are employed for modulating the baseband signal $\mathbf{x}$, where the reflection coefficient Φ_n of the nth RIS element is configured based on the complex value x_n. In the following, the details of configuring the reflection coefficients, including the amplitude and phase shift, are presented.

In each RIS reflecting element, the reflection coefficient Φ_n is controlled by the configurable load impedance $Z_L^{(n)}$, given by [34]

$$\Phi_n = \frac{Z_L^{(n)} - Z_0}{Z_L^{(n)} + Z_0}, \tag{6}$$

where Z_0 is the free space impedance, usually set as $Z_0 = 50\Omega$ [35]. The corresponding amplitude β_n and phase shift θ_n of the reflection coefficient Φ_n can be written as

$$\beta_n = \Big|\frac{Z_L^{(n)} - Z_0}{Z_L^{(n)} + Z_0}\Big|, \tag{7}$$

$$\theta_n = \arctan\Big(\frac{Z_L^{(n)} - Z_0}{Z_L^{(n)} + Z_0}\Big). \tag{8}$$

Based on [36], the circuit of configurable load impedance $Z_L^{(n)}$ is shown in Fig. 1, including a bottom layer inductance L_{bot}, a top layer inductance L_{top}, a variable resistance R_n, and a variable capacitance C_n. The equivalent load impedance $Z_L^{(n)}$ is given by

$$Z_L^{(n)} = \frac{j\omega L_{\text{bot}}(j\omega L_{\text{top}} + \frac{1}{j\omega C_n} + R_n)}{j\omega L_{\text{bot}} + j\omega L_{\text{top}} + \frac{1}{j\omega C_n} + R_n}. \tag{9}$$

The value of the load impedance $Z_L^{(n)}$ can be configured by controlling the variable resistance R_n and the variable capacitance C_n. To independently control the amplitude β_n in (7) and the phase shift in (8) by configuring the load impedance $Z_L^{(n)}$, we can employ the following two methods.

First, according to [25], [26], a non-linear modulation technique maybe employed for configuring the reflection coefficients. Specifically, in each time slot (TS) of duration T_s, the baseband signal $x_n(t)$ on the nth RIS element is defined under the constant envelope constraint as $x_n(t) = \mathrm{e}^{j\frac{\Delta\varphi}{T_s}(t+T_s-t_0)}$ for $0 \le t \le t_0$, and $x_n(t) = \mathrm{e}^{j\frac{\Delta\varphi}{T_s}(t-t_0)}$ for $t_0 < t \le T_s$, where t_0 is the circular time shift and $\frac{\Delta\varphi}{T_s}$ characterizes the changing rate of the phase that varies linearly with time. The reflection coefficient Φ_n is determined by the first harmonic of $x_n(t)$, as detailed in [26]. More explicitly, the amplitude of the first harmonic is determined by adjusting $\Delta\varphi$ and its phase shift by adjusting t_0. Since the amplitude of the harmonic of $x_n(t)$ is given by $|\mathrm{sinc}(\frac{\Delta\varphi}{2} - \pi)|$, the amplitude of the reflection coefficient would satisfy $0 \le \beta_n \le 1$.

The second method is based on [37], [38], where the amplitude β_n and phase shift θ_n were shown to be controlled independently by configuring the resistance R_n and capacitance C_n of the configurable load impedance circuit in the RIS element. Therefore, we opt for employing the following method for independently controlling the amplitude β_n and phase shift θ_n by configuring the load impedance $Z_L^{(n)}$. We denote the resistance component and the reactance component of the load impedance $Z_L^{(n)}$ as $R_L^{(n)}$ and $X_L^{(n)}$, respectively, i.e. $Z_L^{(n)} = R_L^{(n)} + jX_L^{(n)}$. According to (6), the reflection coefficient Φ_n is given by

$$\Phi_n = \frac{R_L^{(n)} - Z_0 + jX_L^{(n)}}{R_L^{(n)} + Z_0 + jX_L^{(n)}}. \tag{10}$$

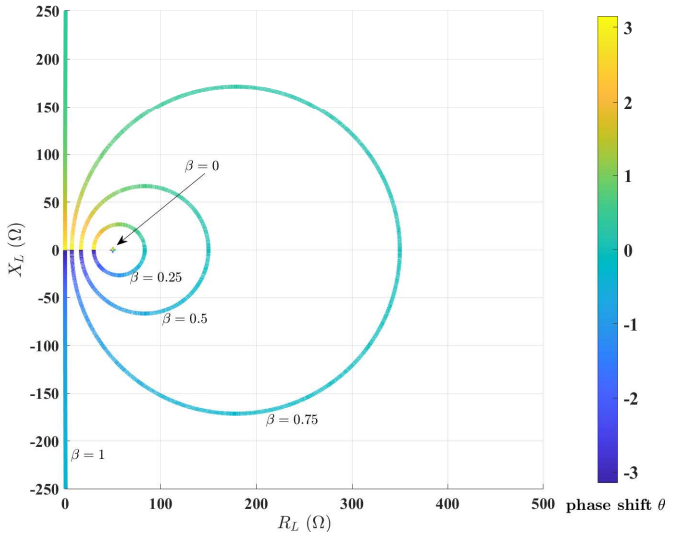

Fig. 2. The relationship between the reflection coefficient Φ_n and the configurable load impedance $Z_L^{(n)}$, where $\Phi_n = \beta_n e^{j2\pi\theta_n}$ and $Z_L^{(n)} = R_L^{(n)} + jX_L^{(n)}$.

Based on (7) and (10), we can get

$$\left(R_L^{(n)} - \frac{1+\beta_n^2}{1-\beta_n^2}Z_0\right)^2 + \left(X_L^{(n)}\right)^2 = \left(\frac{2\beta_n}{1-\beta_n^2}Z_0\right)^2, \quad (11)$$

which means that for a given amplitude β_n, the load impedance $Z_L^{(n)}$ should be configured on a circle having the center of $\left(\frac{1+\beta_n^2}{1-\beta_n^2}Z_0, 0\right)$ and radius of $\frac{2\beta_n}{1-\beta_n^2}Z_0$ for getting different phase shifts θ_n ranging from $-\pi$ to π. Note that the load impedance $Z_L^{(n)}$ is located on the point $(Z_0, 0)$ for $\beta_n = 0$, and distributed on the imaginary axis $R_L^{(n)} = 0$ for $\beta_n = 1$. Fig. 2 shows the relationship between the reflection coefficient Φ_n and the configurable load impedance $Z_L^{(n)}$, where we have $\Phi_n = \beta_n e^{j2\pi\theta_n}$ and $Z_L^{(n)} = R_L^{(n)} + jX_L^{(n)}$. As shown in Fig. 2, the amplitude and the phase shift of the reflection coefficient Φ_n can be controlled independently by configuring the resistance component $R_L^{(n)}$ and the reactance component $X_L^{(n)}$ of the load impedance $Z_L^{(n)}$, as reported in [38]. Additionally, given that the resistance $R_L^{(n)}$ is a non-negative value, the amplitude of β_n would satisfy $0 \le \beta_n \le 1$.

In summary, in this section we discussed two methods of independently controlling the amplitude and the phase shift of the reflection coefficient with the values in the following ranges:

$$0 \le \beta_n \le 1, \quad (12)$$

$$-\pi \le \theta_n < \pi. \quad (13)$$

The reflection coefficients of all RIS elements can be described as a vector $\mathbf{\Phi} \in \mathbb{C}^{N\times 1}$, given by

$$\mathbf{\Phi} = [\Phi_1, \cdots, \Phi_N]^{\mathrm{T}} = [\beta_1 e^{j\theta_1}, \cdots, \beta_N e^{j\theta_N}]^{\mathrm{T}}. \quad (14)$$

In the following sections, we will show how the reflection coefficients are configured for realizing multi-user information transmission.

C. Transmit Precoding Scheme

In this section, we present the design of the baseband signal $\mathbf{x}$ on the RIS to realize the transmission of the multi-user information $\mathbf{s}$. In the conventional MIMO system, the baseband signal can be represented by $\mathbf{x} = \mathbf{P\Lambda s}$, with the TPC $\mathbf{P}$ and power allocation $\mathbf{\Lambda}$ applied to the multi-user information $\mathbf{s}$, while employing multiple RF-chains. In our RIS-aided single-RF system model, the aim is to design the baseband signal on the RIS to achieve a similar effect to that of $\mathbf{x} = \mathbf{P\Lambda s}$, while employing a single-RF aided design, rather than the conventional multiple-RF MIMO design.

In the following, we employ the ZF TPC method as a design example, but other TPC schemes may also be readily harnessed. The ZF TPC matrix is given by [39]

$$\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_K] = \sqrt{N}\mathbf{W}^{\mathrm{H}}(\mathbf{W}\mathbf{W}^{\mathrm{H}})^{-1}, \quad (15)$$

where $\mathbf{p}_k$ is the TPC vector configured for the kth user, and the constant parameter $\sqrt{N}$ is employed for ensuring $\mathbb{E}[\|\mathbf{p}_k\|^2] = 1$. The baseband signal $\mathbf{x}$ can be formulated as

$$\mathbf{x} = \sum_{k=1}^{K} \sqrt{\lambda_k}\mathbf{p}_k s_k, \quad (16)$$

where s_k is the M-ary PSK information symbol destined for the kth user[2], i.e., $s_k \in \mathbb{S}_{M-\mathrm{PSK}}$, where $\mathbb{S}_{M-\mathrm{PSK}}$ is the set of M-PSK modulated symbols, and λ_k is the power allocated to the kth user.

In our system model, the single RF generator of Fig. 1 distributes the power of E_s evenly among the N RIS elements, followed by the phase and amplitude configuration. Then the signal is reflected to the users. Thus, the power impinging on each RIS element is $\frac{E_s}{N}$, while the power reflected from the nth RIS element is $\beta_n^2\frac{E_s}{N}$. The total power reflected from all N RIS elements, denoted as E_r, is given by

$$E_r = \sum_{n=1}^{N}\left(\beta_n^2\frac{E_s}{N}\right) = \frac{E_s}{N}\sum_{n=1}^{N}\beta_n^2 = E_s\xi, \quad (17)$$

where $\xi = \frac{1}{N}\sum_{n=1}^{N}\beta_n^2$ is the average power reflectance of all RIS elements. Since $0 \le \beta_n \le 1$, the power reflectance should satisfy $0 \le \xi \le 1$. Therefore, we get

$$\lambda_1 + \lambda_2 + \cdots + \lambda_K = E_r = E_s\xi. \quad (18)$$

We now introduce $\lambda_k = E_s\xi\lambda'_k$, where λ'_k is the power sharing ratio of the kth user, satisfying

$$\lambda'_1 + \lambda'_2 + \cdots + \lambda'_K = 1. \quad (19)$$

Then, the baseband signal $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \sum_{k=1}^{K}\sqrt{E_s\xi\lambda'_k}\mathbf{p}_k s_k. \quad (20)$$

The operation in (20), including the design of TPC matrix and the power allocation algorithm, is completed by the RIS controller, while further details of the power allocation are included in Section III.

[2]Other modulation methods, such as QAM, are also readily applicable to our proposed system model.

Once the baseband signal $\mathbf{x}$ is evaluated in the RIS controller using (20), given that the power impinging on each RIS element is $\frac{E_s}{N}$, the RIS elements are appropriately configured to modulate the baseband signal $\mathbf{x}$ by adjusting the reflection coefficients as

$$\sqrt{\frac{E_s}{N}}\mathbf{\Phi} = \mathbf{x}. \tag{21}$$

According to (12), since the RIS reflecting elements are passive, the power reflectance ξ and the power sharing ratio $\lambda'_1, \cdots, \lambda'_K$ should satisfy the condition that the maximum amplitude of all elements in $\mathbf{\Phi}$ is not higher than unity for all possible information symbol vectors $\mathbf{s}$, i.e.,

$$\max_{s_k \in \mathbb{S}_{M-\text{PSK}}} \Big\| \sum_{k=1}^{K} \sqrt{E_s \xi \lambda'_k}\mathbf{p}_k s_k \Big\|_\infty \leq \sqrt{\frac{E_s}{N}}, \tag{22}$$

which can be simplified as

$$\max_{s_k \in \mathbb{S}_{M-\text{PSK}}} \Big\| \sum_{k=1}^{K} \sqrt{N \xi \lambda'_k}\mathbf{p}_k s_k \Big\|_\infty \leq 1. \tag{23}$$

It can be shown that the power reflectance ξ is dependent on the TPC matrix $\mathbf{P}$, which is determined by the channel vector between the RIS and the users as well as by the TPC method used, plus by the power sharing ratio $\lambda'_1, \cdots, \lambda'_K$ of each user. These ratios are determined by the power allocation method, but they are independent of the specific transmitted information. However, (23) has a high computational complexity, when determining the amplitude of the transmitted signal in $\mathbf{x}$ for all possible M^K information symbol vector $\mathbf{s}$, especially when the number of users K and the modulation order M increases. Since the PSK-modulated signal s_k has a constant envelope, it can be observed that $|\mathbf{p}_k s_k| \leq |\mathbf{p}_k|$. According to the Cauchy-Schwarz inequality, we can get

$$\max_{s_k \in \mathbb{S}_{M-\text{PSK}}} \Big\| \sum_{k=1}^{K} \sqrt{N \xi \lambda'_k}\mathbf{p}_k s_k \Big\|_\infty \leq \Big\| \sum_{k=1}^{K} \sqrt{N \xi \lambda'_k}|\mathbf{p}_k| \Big\|_\infty, \tag{24}$$

where the equality is established when the modulation order obeys $M \to \infty$. Therefore, the constraint in (23) can be simplified as

$$\Big\| \sum_{k=1}^{K} \sqrt{N \xi \lambda'_k}|\mathbf{p}_k| \Big\|_\infty \leq 1. \tag{25}$$

In the following section, we present the details of designing the power sharing, i.e. determining the power sharing ratios $\lambda'_1, \cdots, \lambda'_K$ and the corresponding power reflectance ξ under the constraints in (19) and (25).

III. Power Allocation Methods

According to (1), (5) and (20), the kth user's received signal is

$$\begin{aligned} y_k &= \sqrt{\upsilon_k}\mathbf{w}_k^{\text{H}}\mathbf{x} + q_k \\ &= \sqrt{\upsilon_k E_s \xi}\mathbf{w}_k^{\text{H}} \sum_{l=1}^{K} \sqrt{\lambda'_l}\mathbf{p}_l s_l + q_k \\ &\overset{\text{(a)}}{=} \sqrt{\upsilon_k E_s \xi \lambda'_k}\mathbf{w}_k^{\text{H}}\mathbf{p}_k s_k + q_k \\ &\overset{\text{(b)}}{=} \sqrt{N \upsilon_k E_s \xi \lambda'_k} s_k + q_k, \end{aligned} \tag{26}$$

where (a) and (b) are true since $\mathbf{w}_k^{\text{H}}\mathbf{p}_l = 0$ when $k \neq l$ and $\mathbf{w}_k^{\text{H}}\mathbf{p}_l = \sqrt{N}$ when $k = l$, respectively, when employing the ZF method. At the receiver, the information destined for the kth user can be detected directly according to the received signal y_k. The signal-to-interference-plus-noise-ratio (SINR) of received information at the kth user is given by

$$\rho_k = \frac{\upsilon_k E_s \xi \lambda'_k \|\mathbf{w}_k^{\text{H}}\mathbf{p}_k\|^2}{\sigma_q^2 + \upsilon_k E_s \xi \sum_{l \neq k} \lambda'_l \|\mathbf{w}_k^{\text{H}}\mathbf{p}_l\|^2} = \frac{N \upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}. \tag{27}$$

Based on (27), the achievable rate of the kth user is expressed as

$$R_k = \log_2(1 + \rho_k) = \log_2\Big(1 + \frac{N \upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\Big). \tag{28}$$

Our aim here is to determine the power sharing ratio $\lambda'_1, \cdots, \lambda'_K$ and the corresponding power reflectance ξ under the constraints in (19) and (25). We then opt for the alternating optimization (AO) technique for iteratively calculating the power sharing ratio $\lambda'_1, \cdots, \lambda'_K$ and the corresponding power reflectance ξ. Specifically, when the power reflectance ξ is fixed, we consider three power allocation methods, including maximizing the sum-rate, maximizing the minimum rate (min-rate) and maximizing the geometric-mean-rate, in order to derive the power allocation ratio $\lambda'_1, \cdots, \lambda'_K$, under the constraint of (19). Then, when the power sharing ratio $\lambda'_1, \cdots, \lambda'_K$ is fixed, we employ the bisection method for calculating the corresponding power reflectance ξ, under the constraint of (25).

1) Maximizing the sum-rate: According to (28), the sum-rate of the K users is given by

$$R_{\text{sum}} = \sum_{k=1}^{K} R_k = \sum_{k=1}^{K} \log_2\Big(1 + \frac{N \upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\Big). \tag{29}$$

The optimization problem of maximizing R_{sum} can then be formulated as

$$\begin{aligned} \text{(P1.a)} \quad & \max_{\xi, \lambda'_1, \cdots, \lambda'_K} \sum_{k=1}^{K} \log_2\Big(1 + \frac{N \upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\Big). \\ \text{s.t.} \quad & \lambda'_1 + \lambda'_2 + \cdots + \lambda'_K = 1, \\ & \Big\| \sum_{k=1}^{K} \sqrt{N \xi \lambda'_k}|\mathbf{p}_k| \Big\|_\infty \leq 1. \end{aligned} \tag{30}$$

Firstly, when we fix the power reflectance ξ and ignore the second constraint in (P1.a), the optimization problem (P1.a) may be written as

$$\begin{aligned} \text{(P1.b)} \quad & \max_{\lambda'_1, \cdots, \lambda'_K} \sum_{k=1}^{K} \log_2\Big(1 + \frac{N \upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\Big). \\ \text{s.t.} \quad & \lambda'_1 + \lambda'_2 + \cdots + \lambda'_K = 1. \end{aligned} \tag{31}$$

The optimization problem (P1.b) may then be solved by the classical water-filling method detailed in [29].

Secondly, when the power sharing ratio $\lambda'_1, \cdots, \lambda'_K$ is

Algorithm 1 Alternating optimization method conceived for maximizing sum-rate/min-rate/geometric-mean-rate power allocation schemes.

Input: $[\mathbf{p}_1, \cdots, \mathbf{p}_K] = \sqrt{N}\mathbf{W}^{\mathrm{H}}(\mathbf{W}\mathbf{W}^{\mathrm{H}})^{-1}$, initial $\xi_{\min}$, initial $\xi_{\max}$, $\frac{E_s}{\sigma_q^2}$, $\upsilon_1, \cdots, \upsilon_K$, ε_ξ, and iteration times T.

For $t = 1$ **to** T
  $\xi^{(t)} = \frac{\xi_{\min}+\xi_{\max}}{2}$.
  $[\lambda_1'^{(t)}, \cdots, \lambda_K'^{(t)}] = f_{\text{sum}}(\frac{E_s}{\sigma_q^2}, \xi^{(t)}, \upsilon_1, \upsilon_2, \cdots, \upsilon_K)$,
  or $[\lambda_1'^{(t)}, \cdots, \lambda_K'^{(t)}] = f_{\min}(\upsilon_1, \cdots, \upsilon_K)$,
  or $[\lambda_1'^{(t)}, \cdots, \lambda_K'^{(t)}] = f_{\text{GM}}(\frac{E_s}{\sigma_q^2}, \xi^{(t)}, \upsilon_1, \upsilon_2, \cdots, \upsilon_K)$.
  Repeat
    if $\left\| \sum_{k=1}^{K} \sqrt{N\xi^{(t)}\lambda_k'^{(t)}}|\mathbf{p}_k| \right\|_\infty < 1 - \varepsilon_\xi$,
      $\xi_{\min} = \xi^{(t)}$;
    elseif $\left\| \sum_{k=1}^{K} \sqrt{N\xi^{(t)}\lambda_k'^{(t)}}|\mathbf{p}_k| \right\|_\infty > 1$,
      $\xi_{\max} = \xi^{(t)}$.
    end
    $\xi^{(t)} = \frac{\xi_{\min}+\xi_{\max}}{2}$.
  Until $1 - \varepsilon_\xi \leq \left\| \sum_{k=1}^{K} \sqrt{N\xi^{(t)}\lambda_k'^{(t)}}|\mathbf{p}_k| \right\|_\infty \leq 1$.
end
Return Power allocation ratio $\lambda_1'^{(t)}, \cdots, \lambda_K'^{(t)}$, and power reflectance $\xi^{(t)}$.

obtained, we can employ the bisection method in order to find the power reflectance ξ that satisfies the second constraint in (P1.a).

The detailed process of the alternating optimization method conceived for maximizing the sum-rate is presented in Algorithm 1, where $\xi_{\min}$ is the lower bound of ξ and the initial $\xi_{\min}$ can be set to 0, $\xi_{\max}$ is the upper bound of ξ and the initial $\xi_{\max}$ can be set to 1. Furthermore, ε_ξ is the maximum tolerable error of ξ and $[\lambda_1', \cdots, \lambda_K'] = f_{\text{sum}}(\frac{E_s}{\sigma_q^2}, \xi, \upsilon_1, \cdots, \upsilon_K)$ is the function of calculating the power sharing ratio of each user by employing the classic water-filling method, given $\frac{E_s}{\sigma_q^2}$, ξ, and $\upsilon_1, \cdots, \upsilon_K$.

2) Maximizing the min-rate: Although maximizing the sum-rate achieves the highest possible throughput of the whole system, it is typically unfair for the users who have poor channel condition. This is because most of the power is allocated to the users having good channel conditions. This often leaves near-zero rate for users having low SNR. By contrast, maximizing the min-rate based power sharing method maximizes the min-rate of all users. According to (28), the minimum rate of all K users, denoted as $R_{\min}$, is given by

$$R_{\min} = \min_{k=1,\cdots,K} R_k = \min_{k=1,\cdots,K} \log_2\left(1 + \frac{N\upsilon_k E_s \xi \lambda_k'}{\sigma_q^2}\right). \tag{32}$$

The problem of maximizing $R_{\min}$ is formulated as

$$\begin{aligned} \text{(P2.a)} \quad & \max_{\xi,\lambda_1',\cdots,\lambda_K'} \min_{k=1,\cdots,K} \log_2\left(1 + \frac{N\upsilon_k E_s \xi \lambda_k'}{\sigma_q^2}\right) \\ \text{s.t.} \quad & \lambda_1' + \lambda_2' + \cdots + \lambda_K' = 1, \\ & \left\| \sum_{k=1}^{K} \sqrt{N\xi\lambda_k'}|\mathbf{p}_k| \right\|_\infty \leq 1. \end{aligned} \tag{33}$$

Firstly, we fix the power reflectance ξ and ignore the second constraint in (P2.a). Then the optimization problem (P2.a) can be written as

$$\begin{aligned} \text{(P2.b)} \quad & \max_{\lambda_1',\cdots,\lambda_K'} \min_{k=1,\cdots,K} \log_2\left(1 + \frac{N\upsilon_k E_s \xi \lambda_k'}{\sigma_q^2}\right) \\ \text{s.t.} \quad & \lambda_1' + \lambda_2' + \cdots + \lambda_K' = 1. \end{aligned} \tag{34}$$

The optimization problem (P2.b) is equivalent to ensuring that the spectral efficiency of all K users is the same. Therefore, we have:

$$\upsilon_1\lambda_1' = \upsilon_2\lambda_2' = \cdots = \upsilon_K\lambda_K'. \tag{35}$$

Since $\lambda_1' + \lambda_2' + \cdots + \lambda_K' = 1$, the power sharing ratio of each user is given by

$$\lambda_k' = \frac{\frac{1}{\upsilon_k}}{\frac{1}{\upsilon_1} + \frac{1}{\upsilon_2} + \cdots + \frac{1}{\upsilon_K}}, \quad k = 1, 2, \cdots, K. \tag{36}$$

According to (36), we can find that λ_k' is entirely determined by $\upsilon_1, \cdots, \upsilon_K$.

Secondly, when the power sharing ratio $\lambda_1', \cdots, \lambda_K'$ is fixed, similar to maximizing the sum-rate, we can employ the bisection method for finding the maximum of the power reflectance ξ satisfying the second constraint in (P2.a). The detailed process of the alternating optimization method conceived for the maximizing the min-rate is shown in Algorithm 1, where $[\lambda_1', \cdots, \lambda_K'] = f_{\min}(\upsilon_1, \cdots, \upsilon_K)$ is the function of calculating the power sharing ratio of each user by maximizing the min-rate according to (36), given $\upsilon_1, \cdots, \upsilon_K$.

3) Maximizing the geometric-mean-rate: It is meaningful to explore maximizing the geometric-mean-rate as well, since it shows a substantially improved rate-fairness amongst the users [12]. According to (28), the geometric-mean of the achievable rate of all K users, denoted as R_{GM}, is given by [12]

$$R_{\text{GM}} = \left(\prod_{k=1}^{K} R_k\right)^{\frac{1}{K}} = \left(\prod_{k=1}^{K} \log_2\left(1 + \frac{N\upsilon_k E_s \xi \lambda_k'}{\sigma_q^2}\right)\right)^{\frac{1}{K}}. \tag{37}$$

The problem of maximizing R_{GM} is formulated as

$$\begin{aligned} \text{(P3.a)} \quad & \max_{\xi,\lambda_1',\cdots,\lambda_K'} \prod_{k=1}^{K} \log_2\left(1 + \frac{N\upsilon_k E_s \xi \lambda_k'}{\sigma_q^2}\right). \\ \text{s.t.} \quad & \lambda_1' + \lambda_2' + \cdots + \lambda_K' = 1, \\ & \left\| \sum_{k=1}^{K} \sqrt{N\xi\lambda_k'}|\mathbf{p}_k| \right\|_\infty \leq 1. \end{aligned} \tag{38}$$

Firstly, when we fix the power reflectance ξ and ignore the second constraint in (P3.a), the optimization problem (P3.a) can be written as

$$\begin{aligned} \text{(P3.b)} \quad & \max_{\lambda_1',\cdots,\lambda_K'} \prod_{k=1}^{K} \log_2\left(1 + \frac{N\upsilon_k E_s \xi \lambda_k'}{\sigma_q^2}\right). \\ \text{s.t.} \quad & \lambda_1' + \lambda_2' + \cdots + \lambda_K' = 1. \end{aligned} \tag{39}$$

Algorithm 2 Double-loop bisection method conceived for calculating the power allocation ratio in maximizing geometric-mean-rate method.

Input: $\eta_1 = \frac{E_s N \upsilon_1 \xi}{\sigma_q^2}, \cdots, \eta_K = \frac{E_s N \upsilon_K \xi}{\sigma_q^2}$, ε_c, and ε_λ.
1: $c_{\min} = \min\limits_{k=1,2,\cdots,K} \frac{\eta_k}{(1+\frac{\eta_k}{K})\ln(1+\frac{\eta_k}{K})}$.
2: $c_{\max} = \max\limits_{k=1,2,\cdots,K} \frac{\eta_k}{(1+\frac{\eta_k}{K})\ln(1+\frac{\eta_k}{K})}$.
3: **Repeat**
4: $c = \frac{c_{\min}+c_{\max}}{2}$.
5: **for** $k = 1$ **to** K
6: $\lambda'_{k\min} = \frac{\sqrt{1+\frac{4\eta_k}{c}}-1}{2\eta_k}$.
7: $\lambda'_{k\max} = \max\left\{\frac{\frac{\eta_k}{c}-1}{\eta_k}, \frac{\mathrm{e}-1}{\eta_k}\right\}$.
8: **Repeat**
9: $\lambda'_k = \frac{\lambda'_{k\min}+\lambda'_{k\max}}{2}$.
10: **if** $\frac{\eta_k}{(1+\eta_k\lambda'_k)\ln(1+\eta_k\lambda'_k)} \geq 1+\varepsilon_\lambda$,
11: $\lambda'_{k\min} = \lambda'_k$;
12: **elseif** $\frac{\eta_k}{(1+\eta_k\lambda'_k)\ln(1+\eta_k\lambda'_k)} \leq 1-\varepsilon_\lambda$,
13: $\lambda'_{k\max} = \lambda'_k$.
14: **end**
15: **Until** $1-\varepsilon_\lambda < \frac{\eta_k}{(1+\eta_k\lambda'_k)\ln(1+\eta_k\lambda'_k)} < 1+\varepsilon_\lambda$.
16: **end**
17: **if** $\lambda'_1+\lambda'_2+\cdots+\lambda'_K \geq 1$,
18: $c_{\min} = c$;
19: **elseif** $\lambda'_1+\lambda'_2+\cdots+\lambda'_K \leq 1-\varepsilon_c$,
20: $c_{\max} = c$.
21: **end**
22: **Until** $1-\varepsilon_c < \lambda'_1+\lambda'_2+\cdots+\lambda'_K < 1$.
23: **Return** Power allocation ratio $\lambda'_1, \cdots, \lambda'_K$.

The problem (P3.b) is equivalent to

$$\text{(P3.c)} \quad \max_{\lambda'_1,\cdots,\lambda'_K} \sum_{k=1}^{K} \ln\ln\left(1+\frac{N\upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\right).$$
$$\text{s.t.} \quad \lambda'_1+\lambda'_2+\cdots+\lambda'_K = 1. \tag{40}$$

The problem (P3.c) may then be solved by using the classic Lagrange multiplier method from the calculus of variations [29]. Hence, we can formulate our problem as

$$J(\lambda'_1,\cdots,\lambda'_K) = \sum_{k=1}^{K} \ln\ln\left(1+\frac{N\upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\right) + c\sum_{k=1}^{K}\lambda'_k, \tag{41}$$

where c is a constant called the *Lagrange multiplier*. Then, we take the partial derivative of J with respect to the power sharing variables to be optimized $\lambda'_1,\cdots,\lambda'_K$ and set them equal to zero, which results in

$$\frac{\partial J}{\partial \lambda'_k} = \frac{\frac{N\upsilon_k E_s \xi}{\sigma_q^2}}{\left(1+\frac{N\upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\right)\ln\left(1+\frac{N\upsilon_k E_s \xi \lambda'_k}{\sigma_q^2}\right)} + c = 0. \tag{42}$$

Upon introducing $\eta_k = \frac{N\upsilon_k E_s \xi}{\sigma_q^2}$, (42) can be written as

$$\frac{\partial J}{\partial \lambda'_k} = \frac{\eta_k}{(1+\eta_k\lambda'_k)\ln(1+\eta_k\lambda'_k)} + c = 0. \tag{43}$$

It is intractable to derive the closed-form solution of (43), hence we resort to numerical methods to find λ'_k as follows.

Since it may be readily observed that $\frac{\eta_k}{(1+\eta_k\lambda'_k)\ln(1+\eta_k\lambda'_k)}$ is a monotonically decreasing function in the interval $\lambda'_k \in [0,1]$, the equation in (43) has a unique solution. We can employ the double-loop based bisection method to determine $\lambda'_1,\cdots,\lambda'_K$ under the constraint of $\lambda'_1+\lambda'_2+\cdots+\lambda'_K = 1$ and (43). The detailed process of the double-loop bisection method conceived for calculating the power sharing ratio in maximizing the geometric-mean-rate is shown in Algorithm 2, where ε_c is the maximum tolerable error of the outside bisection loop (from line 3 to line 22), and ε_λ is that of the inside bisection loop (from line 8 to line 15). In the inner loop, we fix the constant value c and find $\lambda'_1,\cdots,\lambda'_K$ by the bisection method. According to the value of $\lambda'_1,\cdots,\lambda'_K$ obtained from the inner bisection loop, we find the value of c in the outer bisection loop. We denote the solution of maximizing the geometric-mean-rate in Algorithm 2 as $[\lambda'_1,\cdots,\lambda'_K] = f_{\text{GM}}(\frac{E_s}{\sigma_q^2}, \xi, \upsilon_1,\cdots,\upsilon_K)$.

Afterwards, we can employ the classic alternating optimization method to find the optimal power sharing ratio $\lambda'_1,\cdots,\lambda'_K$ and the corresponding power reflectance ξ for maximizing the geometric-mean-rate. The detailed process of the alternating optimization method conceived for maximizing the geometric-mean-rate is shown in Algorithm 1, where $[\lambda'_1,\cdots,\lambda'_K] = f_{\text{GM}}(\frac{E_s}{\sigma_q^2}, \xi, \upsilon_1,\cdots,\upsilon_K)$ is the function of calculating the power sharing ratio of each user by employing the double-loop bisection method in Algorithm 2, given $\frac{E_s}{\sigma_q^2}$, ξ, and $\upsilon_1,\cdots,\upsilon_K$.

Proposition 1. The objective value of (P1.a), (P2.a) and (P3.a) converges over multiple iterations by applying Algorithm 1.

Proof: Here we present the proof for the convergence of the objective value of (P1.a) as follows, and that of (P2.a) and (P3.a) can be attained similarly. We denote the objective value of (P1.b) based on a feasible solution $(\xi, \lambda'_1,\cdots,\lambda'_K)$ as $g(\xi,\lambda'_1,\cdots,\lambda'_K)$. As shown in Algorithm 1, if there exists a feasible solution to the optimization between line 6 and line 13, i.e. $(\xi^{(t+1)}, \lambda_1'^{(t)},\cdots,\lambda_K'^{(t)})$ exists, it is also feasible for the optimization of problem (P1.b). Thus, $(\xi^{(t)}, \lambda_1'^{(t)},\cdots,\lambda_K'^{(t)})$ and $(\xi^{(t+1)}, \lambda_1'^{(t+1)},\cdots,\lambda_K'^{(t+1)})$ in line 3 are feasible solutions of the problem (P1.b) in the tth and $(t+1)$th iterations, respectively. Then, we can show that $g(\xi^{(t+1)}, \lambda_1'^{(t+1)},\cdots,\lambda_K'^{(t+1)}) \overset{(a)}{\geq} g(\xi^{(t+1)}, \lambda_1'^{(t)},\cdots,\lambda_K'^{(t)}) \overset{(b)}{=} g(\xi^{(t)}, \lambda_1'^{(t)},\cdots,\lambda_K'^{(t)})$, where (a) is established due to the fact that $\lambda_1'^{(t+1)},\cdots,\lambda_K'^{(t+1)}$ is the optimal solution to the problem (P1.b) for a given $\xi^{(t+1)}$, and (b) is established since the objective function of (P1.b) is independent of ξ and only depends on $\lambda'_1,\cdots,\lambda'_K$. ■

IV. Performance Analysis

In this section, we theoretically derive the ergodic rate, symbol error probability, outage probability, and coverage range for our proposed RIS-aided single-RF downlink multi-user transmitter architecture, and compare it to the corresponding performance in the conventional MIMO systems.

The derivation of the closed-form performance analysis for our proposed architecture is intractable, because we do not have the closed-form expression of the power allocation. For ease of comparison with the conventional MIMO systems, we assume that the entries in the channel vector $\mathbf{h}_k$ are uncorrelated, i.e. $\mathbf{R} = \mathbf{I}_N$, which can be viewed as the upper bound of the correlated channel case as shown in the simulation results. We also assume that all the users have the same large-scale fading, i.e. $\upsilon_1 = \upsilon_2 = \cdots = \upsilon_K = \upsilon$, which reveals that the power is evenly allocated to all the users, i.e. $\lambda'_1 = \lambda'_2 = \cdots = \lambda'_K = \frac{1}{K}$, and all users have the same SINR of the received information, i.e. $\rho_1 = \rho_2 = \cdots = \rho_K = \rho$.

According to (27), the SINR of the received information at each user is given by

$$\rho = \frac{N\upsilon E_s \xi}{\sigma_q^2 K}, \tag{44}$$

and the constraint in (25) is given by

$$\Big\| \sum_{k=1}^{K} \sqrt{\xi} \sqrt{\frac{N}{K}} |\mathbf{p}_k| \Big\|_\infty \leq 1. \tag{45}$$

Therefore, the maximum value of the power reflectance ξ is

$$\xi = \frac{K}{N\|\mathbf{P}\|_\infty^2} = \frac{K}{N^2 \big\|\mathbf{W}^{\mathrm{H}}(\mathbf{W}\mathbf{W}^{\mathrm{H}})^{-1}\big\|_\infty^2}. \tag{46}$$

When the number N of RIS elements is large enough, the row vectors in $\mathbf{W}$ are approximately orthogonal due to the channel hardening, i.e. $\mathbf{W}\mathbf{W}^{\mathrm{H}} = N\mathbf{I}_K$ [39]. Hence, we have

$$\xi = \frac{K}{\big\|\mathbf{W}^{\mathrm{H}}\big\|_\infty^2} = \frac{1}{\big\|\frac{1}{\sqrt{K}}\mathbf{W}^{\mathrm{H}}\big\|_\infty^2}. \tag{47}$$

We denote the sum of the absolute values of the entries in the nth row of the matrix $\frac{1}{\sqrt{K}}\mathbf{W}^{\mathrm{H}}$ as U_n, i.e.

$$U_n = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} |w_{n,k}|, \tag{48}$$

where the RIS index is $n = 1, 2, \cdots, N$, the user index is $k = 1, 2, \cdots, K$, and $w_{n,k}$ represents the (n,k)th element in the matrix $\mathbf{W}$. Therefore, we have

$$\xi = \frac{1}{U_{\max}^2}, \tag{49}$$

where $U_{\max} = \max_{n=1,2,\cdots,N} U_n$. Since all the elements in $\mathbf{W}$ independently follow identical complex Gaussian distributions with unity variance, the amplitude of $w_{k,n}$ follows the Rician distribution with the shape parameter κ and the scale parameter of 1 [13]. We employ the method of moments for approximating the distribution of U_n as follows. The first and second moment of $|w_{k,n}|$ are given by

$$\mathbb{E}[|w_{n,k}|] = \sqrt{\frac{\pi}{4(1+\kappa)}} L_{\frac{1}{2}}(-\kappa), \tag{50}$$

and

$$\mathbb{E}[|w_{n,k}|^2] = 1, \tag{51}$$

where $L_{\frac{1}{2}}(\cdot)$ denotes a Laguerre polynomial. The mean of U_n is given by

$$\mathbb{E}[U_n] = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \mathbb{E}[|w_{n,k}|] = \sqrt{\frac{K\pi}{4(1+\kappa)}} L_{\frac{1}{2}}(-\kappa). \tag{52}$$

The second moment of U_n is formulated as:

$$\begin{aligned}
&\mathbb{E}[U_n^2] \\
=&(\frac{1}{\sqrt{K}})^2 \mathbb{E}\left[\sum_{k=1}^{K} |w_{n,k}|^2 + \sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^{K} |w_{n,k_1}||w_{n,k_2}| \right] \\
=&\frac{1}{K}\left[\sum_{k=1}^{K} \mathbb{E}[|w_{n,k}|^2] + \sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^{K} \mathbb{E}[|w_{n,k_1}|]\mathbb{E}[|w_{n,k_2}|] \right] \\
=&1 + \frac{(K-1)\pi}{4(1+\kappa)} L_{\frac{1}{2}}^2(-\kappa).
\end{aligned} \tag{53}$$

Therefore, based on the moment-matching method, U_n approximately obeys the Gamma distribution

$$U_n \sim \mathrm{Gamma}(\mu, \nu), \tag{54}$$

where μ is the shape parameter given by

$$\mu = \frac{(\mathbb{E}[U_n])^2}{\mathbb{E}[U_n^2] - (\mathbb{E}[U_n])^2} = \frac{\frac{K\pi}{(1+\kappa)} L_{\frac{1}{2}}^2(-\kappa)}{4 - \frac{\pi}{(1+\kappa)} L_{\frac{1}{2}}^2(-\kappa)}, \tag{55}$$

and ν is the scale parameter given by

$$\nu = \frac{\mathbb{E}[U_n^2] - (\mathbb{E}[U_n])^2}{\mathbb{E}[U_n]} = \frac{4 - \frac{\pi}{(1+\kappa)} L_{\frac{1}{2}}^2(-\kappa)}{2\sqrt{\frac{K\pi}{(1+\kappa)}} L_{\frac{1}{2}}(-\kappa)}. \tag{56}$$

The PDF and the CDF of U_n are given by

$$f_{U_n}(u) = \frac{u^{\mu-1} \exp(-\frac{u}{\nu})}{\Gamma(\mu)\nu^\mu}, \tag{57}$$

$$F_{U_n}(u) = \frac{\Gamma_{\mathrm{li}}(\frac{u}{\nu}; \mu)}{\Gamma(\mu)}, \tag{58}$$

where $\Gamma(\mu)$ is the gamma function, and $\Gamma_{\mathrm{li}}(\cdot;\mu)$ is the lower incomplete gamma function with parameter μ [40]. According to order statistics, the PDF and CDF of $U_{\max}$ is given by [41]

$$\begin{aligned}
f_{U_{\max}}(u) &= N(F_{U_n}(u))^{N-1} f_{U_n}(u) \\
&= N\left[\frac{\Gamma_{\mathrm{li}}(\frac{u}{\nu};\mu)}{\Gamma(\mu)}\right]^{N-1} \frac{u^{\mu-1}\exp(-\frac{u}{\nu})}{\Gamma(\mu)\nu^\mu},
\end{aligned} \tag{59}$$

and

$$F_{U_{\max}}(u) = [F_{U_n}(u)]^N = \left[\frac{\Gamma_{\mathrm{li}}(\frac{u}{\nu};\mu)}{\Gamma(\mu)}\right]^N, \tag{60}$$

respectively.

A. Ergodic Rate

The egodic rate of each user is given by

$$\begin{aligned}
R &= \mathbb{E}[\log_2(1+\rho)] \\
&= \mathbb{E}\Big[\log_2\Big(1 + \frac{N\upsilon E_s \xi}{\sigma_q^2 K}\Big)\Big]
\end{aligned}$$

$$\overset{(a)}{=} \int_0^{\infty} f_{U_{\max}}(u) \log_2 \Big(1 + \frac{N\upsilon E_s}{\sigma_q^2 K u^2}\Big) \mathrm{d}u, \tag{61}$$

where (a) is based on (49). Since $f(x) = \log_2(1 + \frac{1}{x})$ is a convex function, we can formulate the lower bound of the ergodic rate of each user as

$$R^{(\mathrm{LB})} = \log_2 \Big(1 + \frac{N\upsilon E_s}{\sigma_q^2 K \mathbb{E}[\frac{1}{\xi}]}\Big) = \log_2 \Big(1 + \frac{N\upsilon E_s}{\sigma_q^2 K \mathbb{E}[U_{\max}^2]}\Big), \tag{62}$$

where $\mathbb{E}[U_{\max}^2] = \int_0^{\infty} u^2 f_{U_{\max}}(u) \mathrm{d}u$ is the second moment of $U_{\max}$, which is a fixed value for a given number of users K, Rician factor κ and number of RIS elements N, and it can be calculated by a numerical method such as the Gauss quadrature [42].

Observe in (62) that when the number of RIS elements obeys $N \to \infty$, the lower bound of the ergodic rate of each user increases exponentially with the transmit power E_s.

On the other hand, for the sake of comparison, in the conventional fully-digital MIMO, we assume that the number of RF-chains is N_{RF}. Then the rate upper bound is given by

$$R_{\mathrm{MIMO}}^{(\mathrm{UB})} = \log_2 \Big(1 + \frac{N_{\mathrm{RF}}\upsilon E_s}{\sigma_q^2 K}\Big), \tag{63}$$

where the upper bound can be obtained when considering a perfectly-known LoS channel.

Hence, based on (62) and (63), it can be shown that our proposed RIS-aided single RF-chain based architecture having N passive RIS elements achieves the same channel capacity as the conventional fully-digital MIMO system employing $N_{\mathrm{RF}} = N\xi = \frac{N}{\mathbb{E}[U_{\max}^2]}$ RF-chains.

B. Symbol Error Probability

Since the M-PSK modulation scheme is employed by our proposed RIS-aided single-RF downlink transmitter, the theoretical symbol error probability is given by [43]

$$\begin{aligned}
P_{\mathrm{e}} &= \mathbb{E}\Big[\frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} \exp\Big(-\rho \cdot \frac{\sin^2(\frac{\pi}{M})}{\sin^2 t}\Big)\mathrm{d}t\Big] \\
&= \mathbb{E}\Big[\frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} \exp\Big(-\frac{N\upsilon E_s \xi}{\sigma_q^2 K} \cdot \frac{\sin^2(\frac{\pi}{M})}{\sin^2 t}\Big)\mathrm{d}t\Big] \\
&= \frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} \int_0^{\infty} f_{U_{\max}}(u) \times \\
&\qquad \exp\Big(-\frac{N\upsilon_r E_s}{\sigma_n^2 K u^2} \cdot \frac{\sin^2(\frac{\pi}{M})}{\sin^2 t}\Big)\mathrm{d}u\mathrm{d}t. \\
&\overset{(a)}{\leq} \frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} \exp\Big(-\frac{N\upsilon E_s}{\sigma_q^2 K \mathbb{E}[U_{\max}^2]} \cdot \frac{\sin^2(\frac{\pi}{M})}{\sin^2 t}\Big)\mathrm{d}t \\
&\overset{(b)}{\leq} \frac{2(M-1)}{M} Q\Bigg(\sqrt{\frac{2N\upsilon E_s}{\sigma_q^2 K \mathbb{E}[U_{\max}^2]}} \cdot \sin(\frac{\pi}{M})\Bigg),
\end{aligned} \tag{64}$$

where (a) holds, since $f(x) = \exp(-\frac{1}{x})$ is a concave function, and (b) applies the Chernoff bound to the Gaussian Q-function [43]. It can be shown that the RIS-based modulation scheme experiences a symbol error performance degradation if either the modulation order M or the number of users K are increased. However, RISs benefit from having large values of N to counteract the detrimental effect of increasing the modulation order. This increases the energy efficiency of Internet of things networks, which can rely on high-order constellations to support extremely high data rates and large number of users. Furthermore, based on (64), when the number of RIS elements obeys $N \to \infty$, the symbol error probability exponentially decreases with the increase of the transmit power E_s.

On the other hand, the symbol error probability lower bound of the high-complexity fully-digital MIMO systems is given by [43]

$$\begin{aligned}
P_{\mathrm{e,MIMO}}^{(\mathrm{LB})} &= \frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} \exp\Big(-\frac{N_{\mathrm{RF}}\upsilon E_s}{\sigma_q^2 K} \cdot \frac{\sin^2(\frac{\pi}{M})}{\sin^2 t}\Big)\mathrm{d}t \\
&\overset{(a)}{\leq} \frac{2(M-1)}{M} Q\Bigg(\sqrt{\frac{2N_{\mathrm{RF}}\upsilon E_s}{\sigma_q^2 K}} \cdot \sin(\frac{\pi}{M})\Bigg),
\end{aligned} \tag{65}$$

where (a) applies the Chernoff bound to the Gaussian Q-function [43].

Similar to the ergodic capacity analysis in (62), it can be shown from (64) and (65) that our proposed RIS-aided single RF-chain architecture having N passive RIS elements achieves the same symbol error probability performance as the high-complexity fully-digital MIMO system employing $N_{\mathrm{RF}} = N\xi = \frac{N}{\mathbb{E}[U_{\max}^2]}$ RF-chains.

C. Outage Probability

The outage probability, denoted as P_{out}, is defined as the probability that the SINR ρ is lower than a given threshold SINR ρ_{th}, i.e. $P_{\mathrm{out}} = \Pr(\rho < \rho_{\mathrm{th}})$.

According to (44) and (60), the outage probability is formulated as

$$\begin{aligned}
P_{\mathrm{out}} &= \Pr(\rho < \rho_{\mathrm{th}}) \\
&= \Pr\Big(\sqrt{\frac{1}{\xi}} > \sqrt{\frac{N\upsilon E_s}{\sigma_q^2 K \rho_{\mathrm{th}}}}\Big) \\
&= 1 - F_{U_{\max}}\Bigg(\sqrt{\frac{N\upsilon E_s}{\sigma_q^2 K \rho_{\mathrm{th}}}}\Bigg) \\
&= 1 - \Bigg[\frac{\Gamma_{\mathrm{li}}\Big(\sqrt{\frac{N\upsilon E_s}{\sigma_q^2 K \nu^2 \rho_{\mathrm{th}}}}; \mu\Big)}{\Gamma(\mu)}\Bigg]^N.
\end{aligned} \tag{66}$$

Observe in (66) that the outage probability tends to 0 when the transmit power obeys $E_s \to \infty$ or the number of RIS elements obeys $N \to \infty$.

For comparison, the outage probability of conventional fully-digital MIMO systems in the LoS channels, which can be viewed as the lower bound outage probability of practical fading channels, denoted as $P_{\mathrm{out,MIMO}}^{(\mathrm{LB})}$, is given by

$$P_{\mathrm{out,MIMO}}^{(\mathrm{LB})} = \begin{cases} 1 & , \ E_s < \frac{\sigma_q^2 K \rho_{\mathrm{th}}}{\upsilon N_{\mathrm{RF}}} \\ 0 & , \ E_s \geq \frac{\sigma_q^2 K \rho_{\mathrm{th}}}{\upsilon N_{\mathrm{RF}}} \end{cases}. \tag{67}$$

D. Coverage Range

The coverage range, denoted as d_{cov}, is defined as the maximum distance, where the SINR ρ is not lower less than

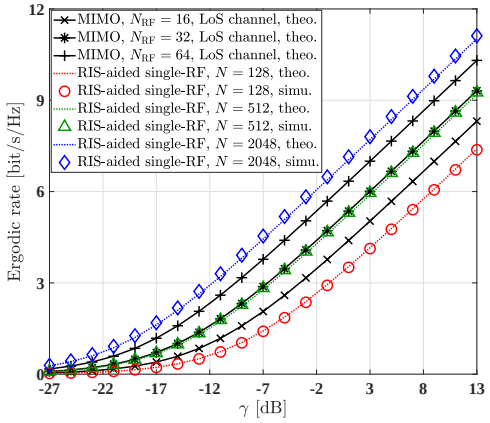

Fig. 3. Comparison of the ergodic rate versus γ for the conventional MIMO and the proposed RIS-based single-RF downlink transmitter system.

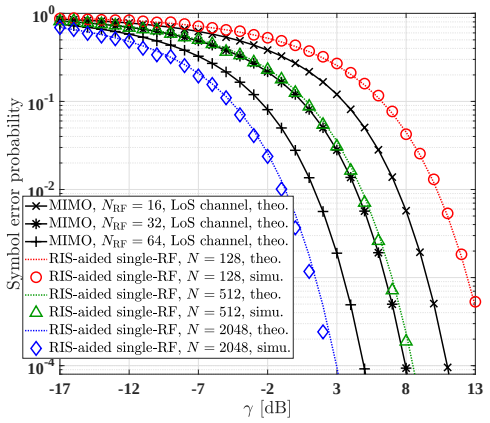

Fig. 4. Comparison of the symbol error probability versus γ for the conventional MIMO and the proposed RIS-based single-RF downlink transmitter system.

a given threshold SINR ρ_{th} with the probability of $(1 - P_{\text{out}})$. Thus,

$$\Pr\left(\frac{E_s \upsilon d_{\text{cov}}^{-\alpha_r} N \xi}{\sigma_q^2 K} \geq \rho_{\text{th}}\right) = 1 - P_{\text{out}}. \tag{68}$$

Then, the coverage range can be formulated as

$$d_{\text{cov}} = \left(\sqrt{\frac{\sigma_q^2 K \rho_{\text{th}}}{E_s \upsilon N}} F_{U_{\max}}^{-1}(1 - P_{\text{out}})\right)^{-\frac{2}{\alpha_r}}, \tag{69}$$

where $F_{U_{\max}}^{-1}(\cdot)$ represents the inverse function of $F_{U_{\max}}(\cdot)$. According to (60) and (69), we can show that

$$F_{U_{\max}}^{-1}(P_{\text{out}}) = \Gamma_{\text{li}}^{-1}(\nu \Gamma(\mu) \sqrt[N]{1 - P_{\text{out}}}; \mu), \tag{70}$$

where $\Gamma_{\text{li}}^{-1}(\cdot; \mu)$ represents the inverse function of $\Gamma_{\text{li}}(\cdot; \mu)$. Then, using (69) and (70), we can formulate the coverage range as

$$d_{\text{cov}} = \left(\frac{E_s \upsilon N}{\sigma_q^2 K \rho_{\text{th}} \left(\Gamma_{\text{li}}^{-1}(\nu \Gamma(\mu) \sqrt[N]{1 - P_{\text{out}}}; \mu)\right)^2}\right)^{\frac{1}{\alpha_r}}. \tag{71}$$

For comparison, the coverage range upper bound of the conventional fully-digital MIMO systems, denoted as $d_{\text{cov,MIMO}}^{(\text{UB})}$, should satisfy that

$$\left(\frac{E_s \upsilon N_{\text{RF}}}{\sigma_q^2 K}\right)\left(d_{\text{cov,MIMO}}^{(\text{UB})}\right)^{-\alpha_r} = \rho_{\text{th}}. \tag{72}$$

According to (72), we arrive at:

$$d_{\text{cov,MIMO}}^{(\text{UB})} = \left(\frac{E_s \upsilon N_{\text{RF}}}{\sigma_q^2 K \rho_{\text{th}}}\right)^{\frac{1}{\alpha_r}}. \tag{73}$$

It can be shown from (71) and (73) that our proposed RIS-aided single RF-chain architecture having N passive RIS elements achieves the same coverage range as the conventional fully-digital MIMO system employing $N_{\text{RF}} = \frac{N}{\left(\Gamma_{\text{li}}^{-1}(\nu \Gamma(\mu) \sqrt[N]{1 - P_{\text{out}}}; \mu)\right)^2}$ RF-chains.

V. Performance Results and Analysis

In this section, the theoretical and simulation results of the ergodic rate, symbol error probability, outage probability, and coverage range of our system are presented, and compared to that of the conventional MIMO systems. Then, we characterize the achievable rate and symbol error probability of three power sharing methods, i.e. of maximizing the sum-rate, maximizing the min-rate and maximizing the geometric-mean-rate, in our proposed multi-user transmitter systems.

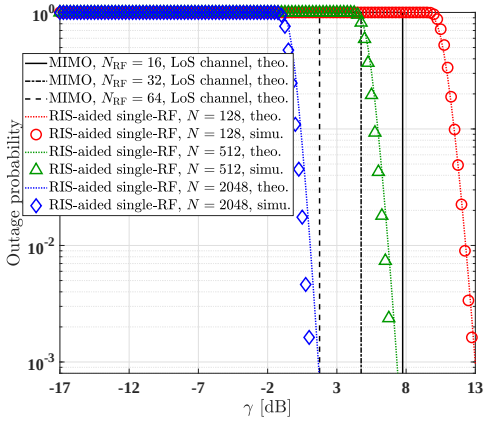

Fig. 5. Comparison of the outage probability versus γ for the conventional MIMO and the proposed RIS-based single-RF downlink transmitter system.

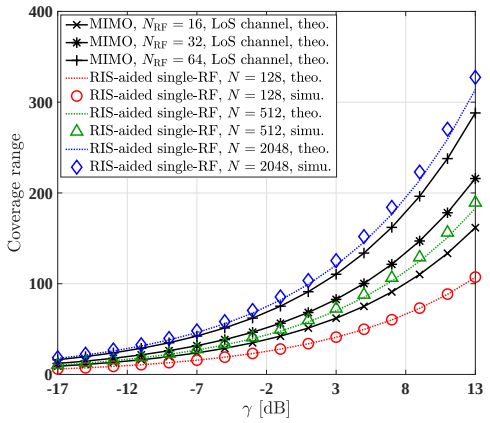

Fig. 6. Comparison of the coverage range versus γ for the conventional MIMO and the proposed RIS-based single-RF downlink transmitter system.

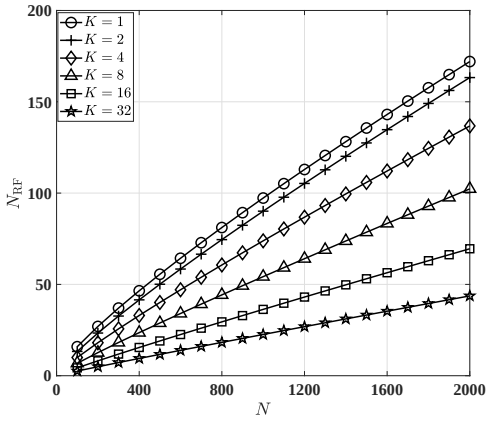

Fig. 7. The relationship between the number of RF-chains N_{RF} in conventional MIMO systems and the number of RIS elements N in our proposed RIS-aided single-RF transmitter scheme when the same ergodic rate is achieved.

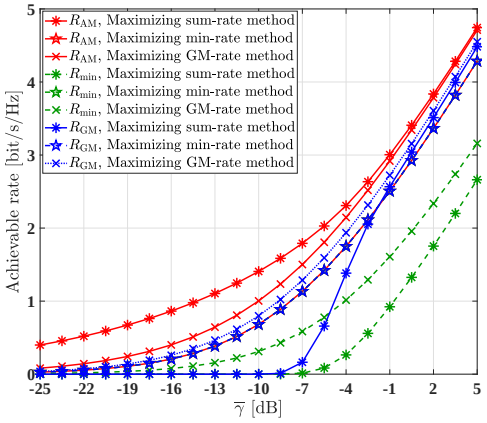

Fig. 8. The simulation comparison of arithmetic-mean-rate R_{AM}, min-rate $R_{\min}$, and geometric-mean-rate R_{GM} versus $\overline{\gamma}$, when considering different power allocation methods.

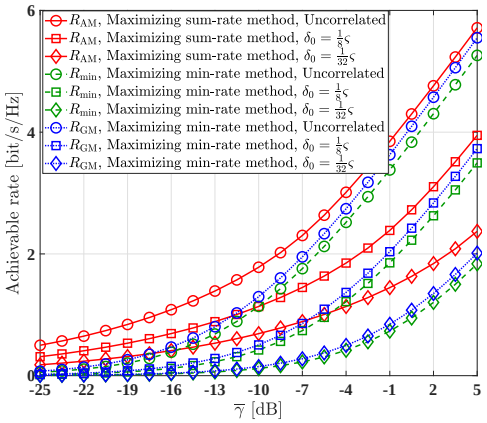

Fig. 9. The simulation comparison of arithmetic-mean-rate R_{AM}, min-rate $R_{\min}$, and geometric-mean-rate R_{GM} versus $\overline{\gamma}$, for different channel correlation levels.

A. Comparison of the Theoretical and Simulation Results

The theoretical performance of our proposed RIS-aided transmitter architecture and the performance upper bound of the conventional MIMO systems, derived in Section IV, are compared in this section. Their accuracy is also verified by simulations. Based on our comparison, we present the number of RIS elements required for achieving the same rate as conventional MIMO systems. Unless otherwise specified, the system parameters employed in this section are set as: the number of users is $K = 8$, the PSK modulation order is $M = 16$, the distance between the BS and the users is $d = 100$m, the path-loss at the reference distance of 1 meter is $\varrho_0 = 30$dB, the path-loss exponent is $\alpha = 2.4$, the noise power is $\sigma_q^2 = -90$dBm, the threshold SINR is $\rho_{\mathrm{th}} = 20$dB, the threshold outage probability is $P_{\mathrm{out}} = 0.001$. The number of RIS elements in our proposed architecture and the number of RF-chains in the conventional MIMO are set to $N = 128, 512, 2048$ and $N_{\mathrm{RF}} = 16, 32, 64$, respectively. Similarly to [19]–[22], we consider the worse channel fading case, i.e. the Rician factor $\kappa = 0$. Furthermore, we define the SNR at the receiver side as $\gamma = \frac{\upsilon E_s}{\sigma_q^2 K}$.

In Figs. 3 - 6 we show the theoretical (theo.) as well as simulation (simu.) results for the ergodic rate, symbol error probability, outage probability, and coverage range versus the SNR at the receiver side γ for both our proposed architecture and for the conventional MIMO systems. As shown in these figures, the simulation results closely match the theoretical analysis. The performance of our proposed scheme improves upon increasing the number of RIS elements. Observe for our proposed method that when the number of RIS elements is doubled, a channel gain of 3dB is achieved, implying that the received power is proportional to the number of RIS elements, i.e. to the RIS area. This is similar to the conventional MIMO systems, where doubling the number of RF-chains brings about 3dB channel gain. When $N = 2048$ RIS elements are employed, our proposed architecture succeeds in outperforming the conventional MIMO systems using $N_{\mathrm{RF}} = 64$ RF-chains. It is important to note that only a single RF-chain is required in our proposed architecture, and the information is conveyed by using the passive RIS. Since the cost of RIS elements is considerably lower than that of active RF-chains, our proposed RIS-aided single RF-chain architecture has lower hardware cost than the conventional MIMO systems at the same performance. It shows that there is a slight difference between the theoretical analysis and the simulation results when the number of RIS elements $N = 2048$, which results from the employment of the moment-matching method for the distribution approximation of the received SINR.

Fig. 7 shows the relationship between the number of RF-chains N_{RF} in conventional MIMO systems and the number of RIS elements N in our proposed scheme, when the same rate is achieved. Based on the analysis of Section IV-A, we showed that $N_{\mathrm{RF}} = N\xi = \frac{N}{\mathbb{E}[U_{\max}^2]}$. Fig. 7 demonstrates that ξ (or $U_{\max}^2$) is determined by the number of users. With the increase of the number of users, more RIS elements are required by our proposed architecture to get the same rate as the conventional MIMO systems. For example, in the case of supporting $K = 32$ users, our proposed single RF-chain architecture having $N = 4000$ RIS elements outperforms the conventional MIMO systems having $N_{\mathrm{RF}} = 40$ RF-chains.

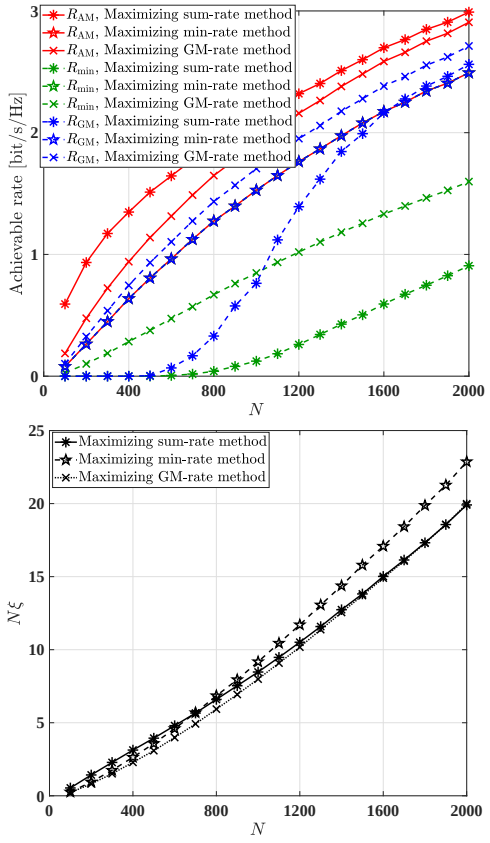

Fig. 10. The simulation comparison of arithmetic-mean-rate $R_{\rm AM}$, min-rate $R_{\rm min}$, geometric-mean-rate $R_{\rm GM}$, as well as $N\xi$, versus the number of RIS elements N for different power allocation methods, where $\overline{\gamma} = -5$dB.

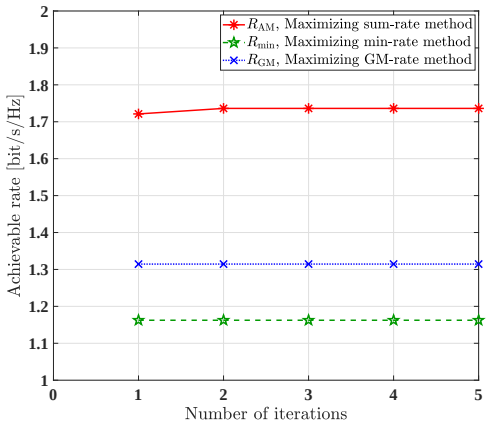

Fig. 11. Achievable rate convergence analysis when considering different power allocation methods.

B. Simulation Based Comparison of Different Power Sharing Methods in Our Proposed Architecture

In this section, we assume that the users are randomly distributed in a circular area, following PPP, as shown in Fig. 1. We denote the number of users K follows Poisson distribution with the mean of $\overline{K}$, i.e. $K \sim \text{Pois}(\overline{K})$, the distance from the RIS to the center of the circular coverage area as $\overline{d}$, the radius of the circular user area as D, and the wavelength of the carrier as ς. Unless otherwise specified, the system parameters employed in this section are set as: $\overline{K} = 50$, $M = 16$, $N = 1024$, $\varrho_0 = -30$dB, $\alpha = 2.4$, $\overline{d} = 100$m, $D = 90$m, the distance between adjacent RIS elements is $\frac{1}{4}\varsigma$, and the correlation level is $\delta_0 = \varsigma$. Similarly to [19]–[22], we consider the worse channel fading case, i.e. the Rician factor $\kappa = 0$. Furthermore, we define the average SNR at the receiver side as $\overline{\gamma} = \frac{\overline{\upsilon} E_s}{\sigma_q^2 \overline{K}}$ with $\overline{\upsilon} = \varrho_0 \overline{d}^{-\alpha}$. Thus, according to (28), the achievable rate of the kth user is given by $R_k = \log_2(1 + N\xi \frac{\upsilon_k}{\overline{\upsilon}} \overline{K} \lambda'_k \overline{\gamma})$, where λ'_k is determined by the channel conditions and the power sharing method employed.

Since maximizing the sum-rate is equivalent to maximizing the arithmetic-mean-rate, we define the arithmetic-mean-rate as $R_{\rm AM} = \frac{1}{K}\sum_{k=1}^{K} R_k$. Fig. 8 compares the simulation results of the arithmetic-mean-rate $R_{\rm AM}$, min-rate $R_{\rm min}$, and geometric-mean-rate $R_{\rm GM}$ versus the average SNR at the receiver side $\overline{\gamma}$, when considering the three power sharing methods. As expected, maximizing the sum-rate achieves the highest arithmetic-mean-rate, while maximizing min-rate achieves the highest min-rate, and maximizing geometric-mean-rate achieves the highest geometric-mean-rate. In the maximizing the sum-rate method, at low SNRs, more power is allocated to the users having good channel conditions, while at high SNRs, power is approximately evenly allocated to all users. Upon maximizing the min-rate, more power is allocated to the users with poor conditions. Maximizing the geometric-mean-rate strikes a tradeoff between maximizing the sum-rate and maximizing the min-rate. This shows that at low SNRs, maximizing geometric-mean-rate tends to the respects of maximizing the min-rate method, where more power is allocated to the users with poor condition. By contrast, at high SNRs, maximizing geometric-mean-rate tends to the respect of maximizing the sum-rate method in which the power is evenly allocated to all users.

Fig. 9 compares the arithmetic-mean-rate $R_{\rm AM}$, min-rate $R_{\rm min}$, and geometric-mean rate $R_{\rm GM}$ versus the average SNR at the receiver side $\overline{\gamma}$ for different channel correlation levels. Observe that the system performance degrades upon decreasing the distance between the adjacent RIS elements.

Fig. 10 compares the arithmetic-mean-rate $R_{\rm AM}$, min-rate $R_{\rm min}$, geometric-mean rate $R_{\rm GM}$, as well as $N\xi$, versus the number of RIS elements N for the different power sharing methods, where $\overline{\gamma} = -5$dB. Observe in Fig. 10 in terms of both the arithmetic-mean-rate and geometric-mean-rate, that the performance of maximizing the geometric-mean-rate tends to that of maximizing the min-rate, when the number of RIS elements N is small. By contrast, its performance tends to that of maximizing the sum-rate when the number of RIS elements N is large. In terms of the min-rate $R_{\rm min}$, naturally, maximizing the min-rate attains the best performance, while maximizing the sum-rate exhibits the worst performance. It may be surmised that $N\xi$ approximately linearly increases with the number of RIS elements N.

Fig. 11 shows the convergence of the achievable rate when considering different power allocation methods at $\overline{\gamma} = -7$dB. It is shown that the achievable rate converges over multiple iterations by applying Algorithm 1. It is worth noting that the achievable rate attained by maximizing the min-rate method

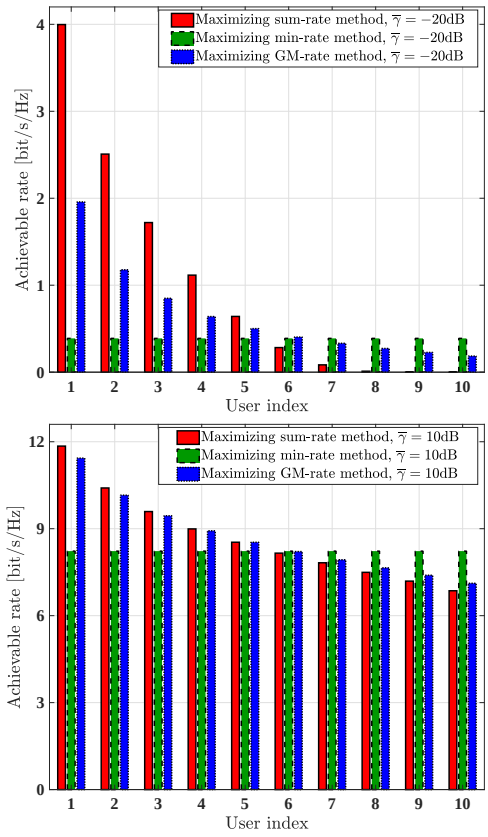

Fig. 12. The simulation comparison of the average achievable rate in each user for different power allocation methods, where the users indices are sorted based on their average achievable rate in the descending order.

converges in the first iteration, since the optimization of the power sharing ratio $\lambda'_1, \cdots, \lambda'_K$ in (P2.b) is independent of ξ.

Finally, Fig. 12 compares the average achievable rate of each user for the different power allocation methods, where the number of users are fixed as $K = 10$, and the user indices are sorted based on their average achievable rate in the descending order. Thus, according to (28), the achievable rate of the kth user is given by $R_k = \log_2(1+N\xi\frac{\upsilon_k}{\upsilon}K\lambda'_k\overline{\gamma})$, where $\lambda_1, \cdots, \lambda_K$ are sorted in a descending order, and the values of $N\xi$ are numerically calculated as approximately 41, 46, 41 for the method maximizing the sum-rate, min-rate and geometric-mean-rate respectively when $\overline{\gamma} = -20$dB. By contrast, they are approximately 58, 46, 42 for the method maximizing the sum-rate, min-rate and geometric-mean-rate respectively, when $\overline{\gamma} = 10$dB. This shows that maximizing the min-rate ensures that all the users have similar achievable rate, while the method maximizing the sum-rate exhibits rather unfair rate allocation across the users. Specifically, in the low SNR region, i.e. $\overline{\gamma} = -20$dB, the users with good channel condition can get high achievable rate, while the users with poor channel condition are effectively disconnected. By contrast, in the high SNR region, i.e. $\overline{\gamma} = 10$dB, the achievable rate of all users tends to be similar. The method maximizing the geometric-rate can be viewed as a compromise between maximizing the sum-rate and the min-rate.

VI. Conclusions

A RIS-aided single-RF downlink scheme supporting multiple users has been proposed. Compared to the conventional massive MIMO systems, our architecture has significantly reduced the hardware complexity, since only a single RF-chain is required for illuminating the RIS. The baseband digital signals are precoded by the ZF method and modulated by configuring the amplitude and phase shift of each RIS element. We assumed that the distribution of multiple users obeys the PPP, and the power allocation algorithms, including maximizing the sum-rate, maximizing the min-rate and maximizing the geometric-mean-rate, are designed based on the alternating optimization method by jointly optimizing the power allocation ratios $\lambda'_1, \cdots, \lambda'_K$ and the total power $E_s\xi$ reflected by the RIS, under the constraint that the amplitude of each RIS element is not higher than unity. We found that the average ergodic rate can be maximized by the method maximizing the sum-rate, while the method maximizing the min-rate can ensure fairness for all users. The method maximizing the geometric-mean-rate strikes a compelling compromise between maximizing the sum-rate and the min-rate. Furthermore, we provided both theoretical analysis and simulation results for characterizing the performance of our proposed RIS-aided single RF-chain based method, and compared it to the performance upper bound of the conventional MIMO systems. Additionally, we theoretically derived the number of RIS elements required by our proposed single RF-chain based architecture to achieve the same channel capacity as the conventional MIMO systems.

References

[1] S. Lu, M. El-Hajjar, and L. Hanzo, "Two-dimensional index modulation for the large-scale multi-user MIMO uplink," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7904–7918, 2019.

[2] H. Liu, Y. Zhang, X. Zhang, M. El-Hajjar, and L.-L. Yang, "Deep learning assisted adaptive index modulation for mmwave communications with channel estimation," *IEEE Transactions on Vehicular Technology*, 2022.

[3] S. Feng, R. Zhang, W. Xu, and L. Hanzo, "Multiple access design for ultra-dense VLC networks: Orthogonal vs non-orthogonal," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2218–2232, 2018.

[4] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G Internet of Things: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 359–383, 2021.

[5] I. F. Akyildiz, C. Han, and S. Nie, "Combating the distance problem in the millimeter wave and Terahertz frequency bands," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 102–108, 2018.

[6] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, F. Chen *et al.*, "Adaptive transmission for reconfigurable intelligent surface-assisted OFDM wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2653–2665, 2020.

[7] G. Yu, X. Chen, C. Zhong, D. W. K. Ng, and Z. Zhang, "Design, analysis, and optimization of a large intelligent reflecting surface-aided B5G cellular Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8902–8916, 2020.

[8] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, "Reconfigurable intelligent surface aided NOMA networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2575–2588, 2020.

[9] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, M. Di Renzo, and F. Chen, "Reconfigurable intelligent surfaces with reflection pattern modulation: Beamforming design and performance analysis," *IEEE transactions on wireless communications*, vol. 20, no. 2, pp. 741–754, 2020.

[10] X. Cao, B. Yang, H. Zhang, C. Huang, C. Yuen, and Z. Han, "Reconfigurable intelligent surface-assisted MAC for wireless networks: Protocol design, analysis, and optimization," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14 171–14 186, 2021.

[11] Y. Chen, M. Wen, E. Basar, Y.-C. Wu, L. Wang, and W. Liu, "Exploiting reconfigurable intelligent surfaces in edge caching: Joint hybrid beamforming and content placement optimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 7799–7812, 2021.

[12] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Maximizing the geometric mean of user-rates to improve rate-fairness: Proper vs. improper gaussian signaling," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 295–309, 2021.

[13] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. Mourad, B. Clerckx, and L. Hanzo, "Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6367–6383, 2022.

[14] S. Xu, J. Liu, and Y. Cao, "Intelligent reflecting surface empowered physical layer security: Signal cancellation or jamming?" *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1265–1275, 2021.

[15] X. Pang, M. Sheng, N. Zhao, J. Tang, D. Niyato, and K.-K. Wong, "When UAV meets IRS: Expanding air-ground networks via passive reflection," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 164–170, 2021.

[16] X. Pang, N. Zhao, J. Tang, C. Wu, D. Niyato, and K.-K. Wong, "IRS-assisted secure UAV transmission via joint trajectory and beamforming design," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1140–1152, 2021.

[17] X. Pang, W. Mei, N. Zhao, and R. Zhang, "Intelligent reflecting surface assisted interference mitigation for cellular-connected UAV," *IEEE Wireless Communications Letters*, 2022.

[18] Z. Wang, Y. Shi, Y. Zhou, H. Zhou, and N. Zhang, "Wireless-powered over-the-air computation in intelligent reflecting surface-aided IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1585–1598, 2020.

[19] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[20] E. Basar, "Transmission through large intelligent surfaces: A new frontier in wireless communications," in *2019 European Conference on Networks and Communications (EuCNC)*. IEEE, 2019, pp. 112–117.

[21] A. Khaleel and E. Basar, "Reconfigurable intelligent surface-empowered MIMO systems," *IEEE Systems Journal*, vol. 15, no. 3, pp. 4358–4366, 2020.

[22] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond MIMO paradigm for 6G," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 3187–3196, 2020.

[23] J. Yuan, M. Wen, Q. Li, E. Basar, G. C. Alexandropoulos, and G. Chen, "Receive quadrature reflecting modulation for RIS-empowered wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 5121–5125, 2021.

[24] W. Tang, J. Y. Dai, M. Chen, X. Li, Q. Cheng, S. Jin, K.-K. Wong, and T. J. Cui, "Programmable metasurface-based RF chain-free 8PSK wireless transmitter," *Electronics Letters*, vol. 55, no. 7, pp. 417–420, 2019.

[25] W. Tang, J. Y. Dai, M. Z. Chen, K.-K. Wong, X. Li, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, "MIMO transmission through reconfigurable intelligent surface: System design, analysis, and implementation," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2683–2699, 2020.

[26] W. Tang, J. Y. Dai, M. Z. Chen, Y. Han, X. Li, C.-K. Wen, S. Jin, Q. Cheng, and T. J. Cui, "Realization of reconfigurable intelligent surface-based Alamouti space-time transmission," in *2020 International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE, 2020, pp. 904–909.

[27] R. Liu, M. Li, Q. Liu, L. Swindlehurst, and Q. Wu, "Intelligent reflecting surface based passive information transmission: A symbol-level precoding approach," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6735–6749, 2021.

[28] J. Wang, H. Wang, Y. Han, S. Jin, and X. Li, "Joint transmit beamforming and phase shift design for reconfigurable intelligent surface assisted MIMO systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 354–368, 2021.

[29] J. R. Hampton, *Introduction to MIMO communications*. Cambridge university press, 2013.

[30] X. Li, Z. Xie, Z. Chu, V. G. Menon, S. Mumtaz, and J. Zhang, "Exploiting benefits of IRS in wireless powered NOMA networks," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 1, pp. 175–186, 2022.

[31] H. Liu, G. Li, X. Li, Y. Liu, G. Huang, and Z. Ding, "Effective capacity analysis of STAR-RIS-assisted NOMA networks," *IEEE Wireless Communications Letters*, vol. 11, no. 9, pp. 1930–1934, 2022.

[32] X. Chen, "Antenna correlation and its impact on multi-antenna system," *Progress In Electromagnetics Research B*, vol. 62, pp. 241–253, 2015.

[33] J. An, C. Xu, L. Gan, and L. Hanzo, "Low-complexity channel estimation and passive beamforming for RIS-assisted MIMO systems relying on discrete phase shifts," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1245–1260, 2021.

[34] D. Pozar, "Rectangular cavity modes," *Microwave Engineering, 3rd ed.; John Wiley and Sons: New York, NY, USA*, p. 120, 2005.

[35] S. Shen, B. Clerckx, and R. Murch, "Modeling and architecture design of reconfigurable intelligent surfaces using scattering parameter network analysis," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1229–1243, 2021.

[36] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849–5863, 2020.

[37] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for intelligent reflecting surface enhanced wireless networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[38] F. Liu, O. Tsilipakos, A. Pitilakis, A. C. Tasolamprou, M. S. Mirmoosa, N. V. Kantartzis, D.-H. Kwon, J. Georgiou, K. Kossifos, M. A. Antoniades *et al.*, "Intelligent metasurfaces with continuously tunable local surface impedance for multiple reconfigurable functions," *Physical Review Applied*, vol. 11, no. 4, p. 044024, 2019.

[39] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[40] H. Jeffreys, *The theory of probability*. OUP Oxford, 1998.

[41] H. A. David and H. N. Nagaraja, *Order statistics*. John Wiley & Sons, 2004.

[42] W. Y. Yang, W. Cao, J. Kim, K. W. Park, H.-H. Park, J. Joung, J.-S. Ro, H. L. Lee, C.-H. Hong, and T. Im, *Applied numerical methods using MATLAB*. John Wiley & Sons, 2020.

[43] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels*. John Wiley & Sons, 2005, vol. 95.