



Fig. 1. An RIS-assisted point-to-point MIMO communication system.

predesigned codebook for RIS-assisted point-to-point MIMO communications. *In contrast to the direct codeword selection from the predesigned codebook, the proposed scheme designs a set of weights for each codeword in the codebook according to the corresponding output results, yielding a new RIS RC vector.* For illustration, we contrast the proposed scheme to its existing channel estimation, passive beamforming and predesigned codebook counterparts in Table I. The simulation results demonstrate that the proposed scheme has improved performance compared to existing passive beamforming and codebook schemes.

II. SYSTEM MODEL

We consider an RIS-assisted point-to-point MIMO system in a single cell as shown in Fig. 1, where a base station (BS) with M_t transmit antennas sends M_s data streams to a user with M_r antennas, with $M_s \leq \min\{M_t, M_r\}$. The RIS consists of N reflecting elements and is equipped with a smart controller capable of adjusting the RCs according to instructions from the BS. The signals from both the cascaded BS-RIS-user link and the direct BS-user link are superimposed at the user. We assume that the frequency-flat baseband equivalent channels spanning from the BS to the RIS, from the RIS to the user, and from the BS to the user are denoted by $\mathbf{H}_t \in \mathbb{C}^{N \times M_t}$, $\mathbf{H}_r \in \mathbb{C}^{M_r \times N}$ and $\mathbf{H}_d \in \mathbb{C}^{M_r \times M_t}$, respectively.

Let $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \dots, \varphi_N]^T$ represent the RIS RC vector, where $\varphi_n = e^{j\theta_n}$ denotes the RC of the n th RIS element with phase shift θ_n , satisfying $\theta_n \in [0, 2\pi)$ for $n = 1, 2, \dots, N$. Thus the composite end-to-end channel $\mathbf{H}_e \in \mathbb{C}^{M_r \times M_t}$ from the BS to the user can be expressed as

$$\mathbf{H}_e = \mathbf{H}_d + \mathbf{H}_r \text{diag}(\boldsymbol{\varphi}) \mathbf{H}_t. \quad (1)$$

During the channel training process in the uplink phase, the user sends the pilot signal $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{M_r}]^T \in \mathbb{C}^{M_r \times \tau}$ to the BS, where $\mathbf{x}_m^T \in \mathbb{C}^{1 \times \tau}$, $m = 1, 2, \dots, M_r$ is the pilot loaded on the m th antenna at the user. The pilot matrix satisfies $\|\mathbf{X}\|_F^2 = \tau p_u$, where p_u is the average pilot power. As we consider a time-division duplexing protocol for both uplink as well as downlink transmissions and assume the channel's reciprocity, the pilot signal received at the BS is given by

$$\mathbf{Y} = \mathbf{H}_e^H \mathbf{X} + \mathbf{N}_{\text{BS}}, \quad (2)$$

where $\mathbf{N}_{\text{BS}} \in \mathbb{C}^{M_t \times \tau}$ denotes the noise matrix at the BS with an average noise power of σ_{BS}^2 , whose i th column vector follows $\mathbf{n}_{\text{BS},i} \sim \mathcal{CN}(\mathbf{0}_{M_t}, \sigma_{\text{BS}}^2 \mathbf{I}_{M_t})$, for $i = 1, 2, \dots, \tau$. We employ mutually orthogonal pilots, and the length of the pilot signal is designed such that $\tau \geq M_r$ [12].

Next, we consider the downlink of data transmission where the BS applies a baseband precoder $\mathbf{W} \in \mathbb{C}^{M_t \times M_s}$ to transmit symbol $\mathbf{s} \in \mathbb{C}^{M_s \times 1}$, with $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_{M_s}$. Furthermore, the precoder satisfies $\|\mathbf{W}\|_F^2 \leq p_d$ and p_d is the total transmit power at the BS. Thus the received signal at the user is obtained as

$$\mathbf{r} = \mathbf{H}_e \mathbf{W} \mathbf{s} + \mathbf{n}_{\text{UE}}, \quad (3)$$

where $\mathbf{n}_{\text{UE}} \in \mathbb{C}^{M_r \times 1}$ is the noise at the user with an average noise power of σ_{UE}^2 , satisfying $\mathbf{n}_{\text{UE}} \sim \mathcal{CN}(\mathbf{0}_{M_r}, \sigma_{\text{UE}}^2 \mathbf{I}_{M_r})$.

Meanwhile, we adopt the Rician channel in this paper. Specifically, the RIS-user channel can be expressed as

$$\mathbf{H}_r = \sqrt{\beta_r} \left(\sqrt{\frac{F_r}{F_r + 1}} \mathbf{H}_r^{\text{LoS}} + \sqrt{\frac{1}{F_r + 1}} \mathbf{H}_r^{\text{NLoS}} \right), \quad (4)$$

where β_r and F_r are the path loss and the Rician factor of RIS-user channel, respectively; $\mathbf{H}_r^{\text{LoS}} \in \mathbb{C}^{M_r \times N}$ and $\mathbf{H}_r^{\text{NLoS}} \in \mathbb{C}^{M_r \times N}$ represent the line-of-sight (LoS) and the non-line-of-sight (NLoS) components of the RIS-user channel, respectively. The element on the m_r th row and the n th column of the NLoS matrix is modeled by Rayleigh fading, which follows $\mathbf{H}_r^{\text{NLoS}}_{m_r, n} \sim \mathcal{CN}(0, 1)$. Similarly, the BS-user channel and BS-RIS channel can be modeled by using (4).

Moreover, we consider a uniform linear array (ULA) at the BS, a ULA at the user, and a uniform planar array (UPA) at the RIS. Let $\mathbf{a}_{\text{BS}}(\delta) \in \mathbb{C}^{M_t \times 1}$, $\mathbf{a}_{\text{UE}}(\delta) \in \mathbb{C}^{M_r \times 1}$ and $\mathbf{a}_{\text{R}}(\zeta, \gamma) \in \mathbb{C}^{N \times 1}$ denote the steering vector of the BS, the user and the RIS, respectively. Specifically, the m_t th entry of \mathbf{a}_{BS} is denoted as $e^{j\frac{2\pi}{\lambda}(m_t-1)d_{\text{BS}}\sin(\delta)}$, $m_t = 1, 2, \dots, M_t$, where d_{BS} denotes the element spacing of the BS, λ denotes the signal wavelength, and $\delta \in [-\pi/2, \pi/2)$ denotes the angle of departure (AoD) or the angle of arrival (AoA). Similarly, the m_r th entry of \mathbf{a}_{UE} is denoted as $e^{j\frac{2\pi}{\lambda}(m_r-1)d_{\text{UE}}\sin(\delta)}$, $m_r = 1, 2, \dots, M_r$, where d_{UE} denotes the element spacing of the user. The n th entry of \mathbf{a}_{R} is denoted as $e^{j2\pi d_{\text{R}}\sin(\gamma)[\lfloor \frac{n-1}{N_x} \rfloor \sin(\zeta) + ((n-1) - \lfloor \frac{n-1}{N_x} \rfloor N_x) \cos(\zeta)]/\lambda}$, $n = 1, 2, \dots, N$, where d_{R} denotes the element spacing of the RIS. N_x is the number of elements deployed at each row of the RIS. $\zeta \in [0, \pi)$ and $\gamma \in [-\pi/2, \pi/2)$ denote the azimuth and elevation AoA/AoD, respectively. Thus, the LoS component of the \mathbf{H}_t , \mathbf{H}_r and \mathbf{H}_d are given by $\mathbf{a}_{\text{R}}(\zeta_t^{\text{AoA}}, \gamma_t^{\text{AoA}}) \mathbf{a}_{\text{BS}}(\delta_t^{\text{AoD}})^H$, $\mathbf{a}_{\text{UE}}(\delta_r^{\text{AoA}}) \mathbf{a}_{\text{R}}(\zeta_r^{\text{AoD}}, \gamma_r^{\text{AoD}})^H$ and $\mathbf{a}_{\text{UE}}(\delta_d^{\text{AoA}}) \mathbf{a}_{\text{BS}}(\delta_d^{\text{AoD}})^H$, respectively, where δ_t^{AoD} , ζ_t^{AoA} and γ_t^{AoA} represent the AoD, the azimuth and elevation AoA from the BS to the RIS, respectively; δ_r^{AoA} , ζ_r^{AoD} and γ_r^{AoD} represent the AoA, the azimuth and elevation AoD from the RIS to the user, respectively; δ_d^{AoA} and δ_d^{AoD} represent the AoA and the AoD from the BS to the user, respectively.

In the next section, unlike traditional passive beamforming and codebook schemes, the proposed scheme maximizes the channel capacity of the point-to-point MIMO systems by