Ergodic Spectral Efficiency Analysis of Intelligent Omni-Surface Aided Systems Suffering From Imperfect CSI and Hardware Impairments

Qingchao Li, *Graduate Student Member, IEEE*, Mohammed El-Hajjar, *Senior Member, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

Abstract—In contrast to the conventional reconfigurable intelligent surfaces (RIS), intelligent omni-surfaces (IOS) are capable of full-space coverage of smart radio environments by simultaneously transmitting and reflecting the incident signals. In this paper, we investigate the ergodic spectral efficiency of IOS-aided systems for transmission over random channel links, while considering both realistic imperfect channel state information (CSI) and transceiver hardware impairments (HWIs). Firstly, we formulate the linear minimum mean square error estimator of the equivalent channel spanning from the user equipments (UEs) to the access point (AP), where the transceiver HWIs are also considered. Then, we apply a two-timescale protocol for designing the beamformer of the IOS-aided system. Specifically, for the active AP beamformer, the minimum mean square error combining method is employed, which relies on the estimated equivalent channels, on the statistical information of the channel estimation error, on the inter-user interference as well as on the HWIs at the AP and UEs. By contrast, the passive IOS beamformer is designed based on the statistical CSI for maximizing the upper bound of the ergodic spectral efficiency. The theoretical analysis and simulation results show that the transceiver HWIs have a significant effect on the ergodic spectral efficiency, especially in the high transmit power region. Furthermore, we show that the HWIs at the AP can be effectively compensated by deploying more AP antennas.

Index Terms—Intelligent omni-surface (IOS), ergodic spectral efficiency, channel estimation, beamforming design, hardware impairment (HWI).

I. INTRODUCTION

DRIVEN by the ever-increasing demand for capacity in the mobile networks, the massive multiple-input and multiple-output (MIMO) technique [1] has been employed in the fifth generation (5G) wireless systems for improving the spectral efficiency. However, the high energy consumption and hardware cost constitute major bottlenecks for the implementation of massive MIMO systems. As a potent technique, the reconfigurable intelligent surface (RIS) concept has been researched as a promising cost-efficient technique for wireless communications [2]–[6].

The RIS is an artificial metamaterial surface comprised of numerous low-cost passive reflecting elements deployed at an elevated position between the access point (AP) and the user equipment (UE) [7]–[11]. By electronically tuning the phase shift of each RIS element, the signal propagation environment can be beneficially configured to realize increased information transmission reliability, especially when the direct AP-UE links are blocked [12], [13].

A. Prior Works

Various aspects of the RIS have been investigated, such as its channel modeling [14], channel estimation [15], [16], resource allocation [17], performance analysis [18], [19], and covert communications capacity [20], [21]. More particularly, by jointly designing the active beamformer of the AP and the passive beamformer constituted by the RIS, the transmission quality can be improved. In [22], Wu and Zhang investigated the problem of minimizing the transmission power by alternatively optimizing the transmit precoder (TPC) of the AP and the RIS phase shift matrix. Specifically, when the active AP TPC metrix is fixed, the RIS phase shift matrix is optimized by the semidefinite relaxation (SDR) algorithm. By contrast, when the RIS phase shift matrix is fixed, the active AP TPC matrix is designed based on the minimum mean squared error (MMSE) criterion. In [23], Li *et al.* researched the problem of deploying multiple RISs for improving the weighted sum-rate of multiple users. The AP TPC and the RIS phase shift were alternately optimized, where the Lagrangian method and the Riemannian manifold conjugate gradient (RMCG) method were employed. The simulation results showed that the performance of multi-RIS aided wireless communication systems is better than that of single-RIS aided systems. In contrast to the RIS phase shift optimization based on instantaneous channel state information (CSI), Han *et al.* [22], [23] focused their attention on designing the phase shift matrix based on the statistics of the angle of departure (AoD) and angle of arrival (AoA), which substantially reduced the channel acquisition overhead [24]. Specifically, the maximum-ratio transmission (MRT) criterion was employed for the design of the active beamformer at the AP, based on which a tight upper bound of the ergodic spectral efficiency was derived. Then, the RIS phase shift can be optimized for maximizing the ergodic spectral efficiency based on the statistical CSI.

B. Motivations and Contributions

Most of the existing treatises focus on the RIS family only capable of reflecting the impinging signals, but not transmitting them. This means that the RIS can work only when the AP and the UEs are at the same side of the RIS, which can only realize $180°$ half-space coverage of the smart radio environment. To deal with this issue, the intelligent omni-surface (IOS), also called a simultaneously transmitting and reflecting (STAR) RIS, has been proposed to realize $360°$ full coverage [25], [26]. Explicitly, in the IOS-aided systems, the incident signals are not only reflected but also refracted by the metasurfaces for full-space coverage. In [27], Mu *et al.* proposed three practical operating protocols for the IOS, namely energy splitting (ES), time switching (TS) and mode stitching (MS). More specifically, in the ES protocol, the energy of signals incident on each IOS element is spitted into two parts, where one of them is configured for reflecting signals to support the receivers at the same side of the transmitter. The other one is configured for transmitting signals to support the receivers at the opposite side of the transmitter. By contrast, in the TS protocol, orthogonal time slots are employed for each IOS element switching between the reflecting mode and the transmitting mode. Finally, in the MS protocol the IOS elements are divided into two partitions, where the elements in one of the partitions is permanently configured to support the receivers at the same side of the transmitter, while the elements of the other partition are configured to support the receivers at the different side of the transmitter. The numerical results show that in terms of minimizing the power consumption, the ES protocol and the TS protocol are preferable for the multicast and unicast methods, respectively. Similar conclusion was also verified by Niu *et al.* in [28], where the block coordinate descent (BCD) algorithm was employed for jointly designing the AP and IOS beamformer to maximize the weighted sum-rate. In [29], Zhang *et al.* solved a specific spectral efficiency maximization problem for the IOS-aided single-input and single-output (SISO) systems, where the closed-form expression and the branch-and-bound algorithm were employed for optimizing the continuous and discrete phase shift of the IOS elements. It was shown that the IOS-aided systems achieve considerably higher spectral efficiency than the state-of-the-art RIS-aided systems and the conventional cellular systems operating without IOS. In [30], the above system model was extended to the case of the AP associated with multi-antennas supporting multiple users, where the active AP beamformer and the passive IOS beamformer were optimized in an iterative manner for maximizing the sum-rate. In [31], Dhok *et al.* considered the IOS-aided systems communicating over spatially-correlated Rician channels, where both the outage probability and channel capacity were derived for the infinite block-length transmissions. Furthermore, the block error rate, the system throughput and the channel capacity were also formulated for realistic finite block-length transmissions. In [32], the sum coverage range maximization problem is formulated for IOS-aided systems, where the coefficients at the IOS elements and the resource allocation at the AP are jointly optimized for both a non-orthogonal multiple access (NOMA) scheme

and an orthogonal multiple access (OMA) scheme. It was verified by the simulation results that the IOS promises wider coverage range than conventional RIS. In [33], [34], an IOS was deployed in NOMA systems. It was shown that IOS-aided NOMA systems outperform the state-of-art RIS-aided systems both in terms of their sum-rate.

The IOS passive beamformer optimization reported in the above contributions are all based on the instantaneous CSI. To cut down the channel acquisition overhead, as well as the IOS phase shift update frequency, the authors of [35], [36] designed the IOS phase shift based on the statistical CSI, which changes slowly and may remain constant over numerous coherence time intervals. In [35], Papazafeiropoulos *et al.* employed the projected gradient ascent method for finding the optimal IOS phase shift based on the statistical CSI for maximizing the coverage probability. In [36], Wu *et al.* employed the two-timescale protocol for IOS-aided systems. Specifically, the passive IOS beamformer is designed based on the statistical CSI, while the active AP beamformer relies on the instantaneous CSI of the equivalent AP-UE channel. This reduces the channel estimation overhead, at the expense of a passive beamforming gain erosion, especially in rich-scattering channel environments.

However, the beamformer design of the IOS-aided systems in the above treatises has the following limitations. Firstly, perfect CSI knowledge is assumed for both the beamformer at the AP and for the IOS elements, which is impractical for finite-length pilot sequences. Furthermore, the transceiver hardware is assumed to be ideal, where the signals can be transmitted and received without distortion, which is unrealistic, especially when a large number of antennas are deployed at the AP. More specifically, in the high transmit power region, the dominant factors determining the system performance are the channel estimation error and hardware impairment (HWI). Hence, beamformers must be designed for IOS-aided systems by considering both the hardware impairment, as well as imperfect CSI acquisition. To deal with the above issues, Table I explicitly contrasts our contributions to the literature.

- Firstly, we formulate the linear minimum mean square error (LMMSE) estimator of the equivalent UE-AP channel for IOS-aided wireless systems, while considering the hardware impairments at both the UEs and the AP. Furthermore, the mean square error (MSE) of the LMMSE estimator is derived theoretically.
- Then, based on the estimated CSI of the equivalent channel, we employ a two-timescale protocol for jointly optimizing the active beamformer at the AP and the passive beamformer at the IOS for maximizing the ergodic spectral efficiency for transmission over random channel links. Specifically, the active beamformer at the AP is designed by the MMSE combination method based on the estimated instantaneous equivalent channel as well on the statistical channel estimation error covariance, the inter-user interference and the transceiver HWIs. Then, the ergodic spectral efficiency is formulated theoretically. Finally, the optimal closed-form IOS phase shift is derived for maximizing the upper bound of the ergodic spectral efficiency, based on the statistical CSI.

TABLE I: Contrasting our contributions to the literature [27]–[36].

	Our paper	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]
IOS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Multiple AP antennas	✔	✔	✔		✔					✔	
IOS beamforming based on statistical CSI	✔									✔	✔
Channel estimation	✔										
Imperfect CSI	✔										
Transceiver hardware impairments	✔										

- The theoretical analysis and the numerical results show that the transceiver HWIs constrain the ergodic spectral efficiency improvement in the high transmit power region. We also show that employing more AP antennas is capable of compensating the HWI at the AP. By contrast, the HWI at the UEs cannot be compensated by harnessing more AP antennas.

C. Organization and Notation

The rest of this paper is organized as follows. In Section II, we present the system model, while channel estimation is described in Section III. Section IV presents our active AP beamformer design and the passive IOS phase shift optimization. Our simulation results are presented in Section V, while we conclude in Section VI.

Notations: $\jmath = \sqrt{-1}$, vectors and matrices are denoted by boldface lower and upper case letters, respectively, $(\cdot)^{\dagger}$ represents the conjugate operation, $\mathbb{C}^{m \times n}$ denotes the space of $m \times n$ complex-valued matrices, a_n represents the nth element in vector $\mathbf{a}$, $\mathbf{0}_N$ is the $N \times 1$ zero vector, $\mathbf{I}_N$ and $\mathbf{O}_{N_1, N_2}$ represents the $N \times N$ identity matrix and $N_1 \times N_2$ zero matrix, respectively, $\mathbf{Diag}\{a_1, a_2, \cdots, a_N\}$ denotes a diagonal matrix having the elements of $a_1, a_2, \cdots, a_N$ in order, $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a circularly symmetric complex Gaussian random vector with the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, $\mathbf{A} \succ \mathbf{0}$ indicates that $\mathbf{A}$ is a positive definite matrix, $\mathbf{A} \succeq \mathbf{0}$ represents $\mathbf{A}$ is a positive semi-definite matrix, $\mathbb{E}[\mathbf{x}]$ represents the mean of the random vector $\mathbf{x}$, the auto-covariance matrix of the random vector $\mathbf{x}$ is denoted as $\mathbf{C_{xx}} = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^{\mathrm{H}}\right]$, the cross-covariance matrix between the random vectors $\mathbf{x}$ and $\mathbf{y}$ is expressed as $\mathbf{C_{xy}} = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^{\mathrm{H}}\right]$, the auto-correlation matrix of the random vector $\mathbf{x}$ is denoted as $\mathbb{E}\left[\mathbf{xx}^{\mathrm{H}}\right]$, the cross-correlation matrix between the random vectors $\mathbf{x}$ and $\mathbf{y}$ is given by $\mathbb{E}\left[\mathbf{xy}^{\mathrm{H}}\right]$, $[\mathbf{a}]_n$ represents the nth element in the vector $\mathbf{a}$ and $[\mathbf{A}]_{n_1, n_2}$ represents the (n_1, n_2)th element in matrix $\mathbf{A}$.

II. System Model

The IOS-aided system model is shown in Fig. 1, including an AP equipped with M uniform linear array (ULA) antennas having an element spacing of d_0, an IOS with N elements, and randomly positioned users. As shown in Fig. 1, we denote
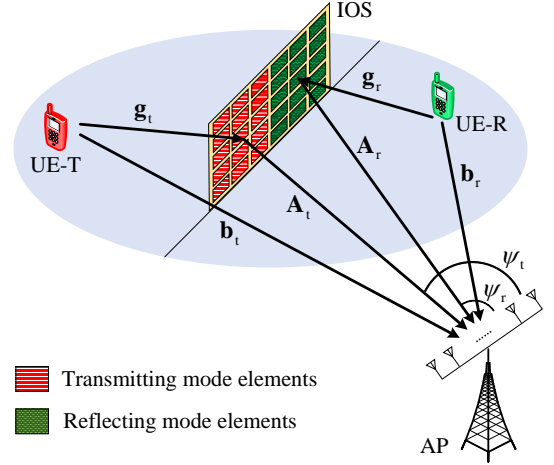

Fig. 1: System model of the considered IOS-aided wireless communication system, including an AP equipped with M ULA antennas, an IOS employing the mode stitching protocol with a total of N elements, a user located at the same side of the IOS, denoted as UE-R, and a user located at the other side of the IOS, denoted as UE-T.

the user located at the same side of the IOS by UE-R, and that at the other side of the IOS by UE-T[1].

A. IOS Architecture

In the ES protocol, additional circuits are required to split the signal energy and some power leakage is inevitable. By contrast, in the TS protocol frequent switching is required between the reflecting and transmitting operation on each element, resulting in high hardware complexity. Fortunately, the MS protocol is easy to implement [38]. Thus, in this paper we focus on the IOS relying on the MS protocol. The part of the IOS configured for reflecting signals is a uniform rectangular planar array (URPA) containing $N_{\mathrm{r}} = N_{\mathrm{r}}^x \times N_{\mathrm{r}}^y$ elements, and that used for transmitting signals is a URPA containing $N_{\mathrm{t}} = N_{\mathrm{t}}^x \times N_{\mathrm{t}}^y$ elements, where N_{r}^x and N_{r}^y represent the numbers of reflecting elements in the horizontal

---

[1]In this paper, we aim to characterize the fundamental ergodic spectral efficiency limits of an IOS aided dual-user communication network. However, it can be extended to a multi-user system by partitioning the IOS into multiple subsurfaces [37]. Specifically, each subsurface is allocated to serve a specific user. Note that in this IOS partitioning method, we need to consider not only the interference between users but also the interference between different subsurfaces. This is a challenge set aside for our future work.

and vertical direction respectively, while N_t^x and N_t^y represent the numbers of transmitting elements in the horizontal and vertical direction respectively. Furthermore, the spacing between adjacent IOS elements in the horizontal and vertical directions are represented by δ_x and δ_y, respectively. Additionally, $\boldsymbol{\Theta}_r$ and $\boldsymbol{\Theta}_t$ denote the phase shift matrices of the reflective mode elements and transmit mode elements, given by

$$\boldsymbol{\Theta}_i = \mathrm{diag}\left\{e^{j\theta_{i,1}}, e^{j\theta_{i,2}}, \cdots, e^{j\theta_{i,N_i}}\right\}, \quad i \in \{r, t\}, \quad (1)$$

where $\theta_{r,n}$ represents the phase shift of the nth reflective mode element and $\theta_{t,n}$ is the phase shift of the nth transmit mode element.

B. Channel Model

Here we consider an uplink scenario, but emphasize that our performance analysis method proposed in this paper is also applicable to a downlink scenario. As shown in Fig. 1, we denote the channel spanning from the UE-R to the AP as $\mathbf{b}_r \in \mathbb{C}^{M \times 1}$, from the UE-T to the AP as $\mathbf{b}_t \in \mathbb{C}^{M \times 1}$, from the UE-R to the reflective mode IOS elements as $\mathbf{g}_r \in \mathbb{C}^{N_r \times 1}$, from the UE-T to the transmit mode IOS elements as $\mathbf{g}_t \in \mathbb{C}^{N_t \times 1}$, and from the IOS to the AP as $\mathbf{A} = [\mathbf{A}_r, \mathbf{A}_t] \in \mathbb{C}^{M \times N}$, where $\mathbf{A}_r \in \mathbb{C}^{M \times N_r}$ and $\mathbf{A}_t^{M \times N_t}$ represent the channel spanning from the reflective mode IOS elements and transmit mode IOS elements to the AP, respectively. Furthermore, $\varrho_{\mathbf{b}_r} = C_0 d_{\mathbf{b}_r}^{-\alpha_{\mathbf{b}_r}}$, $\varrho_{\mathbf{b}_t} = C_0 d_{\mathbf{b}_t}^{-\alpha_{\mathbf{b}_t}}$, $\varrho_{\mathbf{g}_r} = C_0 d_{\mathbf{g}_r}^{-\alpha_{\mathbf{g}_r}}$, $\varrho_{\mathbf{g}_t} = C_0 d_{\mathbf{g}_t}^{-\alpha_{\mathbf{g}_t}}$, $\varrho_{\mathbf{A}_r} = C_0 d_{\mathbf{A}_r}^{-\alpha_{\mathbf{A}_r}}$ and $\varrho_{\mathbf{A}_t} = C_0 d_{\mathbf{A}_t}^{-\alpha_{\mathbf{A}_t}}$ represent the path loss in the corresponding links, where C_0 is the path loss at the reference distance of 1 meter, $d_{\mathbf{b}_r}$, $d_{\mathbf{b}_t}$, $d_{\mathbf{g}_r}$, $d_{\mathbf{g}_t}$, $d_{\mathbf{A}_r}$ and $d_{\mathbf{A}_t}$ denote the distance in the corresponding links, and $\alpha_{\mathbf{b}_r}$, $\alpha_{\mathbf{b}_t}$, $\alpha_{\mathbf{g}_r}$, $\alpha_{\mathbf{g}_t}$, $\alpha_{\mathbf{A}_r}$ and $\alpha_{\mathbf{A}_t}$ represent the path loss exponent in the corresponding links.

In this paper, we consider the far-field channel model. However, our performance analysis approach is also valid, when the users are located in the near-field region of the IOS. Since the IOS is harnessed for creating additional paths for signal transmission among the AP and users, it is reasonable to assume that $\mathbf{g}_r$, $\mathbf{g}_t$, $\mathbf{A}_r$ and $\mathbf{A}_t$ experience Rician fading [39], while $\mathbf{b}_r$ and $\mathbf{b}_t$ face Rayleigh fading [40]. Thus, in the AP-UE links, we have $\mathbf{b}_r \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ and $\mathbf{b}_t \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$. In the IOS-UE links, $\mathbf{g}_r$ and $\mathbf{g}_t$ are

$$\mathbf{g}_i = \sqrt{\frac{\kappa_{\mathbf{g}_i}}{1 + \kappa_{\mathbf{g}_i}}}\overline{\mathbf{g}}_i + \sqrt{\frac{1}{1 + \kappa_{\mathbf{g}_i}}}\widetilde{\mathbf{g}}_i, \quad i \in \{r, t\}, \quad (2)$$

where $\kappa_{\mathbf{g}_r}$ and $\kappa_{\mathbf{g}_t}$ are the Rician factor of $\mathbf{g}_r$ and $\mathbf{g}_t$ respectively. The NLoS component obeys $\widetilde{\mathbf{g}}_r \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ and $\widetilde{\mathbf{g}}_t \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$. The LoS components $\overline{\mathbf{g}}_r$ and $\overline{\mathbf{g}}_t$ are formulated as [41]

$$\overline{\mathbf{g}}_i = \left[1, \cdots, e^{-j\frac{2\pi}{\lambda}(\delta_x n_x \sin\phi_i \cos\varphi_i + \delta_y n_y \cos\phi_i)}, \cdots, \right.$$
$$\left. e^{-j\frac{2\pi}{\lambda}(\delta_x(N_i^x-1)\sin\phi_i\cos\varphi_i + \delta_y(N_i^y-1)\cos\phi_i)}\right]^H,$$
$$i \in \{r, t\}, \quad (3)$$

where λ is the wavelength, ϕ_r and φ_r are the elevation and azimuth AoA with respect to the reflective IOS elements respectively, while ϕ_t and φ_t are the elevation and azimuth

AoA with respect to the transmit IOS elements respectively. In the AP-IOS link, we have

$$\mathbf{A}_i = \sqrt{\frac{\kappa_{\mathbf{A}_i}}{1 + \kappa_{\mathbf{A}_i}}}\overline{\mathbf{A}}_i + \sqrt{\frac{1}{1 + \kappa_{\mathbf{A}_i}}}\widetilde{\mathbf{A}}_i, \ i \in \{r, t\}, \quad (4)$$

where $\kappa_{\mathbf{A}_i}$ is the Rician factor of $\mathbf{A}_i$. The NLoS component $\widetilde{\mathbf{A}}_i = [\widetilde{\mathbf{a}}_{i,1}^H, \widetilde{\mathbf{a}}_{i,2}^H, \cdots, \widetilde{\mathbf{a}}_{i,M}^H]^H$, associated with $\widetilde{\mathbf{a}}_{i,m} \sim \mathcal{CN}(\mathbf{0}_N, \mathbf{I}_N)$ corresponds to the mth AP antenna. The LoS component $\overline{\mathbf{A}}_i$ obeys:

$$\overline{\mathbf{A}}_i = \overline{\mathbf{a}}_i^{(AP)}\overline{\mathbf{a}}_i^{(IOS)H}. \quad (5)$$

In (5), $\overline{\mathbf{a}}_i^{(IOS)}$ is given by [41]

$$\overline{\mathbf{a}}_i^{(IOS)} = \left[1, \cdots, e^{-j\frac{2\pi}{\lambda}(\delta_x n_x \sin\omega_i \cos\varpi_i + \delta_y n_y \cos\omega_i)}, \cdots, \right.$$
$$\left. e^{-j\frac{2\pi}{\lambda}(\delta_x(N_i^x-1)\sin\omega_i\cos\varpi_i + \delta_y(N_i^y-1)\cos\omega_i)}\right]^H, \quad (6)$$

where ω_i and ϖ_i are the elevation and the azimuth AoD with respect to the IOS elements, respectively. Furthermore, in (5), $\overline{\mathbf{a}}_i^{(AP)} = \left[1, e^{-j\frac{2\pi}{\lambda}d_0\cos\psi_i}, \cdots, e^{-j\frac{2\pi}{\lambda}d_0(M-1)\cos\psi_i}\right]^H$ [41], where ψ_i is the AoA of the signals spanning from the IOS elements to the AP antennas.

Therefore, the equivalent channel emerging from the UE-R to the AP is given by $\mathbf{h}_r = \overline{\mathbf{h}}_r + \widetilde{\mathbf{h}}_r$, where $\overline{\mathbf{h}}_r$ is the known component and $\widetilde{\mathbf{h}}_r$ is the uncertainty. Similarly, the equivalent channel arriving from the UE-T to the AP is given by $\mathbf{h}_t = \overline{\mathbf{h}}_t + \widetilde{\mathbf{h}}_t$, where $\overline{\mathbf{h}}_t$ is the known component and $\widetilde{\mathbf{h}}_t$ is the uncertainty. Hence, we have

$$\mathbf{h}_i = \sqrt{\varrho_{\mathbf{A}_i}\varrho_{\mathbf{g}_i}}\mathbf{A}_i\boldsymbol{\Theta}_i\mathbf{g}_i + \sqrt{\varrho_{\mathbf{b}_i}}\mathbf{b}_i, \quad (7)$$

$$\overline{\mathbf{h}}_i = \sqrt{\varrho_{\mathbf{A}_i}\varrho_{\mathbf{g}_i}}\sqrt{\frac{\kappa_{\mathbf{A}_i}\kappa_{\mathbf{g}_i}}{(1+\kappa_{\mathbf{A}_i})(1+\kappa_{\mathbf{g}_i})}}\overline{\mathbf{A}}_i\boldsymbol{\Theta}_i\overline{\mathbf{g}}_i, \quad (8)$$

$$\widetilde{\mathbf{h}}_i = \sqrt{\frac{\varrho_{\mathbf{A}_i}\varrho_{\mathbf{g}_i}}{(1+\kappa_{\mathbf{A}_i})(1+\kappa_{\mathbf{g}_i})}}\left(\sqrt{\kappa_{\mathbf{A}_i}}\overline{\mathbf{A}}_i\boldsymbol{\Theta}_i\widetilde{\mathbf{g}}_i + \sqrt{\kappa_{\mathbf{g}_i}}\widetilde{\mathbf{A}}_i\boldsymbol{\Theta}_i\overline{\mathbf{g}}_i \right.$$
$$\left. + \widetilde{\mathbf{A}}_i\boldsymbol{\Theta}_i\widetilde{\mathbf{g}}_i\right) + \sqrt{\varrho_{\mathbf{b}_i}}\mathbf{b}_i, \quad (9)$$

where $i \in \{r, t\}$. Upon exploiting (2) and (4), we can formulate the mean and auto-covariance matrix of $\mathbf{h}_r$ and $\mathbf{h}_t$ as

$$\mathbb{E}[\mathbf{h}_i] = \overline{\mathbf{h}}_i = \sqrt{\varrho_{\mathbf{A}_i}\varrho_{\mathbf{g}_i}}\sqrt{\frac{\kappa_{\mathbf{A}_i}\kappa_{\mathbf{g}_i}}{(1+\kappa_{\mathbf{A}_i})(1+\kappa_{\mathbf{g}_i})}}\overline{\mathbf{A}}_i\boldsymbol{\Theta}_i\overline{\mathbf{g}}_i, \quad (10)$$

$$\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i} = \mathbf{C}_{\widetilde{\mathbf{h}}_i\widetilde{\mathbf{h}}_i} = \varrho_{\mathbf{A}_i}\varrho_{\mathbf{g}_i}N_i\frac{(1+\kappa_{\mathbf{g}_i})\mathbf{I}_M + \kappa_{\mathbf{A}_i}\overline{\mathbf{a}}_i^{(AP)}\overline{\mathbf{a}}_i^{(AP)H}}{(1+\kappa_{\mathbf{A}_i})(1+\kappa_{\mathbf{g}_i})}$$
$$+ \varrho_{\mathbf{b}_i}\mathbf{I}_M, \quad (11)$$

where $i \in \{r, t\}$.

C. Channel Estimation and Beamforming Protocol

The two-timescale channel estimation and beamforming protocol conceived is illustrated in Fig. 2. In each statistical block, the statistical CSI, e.g. $\overline{\mathbf{g}}_r$, $\overline{\mathbf{g}}_t$, $\overline{\mathbf{a}}_r^{(AP)}$, $\overline{\mathbf{a}}_t^{(AP)}$, $\overline{\mathbf{a}}_r^{(IOS)H}$ and $\overline{\mathbf{a}}_t^{(IOS)H}$, is fixed since they change slowly [5]. The statistical CSI is estimated at the beginning of each statistical
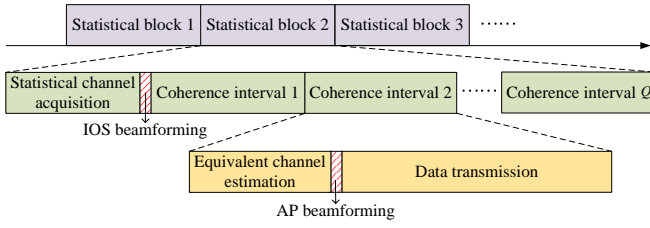
Fig. 2: Illustration of the employed two-timescale channel estimation and beamforming protocol.

block, and the remaining part of each statistical block includes Q coherence intervals. Since the statistical CSI changes slowly, the pilot overhead of the statistical CSI acquisition can be omitted [5]. Each coherence interval is composed of several symbol slots, which are assumed to have the same instantaneous CSI. The IOS beamforming is designed based on the statistical CSI. Thus, the IOS phase shift is fixed within each statistical block, and the equivalent UE-AP channel is estimated at the beginning of each coherence interval. Then, the BS beamformer is designed based on the instantaneous equivalent UE-AP channel, followed by data transmission.

In the following, we formulate the LMMSE estimation of the equivalent channels $\mathbf{h}_r$ and $\mathbf{h}_t$ defined in (7), considering the signal distortions resulting from HWIs at both the APs and UEs.

III. CHANNEL ESTIMATION

In this section, we formulate the LMMSE estimation of the equivalent channel. For making efficient use of the large number of AP antennas, an uplink channel estimation scenario is considered. To eliminate the pilot contamination between the UE-R and UE-T, a pair of orthogonal pilot sequences are employed, denoted as $\left[\tau_r^{(1)}; \tau_r^{(2)}; \cdots; \tau_r^{(K)}\right]$ and $\left[\tau_t^{(1)}; \tau_t^{(2)}; \cdots; \tau_t^{(K)}\right]$, respectively. In previous IOS papers, the transceiver hardware is assumed to be ideal [27]–[36]. However, practical circuits of the transceivers generally suffer from hardware impairments, including power amplifier non-linearities, amplitude/phase imbalance in the In-phase/Quadrature mixers, phase error in the local oscillator, sampling jitter and finite-resolution quantization in the analog-to-digital converters [42]. To characterize the impact of transceiver hardware impairments, the non-ideal hardware circuits at the transmitter and the receiver can be modelled as non-linear memoryless filters [43]. Specifically, the key modeling characteristics in this non-linear memoryless filter are that the desired signal is scaled by a deterministic factor and that an uncorrelated memoryless signal distortion term is added, which follows the Gaussian distribution in the worst case. In the kth $(k = 1, 2, \cdots, K)$ symbol slot, the pilot $\tau_r^{(k)}$ at the UE-R and the pilot $\tau_t^{(k)}$ at the UE-T are transmitted simultaneously, leading to the received signal at the AP as [42]

$$\mathbf{x}^{(k)} = \underbrace{\sqrt{\rho_r \varepsilon_{\mathbf{v}} \varepsilon_{u,r}} \mathbf{h}_r \tau_r^{(k)}}_{\text{Desired UE-R pilot}} + \underbrace{\sqrt{\rho_t \varepsilon_{\mathbf{v}} \varepsilon_{u,t}} \mathbf{h}_t \tau_t^{(k)}}_{\text{Desired UE-T pilot}}$$

$$+ \underbrace{\sqrt{\rho_r \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,r})} \mathbf{h}_r u_r^{(k)}}_{\text{UE-R HWI distortion}} + \underbrace{\sqrt{\rho_t \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,t})} \mathbf{h}_t u_t^{(k)}}_{\text{UE-T HWI distortion}}$$

$$+ \underbrace{\sqrt{\rho_r (1 - \varepsilon_{\mathbf{v}})} \mathbf{h}_r \odot \mathbf{v}_r^{(k)} + \sqrt{\rho_t (1 - \varepsilon_{\mathbf{v}})} \mathbf{h}_t \odot \mathbf{v}_t^{(k)}}_{\text{AP HWI distortion}}$$

$$+ \underbrace{\mathbf{w}^{(k)}}_{\text{Additive noise}}, \tag{12}$$

where $\varepsilon_{\mathbf{v}}$, $\varepsilon_{u,r}$ and $\varepsilon_{u,t}$ represent the hardware quality factor of the AP, UE-R and UE-T, respectively, with $0 \leq \varepsilon_{\mathbf{v}} \leq 1$, $0 \leq \varepsilon_{u,r} \leq 1$ and $0 \leq \varepsilon_{u,t} \leq 1$. The hardware quality factor of 1 means that the hardware is ideal, while 0 means that the hardware is completely inadequate. Furthermore, ρ_r and ρ_t are the power of transmitted symbols at the UE-R and UE-T, respectively. We have $u_r^{(k)} \sim \mathcal{CN}(0, 1)$, $u_t^{(k)} \sim \mathcal{CN}(0, 1)$, $\mathbf{v}_r^{(k)} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$, $\mathbf{v}_t^{(k)} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$, and $\mathbf{w}^{(k)} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_w^2 \mathbf{I}_M)$.

Firstly, to estimate $\mathbf{h}_r$, the conjugate transpose of the UE-R pilot sequence, i.e. $\left[\tau_r^{(1)\dagger}, \tau_r^{(2)\dagger}, \cdots, \tau_r^{(K)\dagger}\right]$, is employed to combine the AP observations $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(K)}$ and we have the processed received observation formulated as

$$\mathbf{x}_r = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \mathbf{x}^{(k)} \tau_r^{(k)\dagger}$$

$$\overset{(a)}{=} \sqrt{K \rho_r \varepsilon_{\mathbf{v}} \varepsilon_{u,r}} \mathbf{h}_r + \sqrt{\rho_r \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,r})} \mathbf{h}_r u_r$$

$$+ \sqrt{\rho_t \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,t})} \mathbf{h}_t u_t + \sqrt{\rho_r (1 - \varepsilon_{\mathbf{v}})} \mathbf{h}_r \odot \mathbf{v}_r$$

$$+ \sqrt{\rho_t (1 - \varepsilon_{\mathbf{v}})} \mathbf{h}_t \odot \mathbf{v}_t + \mathbf{w}, \tag{13}$$

where $u_r = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} u_r^{(k)} \tau_r^{(k)\dagger}$, $u_t = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} u_t^{(k)} \tau_r^{(k)\dagger}$, $\mathbf{v}_r = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \mathbf{v}_r^{(k)} \tau_r^{(k)\dagger}$, $\mathbf{v}_t = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \mathbf{v}_t^{(k)} \tau_r^{(k)\dagger}$, $\mathbf{w} = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \mathbf{w}^{(k)} \tau_r^{(k)\dagger}$, In (13), (a) is based on the orthogonality of the pilot sequence, i.e. $\sum_{k=1}^{K} \tau_r^{(k)} \tau_r^{(k)\dagger} = \sum_{k=1}^{K} \tau_t^{(k)} \tau_t^{(k)\dagger} = K$ and $\sum_{k=1}^{K} \tau_r^{(k)} \tau_t^{(k)\dagger} = \sum_{k=1}^{K} \tau_t^{(k)} \tau_r^{(k)\dagger} = 0$. Due to the independence of $u_r^{(k)}$, $u_t^{(k)}$, $\mathbf{v}_r^{(k)}$, $\mathbf{v}_t^{(k)}$ and $\mathbf{w}^{(k)}$ for $k = 1, 2, \cdots, K$, we have $u_r \sim \mathcal{CN}(0, 1)$, $u_t \sim \mathcal{CN}(0, 1)$, $\mathbf{v}_r \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$, $\mathbf{v}_t \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}_M, \sigma_w^2 \mathbf{I}_M)$.

Based on (13), we have

$$\mathbb{E}[\mathbf{x}_r] = \sqrt{K \rho_r \varepsilon_{\mathbf{v}} \varepsilon_{u,r}} \overline{\mathbf{h}}_r, \tag{14}$$

$$\mathbf{C}_{\mathbf{h}_r \mathbf{x}_r} = \sqrt{K \rho_r \varepsilon_{\mathbf{v}} \varepsilon_{u,r}} \mathbf{C}_{\mathbf{h}_r \mathbf{h}_r} \tag{15}$$

and

$$\mathbf{C}_{\mathbf{x}_r \mathbf{x}_r} = \rho_r \varepsilon_{\mathbf{v}} (1 + (K - 1) \varepsilon_{u,r}) \mathbf{C}_{\mathbf{h}_r \mathbf{h}_r} + \rho_t \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,t}) \cdot$$
$$\mathbf{C}_{\mathbf{h}_t \mathbf{h}_t} + \rho_r (1 - \varepsilon_{\mathbf{v}}) \mathbf{C}_{\mathbf{h}_r \mathbf{h}_r} \odot \mathbf{I}_M + \rho_t (1 - \varepsilon_{\mathbf{v}}) \mathbf{C}_{\mathbf{h}_t \mathbf{h}_t} \odot \mathbf{I}_M$$
$$+ \rho_r \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,r}) \overline{\mathbf{h}}_r \overline{\mathbf{h}}_r^H + \rho_t \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,t}) \overline{\mathbf{h}}_t \overline{\mathbf{h}}_t^H$$
$$+ \left(\rho_r (1 - \varepsilon_{\mathbf{v}}) \varrho_{\mathbf{A}_r} \varrho_{\mathbf{g}_r} \frac{\kappa_{\mathbf{A}_r} \kappa_{\mathbf{g}_r} \left\| \overline{\mathbf{a}}_r^{(\text{IOS})H} \boldsymbol{\Theta}_r \overline{\mathbf{g}}_r \right\|^2}{(1 + \kappa_{\mathbf{A}_r})(1 + \kappa_{\mathbf{g}_r})} \right.$$

$$+\rho_{\mathrm{t}}\left(1-\varepsilon_{\mathbf{v}}\right)\varrho_{\mathbf{A}_{\mathrm{t}}}\varrho_{\mathbf{g}_{\mathrm{t}}}\frac{\kappa_{\mathbf{A}_{\mathrm{t}}}\kappa_{\mathbf{g}_{\mathrm{t}}}\|\overline{\mathbf{a}}_{\mathrm{t}}^{(\mathrm{IOS})\mathrm{H}}\boldsymbol{\Theta}_{\mathrm{t}}\overline{\mathbf{g}}_{\mathrm{t}}\|^{2}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}\right)}+\sigma_{w}^{2}\Bigg)\mathbf{I}_{M}. \tag{16}$$

Therefore, the LMMSE estimator of $\mathbf{h}_{\mathrm{r}}$ is given by

$$\hat{\mathbf{h}}_{\mathrm{r}}=\overline{\mathbf{h}}_{\mathrm{r}}+\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}\mathbf{C}_{\mathbf{x}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}^{-1}\left(\mathbf{x}_{\mathrm{r}}-\mathbb{E}[\mathbf{x}_{\mathrm{r}}]\right). \tag{17}$$

Hence, the estimation error is $\check{\mathbf{h}}_{\mathrm{r}}=\mathbf{h}_{\mathrm{r}}-\hat{\mathbf{h}}_{\mathrm{r}}$, and the estimation covariance matrix and estimation error covariance matrix are

$$\mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}}=\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}\mathbf{C}_{\mathbf{x}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}^{\mathrm{H}}, \tag{18}$$

$$\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{r}}\check{\mathbf{h}}_{\mathrm{r}}}=\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}}-\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}\mathbf{C}_{\mathbf{x}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}^{\mathrm{H}}. \tag{19}$$

Similarly, we can formulate the LMMSE estimator of $\mathbf{h}_{\mathrm{t}}$ and its corresponding estimation covariance matrix and estimation error covariance matrix as:

$$\hat{\mathbf{h}}_{\mathrm{t}}=\overline{\mathbf{h}}_{\mathrm{t}}+\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}\mathbf{C}_{\mathbf{x}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}^{-1}\left(\mathbf{x}_{\mathrm{t}}-\mathbb{E}[\mathbf{x}_{\mathrm{t}}]\right), \tag{20}$$

$$\mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{t}}\hat{\mathbf{h}}_{\mathrm{t}}}=\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}\mathbf{C}_{\mathbf{x}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}^{\mathrm{H}}, \tag{21}$$

$$\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{t}}\check{\mathbf{h}}_{\mathrm{t}}}=\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{h}_{\mathrm{t}}}-\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}\mathbf{C}_{\mathbf{x}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}^{\mathrm{H}}, \tag{22}$$

where $\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}$ and $\mathbf{C}_{\mathbf{x}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}$ can be evaluated similarly to the case, when $\mathbf{h}_{\mathrm{r}}$ is estimated. According to (19) and (22), the normalized mean square error (N-MSE) is

$$\epsilon=\frac{1}{2}\left(\frac{\mathrm{Tr}\left[\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}}-\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}\mathbf{C}_{\mathbf{x}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{x}_{\mathrm{r}}}^{\mathrm{H}}\right]}{\mathrm{Tr}\left[\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}}\right]}\right.$$
$$\left.+\frac{\mathrm{Tr}\left[\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{h}_{\mathrm{t}}}-\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}\mathbf{C}_{\mathbf{x}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{x}_{\mathrm{t}}}^{\mathrm{H}}\right]}{\mathrm{Tr}\left[\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{h}_{\mathrm{t}}}\right]}\right). \tag{23}$$

IV. BEAMFORMING DESIGN

In the section, we present the AP beamforming based on the estimated equivalent channels $\hat{\mathbf{h}}_{\mathrm{r}}$ and $\hat{\mathbf{h}}_{\mathrm{t}}$ defined in (17) and (20), respectively. Then, the ergodic spectral efficiency upper bound is theoretically derived with respect to the IOS design. Finally, we optimize the beamforming design of the IOS by maximizing the ergodic spectral efficiency upper bound.

A. AP Beamforming

We denote the information symbol at the UE-R and UE-T as $s_{\mathrm{r}}\sim\mathcal{CN}\left(0,1\right)$ and $s_{\mathrm{t}}\sim\mathcal{CN}\left(0,1\right)$, respectively. Then, the signal received at the AP is given by

$$\mathbf{y}=\underbrace{\sqrt{\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}}\hat{\mathbf{h}}_{\mathrm{r}}s_{\mathrm{r}}}_{s_{\mathrm{r}}\text{ over estimated channel}}+\underbrace{\sqrt{\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{t}}}\hat{\mathbf{h}}_{\mathrm{t}}s_{\mathrm{t}}}_{s_{\mathrm{t}}\text{ over estimated channel}}$$
$$+\underbrace{\sqrt{\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}}\check{\mathbf{h}}_{\mathrm{r}}s_{\mathrm{r}}}_{s_{\mathrm{r}}\text{ over unknown channel}}+\underbrace{\sqrt{\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{t}}}\check{\mathbf{h}}_{\mathrm{t}}s_{\mathrm{t}}}_{s_{\mathrm{t}}\text{ over unknown channel}}$$
$$+\underbrace{\sqrt{\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{r}}\right)}\mathbf{h}_{\mathrm{r}}u_{\mathrm{r}}}_{\text{UE-R HWI distortion}}+\underbrace{\sqrt{\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{t}}\right)}\mathbf{h}_{\mathrm{t}}u_{\mathrm{t}}}_{\text{UE-T HWI distortion}}$$
$$+\underbrace{\sqrt{\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)}\mathbf{h}_{\mathrm{r}}\odot\mathbf{v}_{\mathrm{r}}+\sqrt{\rho_{\mathrm{t}}\left(1-\varepsilon_{\mathbf{v}}\right)}\mathbf{h}_{\mathrm{t}}\odot\mathbf{v}_{\mathrm{t}}}_{\text{AP HWI distortion}}$$
$$+\underbrace{\mathbf{w}}_{\text{Additive noise}}, \tag{24}$$

where $u_{\mathrm{r}}\sim\mathcal{CN}\left(0,1\right)$, $u_{\mathrm{t}}\sim\mathcal{CN}\left(0,1\right)$, $\mathbf{v}_{\mathrm{r}}\sim\mathcal{CN}(\mathbf{0}_{M},\mathbf{I}_{M})$, $\mathbf{v}_{\mathrm{t}}\sim\mathcal{CN}\left(\mathbf{0}_{M},\mathbf{I}_{M}\right)$, and $\mathbf{w}\sim\mathcal{CN}\left(\mathbf{0}_{M},\sigma_{w}^{2}\mathbf{I}_{M}\right)$.

The MMSE combiner is employed at the AP to recover the information s_{r} and s_{t} based on the observation $\mathbf{y}$.

***Theorem* 1.** The MMSE combining vectors $\mathbf{q}_{\mathrm{r}}$ and $\mathbf{q}_{\mathrm{t}}$ of the UE-R and UE-T, respectively, are given by [42]

$$\mathbf{q}_{\mathrm{r}}=\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}\left(\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}+\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{t}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}+\mathbf{R}\right)^{-1}\hat{\mathbf{h}}_{\mathrm{r}}, \tag{25}$$

$$\mathbf{q}_{\mathrm{t}}=\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{t}}\left(\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}+\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{t}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}+\mathbf{R}\right)^{-1}\hat{\mathbf{h}}_{\mathrm{t}}, \tag{26}$$

where we have:

$$\mathbf{R}=\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}(1-\varepsilon_{u,\mathrm{r}})\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{r}}\check{\mathbf{h}}_{\mathrm{r}}}+\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{t}}\right)\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{t}}\check{\mathbf{h}}_{\mathrm{t}}}$$
$$+\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\frac{\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)N_{\mathrm{r}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)}+\varrho_{\mathbf{b}_{\mathrm{r}}}\right.$$
$$\left.+\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\frac{\kappa_{\mathbf{A}_{\mathrm{r}}}\kappa_{\mathbf{g}_{\mathrm{r}}}\|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\boldsymbol{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\|^{2}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)}\right)\mathbf{I}_{M}$$
$$+\rho_{\mathrm{t}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\varrho_{\mathbf{A}_{\mathrm{t}}}\varrho_{\mathbf{g}_{\mathrm{t}}}\frac{\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)N_{\mathrm{t}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}\right)}+\varrho_{\mathbf{b}_{\mathrm{t}}}\right.$$
$$\left.+\varrho_{\mathbf{A}_{\mathrm{t}}}\varrho_{\mathbf{g}_{\mathrm{t}}}\frac{\kappa_{\mathbf{A}_{\mathrm{t}}}\kappa_{\mathbf{g}_{\mathrm{t}}}\|\overline{\mathbf{a}}_{\mathrm{t}}^{(\mathrm{IOS})\mathrm{H}}\boldsymbol{\Theta}_{\mathrm{t}}\overline{\mathbf{g}}_{\mathrm{t}}\|^{2}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}\right)}\right)\mathbf{I}_{M}+\sigma_{w}^{2}\mathbf{I}_{M}. \tag{27}$$

Based on the MMSE combiner at the AP, the instantaneous SINR for the received symbols of the UE-R and UE-T are

$$\gamma_{\mathrm{r}}=\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}\left(\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{r}}\right)\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}+\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{t}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}\right.$$
$$\left.+\mathbf{R}\right)^{-1}\hat{\mathbf{h}}_{\mathrm{r}}, \tag{28}$$

$$\gamma_{\mathrm{t}}=\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{t}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}\left(\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{t}}\right)\hat{\mathbf{h}}_{\mathrm{t}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}+\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}\right.$$
$$\left.+\mathbf{R}\right)^{-1}\hat{\mathbf{h}}_{\mathrm{t}}. \tag{29}$$

Proof: See Appendix A. ∎

***Theorem* 2.** Upon using the MMSE combiner at the AP, the instantaneous spectral efficiency of the UE-R and UE-T, denoted as $\mathcal{R}_{\mathrm{ins,r}}$ and $\mathcal{R}_{\mathrm{ins,t}}$, respectively, becomes

$$\mathcal{R}_{\mathrm{ins,i}}=\log_{2}\left(1+\frac{\rho_{\mathrm{i}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{i}}\zeta_{\mathrm{i}}}{1+\rho_{\mathrm{i}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{i}}\right)\zeta_{\mathrm{i}}}\right),\;\mathrm{i}\in\{\mathrm{r},\mathrm{t}\}, \tag{30}$$

where $\zeta_{\mathrm{r}}=\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{r}}-\frac{\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{t}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{r}}}{1+\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{t}}}$ and $\zeta_{\mathrm{t}}=\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{t}}-\frac{\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{t}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{t}}}{1+\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_{\mathrm{r}}^{\mathrm{H}}\mathbf{R}^{-1}\hat{\mathbf{h}}_{\mathrm{r}}}$.

Proof: See Appendix B. ∎

Therefore, according to *Theorem* 2, we can express the ergodic spectral efficiency of UE-R and UE-T as $\mathcal{R}_{\mathrm{erg,i}}=\mathbb{E}\left[\mathcal{R}_{\mathrm{ins,i}}\right]=\mathbb{E}\left[\log_{2}\left(1+\frac{\rho_{\mathrm{i}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{i}}\zeta_{\mathrm{i}}}{1+\rho_{\mathrm{i}}\varepsilon_{\mathbf{v}}(1-\varepsilon_{u,\mathrm{i}})\zeta_{\mathrm{i}}}\right)\right]$, with $\mathrm{i}\in\{\mathrm{r},\mathrm{t}\}$.

***Theorem* 3.** When $\cos\psi_{\mathrm{t}}\neq\cos\psi_{\mathrm{r}}$, the ergodic spectral efficiency upper bound of UE-R and UE-T for transmission over random channel links, denoted as $\ddot{\mathcal{R}}_{\mathrm{erg,r}}$ and $\ddot{\mathcal{R}}_{\mathrm{erg,t}}$,

respectively, can be written as

$$\ddot{\mathcal{R}}_{\text{erg,i}} = \log_2 \left(1 + \frac{\rho_i \varepsilon_{\mathbf{v}} \varepsilon_{u,i} \ddot{\zeta}_i}{1 + \rho_i \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,i}) \ddot{\zeta}_i} \right), \ i \in \{r, t\}, \tag{31}$$

where $\ddot{\zeta}_i$ can be derived as follows:

$$\begin{aligned}
\ddot{\zeta}_i =& \mathbb{E} \left[\hat{\mathbf{h}}_i^H \mathbf{R}^{-1} \hat{\mathbf{h}}_i \right] \\
=& \text{Tr} \left[\mathbf{R}^{-1} \left(\overline{\mathbf{h}}_i \overline{\mathbf{h}}_i^H + \mathbf{C}_{\hat{\mathbf{h}}_i \hat{\mathbf{h}}_i} \right) \right] \\
=& \frac{\varrho_{\mathbf{A}_i} \varrho_{\mathbf{g}_i} \kappa_{\mathbf{A}_i} \kappa_{\mathbf{g}_i} \|\overline{\mathbf{a}}_i^{(\text{IOS})H} \mathbf{\Theta}_i \overline{\mathbf{g}}_i\|^2}{(1 + \kappa_{\mathbf{A}_i})(1 + \kappa_{\mathbf{g}_i})} \cdot \overline{\mathbf{a}}_i^{(\text{AP})H} \mathbf{R}^{-1} \overline{\mathbf{a}}_i^{(\text{AP})} \\
& + \text{Tr} \left[\mathbf{R}^{-1} \mathbf{C}_{\hat{\mathbf{h}}_i \hat{\mathbf{h}}_i} \right].
\end{aligned} \tag{32}$$

The upper bound can be attained by letting $M \to \infty$.

Proof: See Appendix C. ∎

B. IOS Beamforming

In this section, we design the reflective IOS beamformer matrix $\mathbf{\Theta}_r$ based on the statistical CSI of UE-R, i.e. $\overline{\mathbf{A}}_r$, $\overline{\mathbf{g}}_r$, and the transmit IOS beamformer matrix $\mathbf{\Theta}_t$ based on the statistical CSI of UE-T, i.e. $\overline{\mathbf{A}}_t$, $\overline{\mathbf{g}}_t$. This is carried out by maximizing the ergodic spectral efficiency upper bound of the UE-R, i.e. $\ddot{\mathcal{R}}_{\text{erg,r}}$, and that of the UE-T, i.e. $\ddot{\mathcal{R}}_{\text{erg,t}}$, respectively. Thus, we formulate the IOS beamforming problem as

$$\begin{aligned}
\text{(P1)} \quad &\max_{\mathbf{\Theta}_i} \ \ddot{\mathcal{R}}_{\text{erg,i}} \\
&\text{s.t.} \quad \theta_{i,n} \in [0, 2\pi), \ n = 1, 2, \cdots, N_i, \ i \in \{r, t\}. \tag{33}
\end{aligned}$$

Theorem 4. In practical systems, we have $\frac{\varepsilon_{u,r}}{1 - \varepsilon_{u,r}} \gg 1$, $\frac{\varepsilon_{u,t}}{1 - \varepsilon_{u,t}} \gg 1$ and $\frac{\varepsilon_{\mathbf{v}}}{1 - \varepsilon_{\mathbf{v}}} \gg 1$. Hence, the problem (P1) can be derived as

$$\begin{aligned}
\text{(P2)} \quad &\max_{\mathbf{\Theta}_i} \ \left\| \overline{\mathbf{A}}_i \mathbf{\Theta}_i \overline{\mathbf{g}}_i \right\|^2 \\
&\text{s.t.} \quad \theta_{i,n} \in [0, 2\pi), \ n = 1, 2, \cdots, N_i, \ i \in \{r, t\}. \tag{34}
\end{aligned}$$

Proof: See Appendix E. ∎

In problem (P2), based on (5) we can exploit that $\left\| \overline{\mathbf{A}}_i \mathbf{\Theta}_i \overline{\mathbf{g}}_i \right\|^2 = M \left\| \overline{\mathbf{a}}_i^{(\text{IOS})H} \mathbf{\Theta}_i \overline{\mathbf{g}}_i \right\|^2 \overset{(a)}{\leq} M N_i^2$, where the relationship in (a) is established when $[\mathbf{\Theta}_r]_{n,n} = \left[\overline{\mathbf{a}}_r^{(\text{IOS})} \right]_n \cdot \left[\overline{\mathbf{g}}_r^H \right]_n$ and $[\mathbf{\Theta}_t]_{n,n} = \left[\overline{\mathbf{a}}_t^{(\text{IOS})} \right]_n \cdot \left[\overline{\mathbf{g}}_t^H \right]_n$. According to (1), (3) and (6), we can express the optimal phase shift of the IOS elements as

$$\begin{aligned}
\theta_{i,n} =& \frac{2\pi}{\lambda} \left(\delta_x n_i^x \left(\sin \phi_i \cos \varphi_i - \sin \omega_i \cos \varpi_i \right) \right. \\
& \left. + \delta_y n_i^y \left(\cos \phi_i - \cos \omega_i \right) \right), \ i \in \{r, t\}, \tag{35}
\end{aligned}$$

with $n_i^x \in \{1, 2, \cdots, N_i^x\}$, $n_i^y \in \{1, 2, \cdots, N_i^y\}$, and $n = (n_i^y - 1) N_i^x + n_i^x$. Therefore, $\ddot{\zeta}_i$ in (31) is given by

$$\ddot{\zeta}_i = \frac{\varrho_{\mathbf{A}_i} \varrho_{\mathbf{g}_i} \kappa_{\mathbf{A}_i} \kappa_{\mathbf{g}_i} N_i^2}{(1 + \kappa_{\mathbf{A}_i})(1 + \kappa_{\mathbf{g}_i})} \overline{\mathbf{a}}_i^{(\text{AP})H} \mathbf{R}^{-1} \overline{\mathbf{a}}_i^{(\text{AP})} + \text{Tr} \left[\mathbf{R}^{-1} \mathbf{C}_{\hat{\mathbf{h}}_i \hat{\mathbf{h}}_i} \right], \tag{36}$$

where $\left\| \overline{\mathbf{a}}_i^{(\text{IOS})H} \mathbf{\Theta}_i \overline{\mathbf{g}}_i \right\|^2 = N_i^2$ in the matrices $\mathbf{R}$ and $\mathbf{C}_{\hat{\mathbf{h}}_i \hat{\mathbf{h}}_i}$.

Theorem 5. To characterize the scaling law of the transceiver HWIs versus the number of AP antennas M, and the number of IOS elements N_r and N_t, we present a loose upper bound of the ergodic spectral efficiency of the UE-R and UE-T, denoted as $\ddot{\mathcal{R}}'_{\text{erg,r}}$ and $\ddot{\mathcal{R}}'_{\text{erg,t}}$ respectively, as

$$\ddot{\mathcal{R}}'_{\text{erg,i}} = \log_2 \left(1 + \frac{\rho_i \varepsilon_{\mathbf{v}} \varepsilon_{u,i} M \eta_i}{\rho_i \varepsilon_{\mathbf{v}} (1 - \varepsilon_{u,i}) M \eta_i + \rho_r (1 - \varepsilon_{\mathbf{v}}) \eta_r + \rho_t (1 - \varepsilon_{\mathbf{v}}) \eta_t + \sigma_w^2} \right), \tag{37}$$

where $i \in \{r, t\}$ and $\eta_i = \frac{\varrho_{\mathbf{A}_i} \varrho_{\mathbf{g}_i} \kappa_{\mathbf{A}_i} \kappa_{\mathbf{g}_i} N_i^2 + (1 + \kappa_{\mathbf{g}_i} + \kappa_{\mathbf{A}_i}) N_i}{(1 + \kappa_{\mathbf{A}_i})(1 + \kappa_{\mathbf{g}_i})} + \varrho_{\mathbf{b}_i}$.

Proof: See Appendix F. ∎

Observe Theorem 5 that increasing the number of AP antennas M or the number of IOS elements N_r and N_t is beneficial for improving the ergodic spectral efficiency in the low signal-to-noise ratio (SNR) region. By contrast, there is a performance floor in the high SNR region, i.e. the ergodic spectral efficiency is limited by the hardware quality of the UEs. This means that deploying more AP antennas or IOS elements improves the ergodic spectral efficiency performance when the SNR is low. However, any ergodic spectral efficiency enhancement in the high SNR region requires high-quality hardware at the UEs.

C. Hardware-Quality Scaling Law

In this section, we present the scaling law of transceiver HWIs, with respect to the number of AP antennas M, and the number of IOS elements N_r and N_t, to gather valuable insights on the hardware efficiency.

1) AP HWI: Observe in (37) that in the low transmit power region, the egordic spectral efficiency degradation resulting from the AP HWI $\frac{(1 - \varepsilon_{\mathbf{v}}) M}{\varepsilon_{\mathbf{v}}}$ can be compensated by harnessing an additional $\frac{(1 - \varepsilon_{\mathbf{v}}) M}{\varepsilon_{\mathbf{v}}}$ number of AP antennas.

2) UE HWI: According to (37), when UE-R and UE-T have HWIs, the loss of ergodic spectral efficiency cannot be completely compensated by increasing either the number of AP antennas or that of the IOS elements. When the transmit power obeys $\rho_i \to \infty$, or $M \to \infty$, or $N_i \to \infty$, the loose egordic spectral efficiency upper bound $\ddot{\mathcal{R}}'_{\text{erg,i}}$ tends to saturate according to $\ddot{\mathcal{R}}'_{\text{erg,i}} \to \log_2 \left(1 + \frac{\varepsilon_{u,i}}{1 - \varepsilon_{u,i}} \right)$.

3) Number of IOS elements: We consider the scenario when the direct UE-AP link is completely blocked, i.e. $\varrho_{\mathbf{b}_i} = \varrho_{\mathbf{b}_i} = 0$. If the signals suffer from Rayleigh fading, i.e. $\kappa_{\mathbf{A}_i} = \kappa_{\mathbf{g}_i} = 0$, doubling the number of IOS elements N_i is equivalent to a 3dB enhancement of the transmit power ρ_i. By contrast, if the channel is perfectly known, i.e. $\kappa_{\mathbf{A}_i} = \kappa_{\mathbf{g}_i} \to \infty$, doubling the number of IOS elements N_i is equivalent to a 6dB enhancement of the transmit power ρ_i.

It is worth noting that, according to (37), increasing the transmit power ρ_i or the number of IOS elements N_i cannot compensate for the effect of the AP HWI or UE HWI since the signal distortion is exacerbated upon increasing of the transmit power and the number of IOS elements used.

V. NUMERICAL AND SIMULATION RESULTS

In this section, our theoretical analysis and simulation results characterizing the ergodic sum-rate of the UE-R and
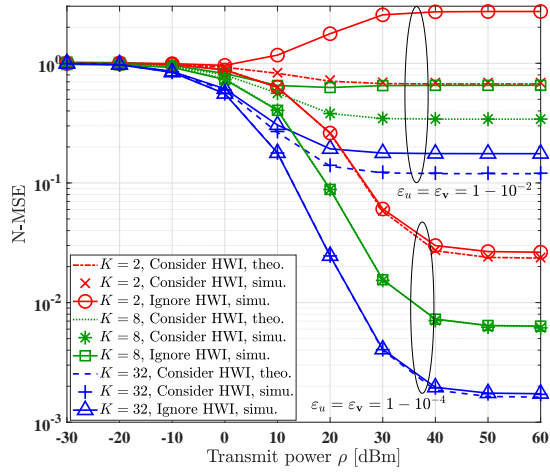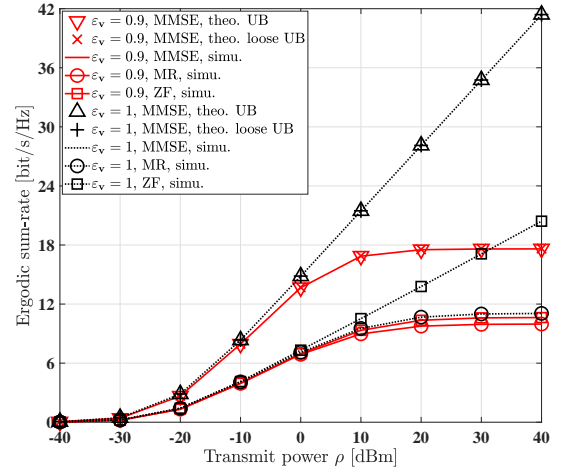
Fig. 3: Comparison of the N-MSE versus the transmit power.

(a) The UE hardware quality factor $\varepsilon_u = 1$

(b) The AP hardware quality factor $\varepsilon_{\mathbf{v}} = 1$

Fig. 4: Theoretical (31), (37) and simulation comparison of the ergodic sum-rate versus the transmit power in various combining methods.

UE-T are presented in the face of transceiver HWIs. Referring to the RIS-aided system setup in [44], the distance between the nodes of the AP, the UEs and the RIS ranges from tens to one hundred meters. Therefore, in the system model of Fig. 1, the AP, UE-R, UE-T, reflective IOS elements and transmit IOS elements are located at the Cartesian coordinates of (0m, -100m, 20m), (0m, -20m, 5m), (0m, 20m, 5m), (2m, 0m, 15m) and (-2m, 0m, 15m), respectively. Furthermore, ϕ_{r}, ϕ_{t}, φ_{r}, φ_{t}, ω_{r}, ω_{t}, ϖ_{r}, ϖ_{t} ψ_{r} and ψ_{t} are randomly generated following a uniform distribution. According to [44], we set $\mathrm{C}_0 = -30\mathrm{dB}$, $\sigma_w^2 = -90\mathrm{dBm}$, $\alpha_{\mathbf{A}_{\mathrm{r}}} = \alpha_{\mathbf{A}_{\mathrm{t}}} = \alpha_{\mathbf{g}_{\mathrm{r}}} = \alpha_{\mathbf{g}_{\mathrm{t}}} = 2.2$. Unless otherwise specified, the other simulation parameters are: $\alpha_{\mathbf{b}_{\mathrm{r}}} = \alpha_{\mathbf{b}_{\mathrm{t}}} = \alpha_{\mathbf{b}} = 4.8$, $K = 16$, $M = 100$, $N_{\mathrm{r}} = N_{\mathrm{t}} = 20 \times 20$, $\delta_x = \delta_y = \frac{\lambda}{2}$, $d_0 = \frac{\lambda}{2}$, $\Delta\psi = \psi_{\mathrm{r}} - \psi_{\mathrm{t}} = 0.1\pi$, $\rho_{\mathrm{r}} = \rho_{\mathrm{t}} = \rho = 20\mathrm{dBm}$, $\varepsilon_{u,\mathrm{r}} = \varepsilon_{u,\mathrm{t}} = \varepsilon_u$, $\kappa_{\mathbf{A}_{\mathrm{r}}} = \kappa_{\mathbf{A}_{\mathrm{t}}} = \kappa_{\mathbf{g}_{\mathrm{r}}} = \kappa_{\mathbf{g}_{\mathrm{t}}} = \kappa = 0\mathrm{dB}$, and the ergodic sum-rate performance variation with the change of these parameters will be evaluated in the following simulation results.

Fig. 3 compares the theoretical analysis (theo.) and the simulation results (simu.) of the N-MSE versus the transmit power ρ under different pilot overhead K with the number of AP antennas $M = 50$, where the legend 'Consider HWI' means the hardware impairment is considered in the LMMSE estimator, while the legend 'Ignore HWI' means the hardware impairment is ignored in the design of the LMMSE estimator. Fig. 3 shows that increasing the pilot overhead can effectively improve the N-MSE performance, while at the cost of transmitting more pilot sequences. With the increase of the transmit power, the N-MSE tends to a constant, since the channel estimation error dominantly results from the effect of HWI instead of the additive noise at the AP. Furthermore, Fig. 3 shows that ignoring the effect of HWI exacerbates the N-MSE performance especially in the low hardware quality factor region and high transmit power region, which demonstrates the necessity of considering the effect of HWI in the design of channel estimator.

Fig. 4 compares the ergodic sum-rate versus the transmit power ρ for various combining methods, including the MMSE, maximum ratio (MR) and zero-forcing (ZF) techniques. Apart from the simulation results, the theoretical upper bound (theo.
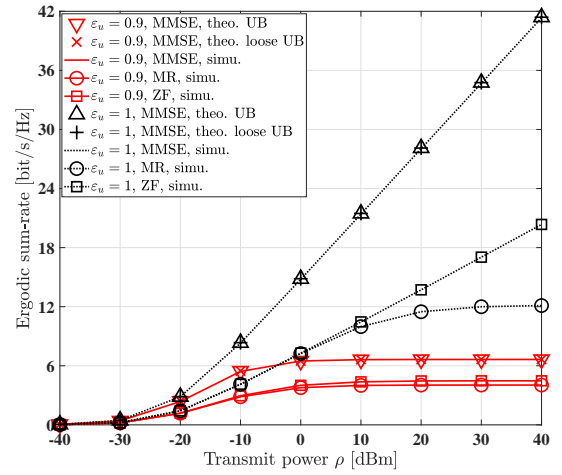
UB) in (31) and the loose theoretical upper bound (theo. loose UB) in (37) of the MMSE combining method are also presented for comparison. Fig. 4 (a) shows that the MMSE combiner promises higher ergodic sum-rate than the MR and ZF combiners. Observe that for the MR method, the design of combining vector for the UE-R only employs the information of the estimated channel, but ignores all other information, including the statistical channel estimation error, the interference arriving from the UE-T, the effect of HWIs and the additive noise at the AP. On the other hand, in the design of the combining vector for UE-R based on the ZF method, the hostile interference emanating from the UE-T is eliminated, but the effects of HWIs, of the channel estimation error and of the additive noise at the AP are amplified. By contrast, the MMSE method comprehensively considers the statistical information concerning the channel estimation error, the inter-user interference, the HWIs and the additive noise at the AP in the design of combining vectors. Observe that when $\varepsilon_u = \varepsilon_{\mathbf{v}} = 1$, the ergodic sum-rate increase linearly with the transmit power, while when the AP hardware quality factor
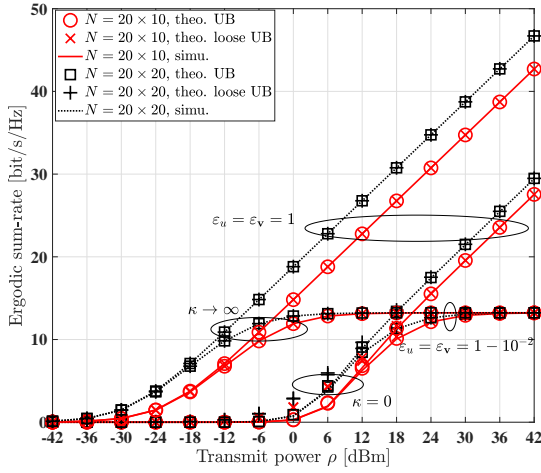
Fig. 5: Theoretical (31), (37) and simulation comparison of the ergodic sum-rate versus the transmit power under different number of IOS elements, hardware quality factor and Rician factor.
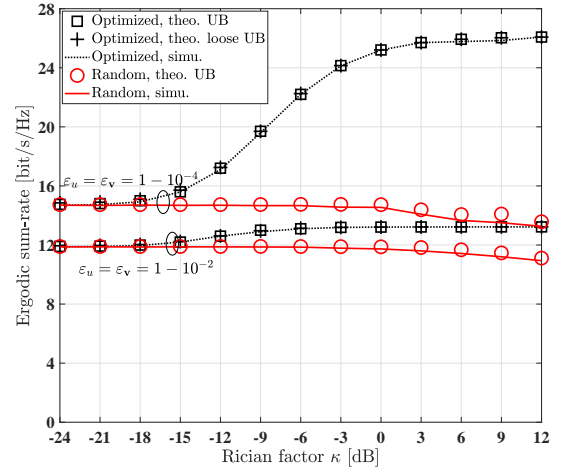

Fig. 7: Theoretical (31), (37) and simulation comparison of the ergodic sum-rate versus the Rician factor.
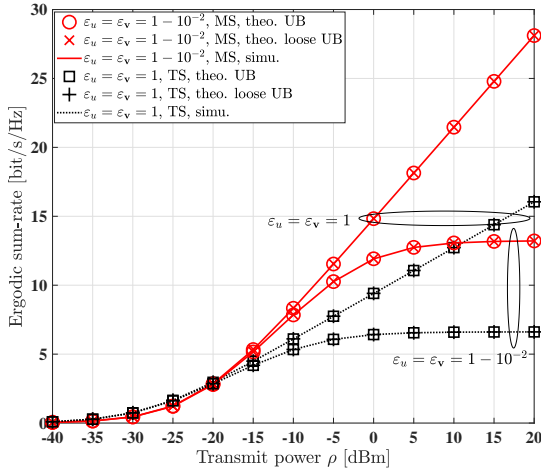

Fig. 6: Theoretical (31), (37) and simulation comparison of the ergodic sum-rate versus the transmit power under different hardware quality factor for the IOS-assisted systems based on the MS protocol and the TS protocol, respectively.

obeys $\varepsilon_{\mathbf{v}} < 1$, the ergodic sum-rate tends to a constant in the high transmit power region, since the signal distortion at the AP is amplified upon increasing of the transmit power. Fig. 4 (b) compares the ergodic sum-rate versus the transmit power ρ for the MMSE, MR and ZF, under different UE hardware quality factors ε_u, along with an ideal AP hardware, i.e. for $\varepsilon_{\mathbf{v}} = 1$. Observe that upon increasing the transmit power, the ergodic sum-rate of the MMSE method approximately equals $\log_2 \left(1 + \frac{\varepsilon_{u,\mathrm{r}}}{1-\varepsilon_{u,\mathrm{r}}}\right) + \log_2 \left(1 + \frac{\varepsilon_{u,\mathrm{t}}}{1-\varepsilon_{u,\mathrm{t}}}\right) = \log_2 \left(1 + \frac{0.9}{1-0.9}\right) + \log_2 \left(1 + \frac{0.9}{1-0.9}\right) \approx 6.644$, which verifies our theoretical analysis.

Fig. 5 compares the ergodic sum-rate versus the transmit power ρ for different number of the IOS elements, hardware quality factors and Rician factors, with the phase shift of IOS elements being optimally designed based on (35). Observe that when $\varepsilon_{\mathbf{v}} = \varepsilon_u = 1$, the ergodic sum-rate tends to

infinity upon increasing the transmit power. However, when $\varepsilon_{\mathbf{v}} = \varepsilon_u = 1 - 10^{-2}$, the ergodic sum-rate saturates at $2\log_2 \left(1 + \frac{1-10^{-2}}{10^{-2}}\right) \approx 13.29$. Furthermore, it is worth noting that for realistic HWIs, increasing the number of IOS elements cannot increase the ergodic sum-rate in the high transmit power region since it is limited by the hardware quality of UE-R and UE-T. Besides, when the channel is dominated by LoS components, i.e. $\kappa \to \infty$, doubling the number of IOS elements can bring approximately 6dB transmit power gain. By contrast, when the channel is dominated by NLoS components, i.e. $\kappa = 0$, doubling the number of IOS elements can bring approximately 3dB transmit power gain.

Fig. 6 compares the ergodic sum-rate versus the transmit power ρ under different hardware quality factors for the IOS-assisted systems based on the MS protocol and the TS protocol, respectively. The ergodic sum-rate based on the MS protocol is almost 2 times that based on the TS protocol. This is due to the fact that the UE-T and the UE-R are allocated to the orthogonal time resources of the TS protocol, while they are instantaneously supported within the same time resource in the MS protocol.

Fig. 7 compares the ergodic sum-rate versus the Rician factor κ under different hardware quality factors. Observe that in the low Rician factor region, i.e. when the NLoS components are dominant, the ergodic sum-rate associated with the optimal IOS phase shift is almost the same as that for random IOS phase shifts. By contrast, in the high Rician factor region, i.e. when the LoS components are dominant, the ergodic sum-rate associated with optimal IOS phase shifts is significantly better than that for random IOS phase shift, since the IOS phase shift is purely designed based on the LoS components.

Fig. 8 (a) compares the ergodic sum-rate versus the number of AP antennas M for different transmit powers and AP hardware quality factors, with the UE hardware quality factor set to $\varepsilon_u = 1$. Observe that in the low transmit power region, such as $\rho = -15$dBm, employing slightly more AP antennas efficiently compensates the spectral efficiency degradation
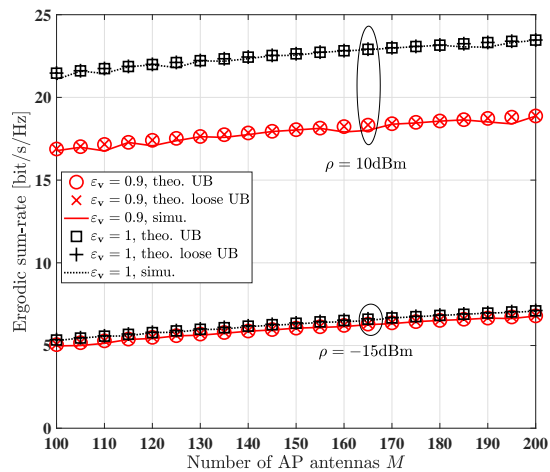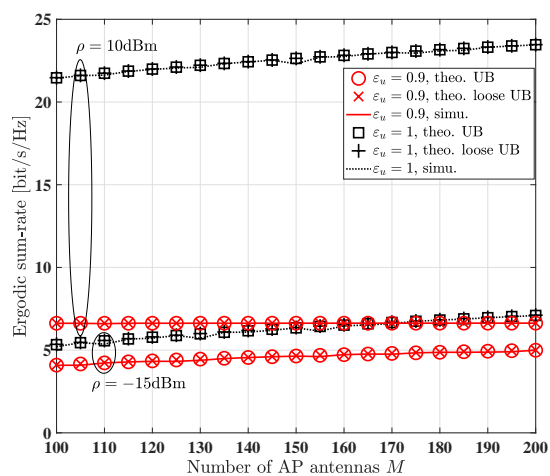
(a) The UE hardware quality factor $\varepsilon_u = 1$

(b) The AP hardware quality factor $\varepsilon_{\mathbf{v}} = 1$

Fig. 8: Theoretical (31), (37) and simulation comparison of the ergodic sum-rate versus the number of AP antennas M.

Fig. 9: Simulation comparison of the ergodic sum-rate versus the AoA difference at the AP $\Delta\psi = \psi_{\mathrm{r}} - \psi_{\mathrm{t}}$.

Fig. 10: Simulation comparison of the ergodic sum-rate versus the path loss exponent of direct UE-AP links $\alpha_{\mathbf{b}}$ for the IOS-assisted system and the conventional system without IOS, respectively.

resulting from HWIs. For example, the ergodic sum-rate for $M = 110$ AP antennas with the AP hardware quality factor of $\varepsilon_{\mathbf{v}} = 0.9$ is almost the same as that for $M = 100$ AP antennas with ideal AP hardware. In the high transmit power region, such as $\rho = 10$dBm, using more AP antennas can also compensate the spectral efficiency degradation resulting from HWIs, but more AP antennas are required than in the low transmit power region. Fig. 8 (b) compares the ergodic sum-rate versus the number of AP antennas M for different transmit power regions ($\rho = 10$dBm and $\rho = -15$dBm) and UE hardware quality factors, along with the AP hardware quality factor of $\varepsilon_{\mathbf{v}} = 1$. In contrast to the systems in Fig. 8 (a) having non-ideal AP hardware and ideal UE hardware, employing more AP antennas has limited benefit in terms of compensating for the UE hardware impairments. For example, in the low transmit power region of $\rho = -15$dBm, the ergodic sum-rate for $M = 200$ AP antennas and the UE hardware quality factor of $\varepsilon_u = 0.9$ is still lower than that for $M = 100$ AP antennas and ideal UE hardware. Furthermore, in the high transmit power region of $\rho = 10$dBm, the ergodic sum-rate for the UE hardware quality factor of $\varepsilon_u = 0.9$ has a large

gap compared to that associated with ideal UE hardware. As a further observation, using more AP antennas fails to compensate for the UE HWIs.

Fig. 9 presents the simulation results of the ergodic sum-rate versus the AoA difference $\Delta\psi = \psi_{\mathrm{r}} - \psi_{\mathrm{t}}$ at the AP for different number of AP antennas and UE hardware quality factors. Observe that the ergodic sum-rate is improved upon increasing the AoA difference at the AP. Furthermore, employing more AP antennas allows the IOS-based wireless system to accommodate a wider AoA difference range. For example, although when the AoA difference at the AP is small, such as $\Delta\psi = \pm\pi/128$, the ergodic sum-rate can be improved upon using more AP antennas.

Fig. 10 presents the simulation results of the ergodic sum-rate versus the path loss exponent of the direct UE-AP links $\alpha_{\mathbf{b}}$ for the IOS-assisted system and the conventional system without IOS, respectively. Observe that when the IOS is absent, the ergodic sum-rate degrades with the increase of the path loss exponent of the direct UE-AP links. Furthermore,

when the path loss exponent of the direct UE-AP links is small, the ergodic sum-rate performance cannot be substantially improved by deploying the IOS elements, which is due to the fact that the direct UE-AP channel links are more reliable than that created by the IOS. By contrast, when the path loss exponent of direct UE-AP links is high, deploying the IOS elements is beneficial for the improvement of the ergodic sum-rate, since reliable channel links are created by the IOS elements, even though the signals suffer from severe attenuation in the direct UE-AP links.

VI. CONCLUSIONS

The ergodic spectral efficiency of the IOS-aided MIMO uplink associated with imperfect CSI and transceiver HWIs was characterized. Firstly the LMMSE estimator of the e-quivalent channel spanning from the UEs to the AP was derived, in the face of transceiver HWIs. Next, a two-timescale protocol was conceived for the joint beamformer design of IOS-aided systems. At the AP, an MMSE combiner was employed based on the estimated equivalent channels, as well as on the statistical channel estimation error, on the inter-user interference and on transceiver HWIs. Then, we derived the closed-form expression of the optimal IOS phase shift based on the statistical CSI for maximizing the upper bound of the ergodic spectral efficiency. Our theoretical analysis and numerical results show that the transceiver HWIs have a grave deleterious impact on the ergodic spectral efficiency. Employing more AP antennas can compensate for the HWIs at the AP, but the HWI at the UEs cannot be compensated by using more AP antennas or by increasing the transmit power.

APPENDIX A
PROOF OF THEOREM 1

In (24), based on the independence of s_r, s_t, u_r, u_t, $\mathbf{v}_\mathrm{r}$, $\mathbf{v}_\mathrm{t}$ and $\mathbf{w}$, the auto-correlation matrix of $\mathbf{y}$ can be formulated as

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{y}\mathbf{y}^\mathrm{H}\right] = {}& \rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\hat{\mathbf{h}}_\mathrm{r}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H} + \rho_\mathrm{t}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{t}}\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H} + \rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\mathbf{C}_{\check{\mathbf{h}}_\mathrm{r}\check{\mathbf{h}}_\mathrm{r}} \\
& + \rho_\mathrm{t}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{t}}\mathbf{C}_{\check{\mathbf{h}}_\mathrm{t}\check{\mathbf{h}}_\mathrm{t}} + \rho_\mathrm{r}\varepsilon_\mathbf{v}\left(1-\varepsilon_{u,\mathrm{r}}\right)\left(\hat{\mathbf{h}}_\mathrm{r}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H} + \mathbf{C}_{\check{\mathbf{h}}_\mathrm{r}\check{\mathbf{h}}_\mathrm{r}}\right) \\
& + \rho_\mathrm{t}\varepsilon_\mathbf{v}\left(1-\varepsilon_{u,\mathrm{t}}\right)\left(\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H} + \mathbf{C}_{\check{\mathbf{h}}_\mathrm{t}\check{\mathbf{h}}_\mathrm{t}}\right) + \left(\rho_\mathrm{r}\left(1-\varepsilon_\mathbf{v}\right)\cdot \right. \\
& \left. \mathbb{E}\left[\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}^\mathrm{H}\right] + \rho_\mathrm{t}\left(1-\varepsilon_\mathbf{v}\right)\mathbb{E}\left[\mathbf{h}_\mathrm{t}\mathbf{h}_\mathrm{t}^\mathrm{H}\right]\right) \odot \mathbf{I}_M + \sigma_w^2\mathbf{I}_M \\
\overset{(a)}{=} {}& \rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\hat{\mathbf{h}}_\mathrm{r}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H} + \rho_\mathrm{t}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{t}}\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H} + \mathbf{R}, \qquad (38)
\end{aligned}
$$

where (a) is based on (10), (11) and (27). The MMSE combining vector and the SINR of UE-R are given by

$$
\mathbf{q}_\mathrm{r} = \rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\left(\mathbb{E}\left[\mathbf{y}\mathbf{y}^\mathrm{H}\right]\right)^{-1}\hat{\mathbf{h}}_\mathrm{r}, \qquad (39)
$$

$$
\gamma_\mathrm{r} = \rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\left(\mathbb{E}[\mathbf{y}\mathbf{y}^\mathrm{H}] - \rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\hat{\mathbf{h}}_\mathrm{r}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\right)^{-1}\hat{\mathbf{h}}_\mathrm{r}. \qquad (40)
$$

According to (38), (39) and (40), we can arrive at (25) and (28). Furthermore, the MMSE combining vector and the SINR of UE-T in (26) and (29) can be similarly evaluated.

APPENDIX B
PROOF OF THEOREM 2

Lemma 1. For the matrices $\mathbf{D} \in \mathbb{C}^{L_1 \times L_1}$, $\mathbf{F}_1 \in \mathbb{C}^{L_1 \times L_2}$, $\mathbf{F}_2 \in \mathbb{C}^{L_2 \times L_2}$, $\mathbf{F}_3 \in \mathbb{C}^{L_2 \times L_1}$, the inverse of $\mathbf{D} + \mathbf{F}_1\mathbf{F}_2\mathbf{F}_3$ is

$\mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{F}_1\left(\mathbf{F}_3\mathbf{D}^{-1}\mathbf{F}_1 + \mathbf{F}_2^{-1}\right)^{-1}\mathbf{F}_3\mathbf{D}^{-1}$ [45].

Based on Lemma 1, letting $L_1 = M$, $L_2 = 1$, $L_3 = M$, $\mathbf{D} = \rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H} + \mathbf{R}$, $\mathbf{F}_1 = \sqrt{\rho_\mathrm{r}\varepsilon_\mathbf{v}\left(1-\varepsilon_{u,\mathrm{r}}\right)}\hat{\mathbf{h}}_\mathrm{r}$, $\mathbf{F}_2 = 1$ and $\mathbf{F}_3 = \sqrt{\rho_\mathrm{r}\varepsilon_\mathbf{v}\left(1-\varepsilon_{u,\mathrm{r}}\right)}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}$ for (28), we can formulate γ_r as

$$
\gamma_\mathrm{r} = \frac{\rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\left(\rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H} + \mathbf{R}\right)^{-1}\hat{\mathbf{h}}_\mathrm{r}}{1 + \rho_\mathrm{r}\varepsilon_\mathbf{v}\left(1-\varepsilon_{u,\mathrm{r}}\right)\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\left(\rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H} + \mathbf{R}\right)^{-1}\hat{\mathbf{h}}_\mathrm{r}}. \qquad (41)
$$

Again, based on Lemma 1, letting $L_1 = M$, $L_2 = 1$, $L_3 = M$, $\mathbf{D} = \mathbf{R}$, $\mathbf{F}_1 = \sqrt{\rho_\mathrm{t}\varepsilon_\mathbf{v}}\hat{\mathbf{h}}_\mathrm{t}$, $\mathbf{F}_2 = 1$ and $\mathbf{F}_3 = \sqrt{\rho_\mathrm{t}\varepsilon_\mathbf{v}}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}$ for (41), we have $\gamma_\mathrm{r} = \frac{\rho_\mathrm{r}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{r}}\zeta_\mathrm{r}}{1+\rho_\mathrm{r}\varepsilon_\mathbf{v}(1-\varepsilon_{u,\mathrm{r}})\zeta_\mathrm{r}}$. Similarly, we can get $\gamma_\mathrm{t} = \frac{\rho_\mathrm{t}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{t}}\zeta_\mathrm{t}}{1+\rho_\mathrm{t}\varepsilon_\mathbf{v}(1-\varepsilon_{u,\mathrm{t}})\zeta_\mathrm{t}}$. Therefore, we can arrive at the instantaneous ergodic spectral efficiency of UE-R and UE-T in (30).

APPENDIX C
PROOF OF THEOREM 3

Since $f(\zeta_\mathrm{i}) = \log_2\left(1 + \frac{\rho_\mathrm{i}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{i}}\zeta_\mathrm{i}}{1+\rho_\mathrm{i}\varepsilon_\mathbf{v}(1-\varepsilon_{u,\mathrm{i}})\zeta_\mathrm{i}}\right)$ is a concave function, according to Jensen's inequality we can get

$$
\mathcal{R}_{\mathrm{erg,i}} \leq \log_2\left(1 + \frac{\rho_\mathrm{i}\varepsilon_\mathbf{v}\varepsilon_{u,\mathrm{i}}\mathbb{E}[\zeta_\mathrm{i}]}{1 + \rho_\mathrm{i}\varepsilon_\mathbf{v}\left(1-\varepsilon_{u,\mathrm{i}}\right)\mathbb{E}[\zeta_\mathrm{i}]}\right). \qquad (42)
$$

In (42), $\mathbb{E}[\zeta_\mathrm{r}]$ and $\mathbb{E}[\zeta_\mathrm{t}]$ are given by

$$
\begin{aligned}
\mathbb{E}[\zeta_\mathrm{r}] = {}& \mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}\right] - \mathbb{E}\left[\frac{\rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\hat{\mathbf{h}}_\mathrm{r}}{1 + \rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}}\right] \\
& \overset{(a)}{\leq} \mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}\right], \qquad (43)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\zeta_\mathrm{t}] = {}& \mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}\right] - \mathbb{E}\left[\frac{\rho_\mathrm{r}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\hat{\mathbf{h}}_\mathrm{t}}{1 + \rho_\mathrm{r}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}}\right] \\
& \overset{(b)}{\leq} \mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}\right], \qquad (44)
\end{aligned}
$$

where the equalities in (a) and (b) are established when $M \to \infty$ in conjunction with $\cos\psi_\mathrm{t} \neq \cos\psi_\mathrm{r}$ and the proof is presented in Appendix D. Furthermore, it can be shown that when $M \to \infty$, the derivative of the function $f(\zeta_\mathrm{i})$ with respect to ζ_i tends to 0. Therefore, (31) is the upper bound of the ergodic spectral efficiency, and upon increasing of M, the ergodic spectral efficiency tends to its upper bound. Furthermore, $\mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}\right]$ in (43) and $\mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}\right]$ in (44) can be exploited to arrive at (32) based on (8), (18) and (21).

APPENDIX D
PROOF OF (42), (43) AND (44)

The proof of $\mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}\right] - \mathbb{E}\left[\frac{\rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}}{1+\rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}}\right]$ tends to $\mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}\right]$ when $M \to \infty$ and $\cos\psi_\mathrm{t} \neq \cos\psi_\mathrm{r}$ is equivalent to the following proposition:

Proposition 1. When $M \to \infty$ and $\cos\psi_\mathrm{t} \neq \cos\psi_\mathrm{r}$, we have

$$
\frac{\mathbb{E}\left[\frac{\rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}}{1+\rho_\mathrm{t}\varepsilon_\mathbf{v}\hat{\mathbf{h}}_\mathrm{t}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{t}}\right]}{\mathbb{E}\left[\hat{\mathbf{h}}_\mathrm{r}^\mathrm{H}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathrm{r}\right]} = 0. \qquad (45)
$$

In (45), let $\mathbf{R} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ represent the eigenvalue decomposition of $\mathbf{R}$ with $\mathbf{P}$ being a unitary matrix and $\boldsymbol{\Lambda} = \mathbf{Diag}\{\lambda_1, \lambda_2, \cdots, \lambda_M\}$. Then, upon defining $\hat{\mathbf{h}}_r' = \sqrt{\boldsymbol{\Lambda}^{-1}}\mathbf{P}^{-1}\hat{\mathbf{h}}_r$, $\hat{\mathbf{h}}_t' = \sqrt{\boldsymbol{\Lambda}^{-1}}\mathbf{P}^{-1}\hat{\mathbf{h}}_t$, $\overline{\mathbf{h}}_r' = \sqrt{\boldsymbol{\Lambda}^{-1}}\mathbf{P}^{-1}\overline{\mathbf{h}}_r$ and $\overline{\mathbf{h}}_t' = \sqrt{\boldsymbol{\Lambda}^{-1}}\mathbf{P}^{-1}\overline{\mathbf{h}}_t$, the numerator of (45) can be expressed as

$$\mathbb{E}\left[\frac{\rho_t \varepsilon_{\mathbf{v}} \hat{\mathbf{h}}_r^H \mathbf{R}^{-1}\hat{\mathbf{h}}_t \hat{\mathbf{h}}_t^H \mathbf{R}^{-1}\hat{\mathbf{h}}_r}{1+\rho_t \varepsilon_{\mathbf{v}} \hat{\mathbf{h}}_t^H \mathbf{R}^{-1}\hat{\mathbf{h}}_t}\right]$$

$$< \mathbb{E}\left[\frac{\hat{\mathbf{h}}_r'^H \hat{\mathbf{h}}_t' \hat{\mathbf{h}}_t'^H \hat{\mathbf{h}}_r'}{\hat{\mathbf{h}}_t'^H \hat{\mathbf{h}}_t'}\right]$$

$$= \mathbb{E}\left[\left|(\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r')^H \frac{\hat{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|} + \overline{\mathbf{h}}_r'^H \frac{\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|} + \overline{\mathbf{h}}_r'^H \frac{\overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right]$$

$$\overset{(a)}{=} \mathbb{E}\left[\left|\left(\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r'\right)^H \frac{\hat{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right] + \mathbb{E}\left[\left|\overline{\mathbf{h}}_r'^H \frac{\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right]$$

$$+ \mathbb{E}\left[\left|\overline{\mathbf{h}}_r'^H \frac{\overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right], \tag{46}$$

where (a) is based on $\mathbb{E}\left[\left(\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r'\right)^H \frac{\hat{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|} \frac{\hat{\mathbf{h}}_t'^H}{\|\hat{\mathbf{h}}_t'\|} \overline{\mathbf{h}}_r'\right] = 0$ and $\mathbb{E}\left[\overline{\mathbf{h}}_r'^H \frac{\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|} \frac{\overline{\mathbf{h}}_t'^H}{\|\hat{\mathbf{h}}_t'\|} \overline{\mathbf{h}}_r'\right] = 0$. Since $\left\|\frac{\hat{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right\| = 1$ and the elements in $\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r'$ have identical covariance, we may write:

$$\mathbb{E}\left[\left|(\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r')^H \frac{\hat{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right] \le \frac{\lambda_{\min}^{-2}}{M}\mathbb{E}\left[\|\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r'\|^2\right]$$

$$\overset{(a)}{\le} \frac{\lambda_{\min}^{-2}}{M}\mathbb{E}\left[\|\hat{\mathbf{h}}_r'\|^2\right], \tag{47}$$

where $\lambda_{\min}$ represents the minimum value of $\lambda_1, \lambda_2, \cdots, \lambda_M$, and (a) is based on $\left\|\hat{\mathbf{h}}_r'\right\|^2 = \left\|\overline{\mathbf{h}}_r'\right\|^2 + \left\|\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r'\right\|^2$. In $\mathbb{E}\left[|\overline{\mathbf{h}}_r'^H \frac{\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}|^2\right]$, since $\left\|\frac{\overline{\mathbf{h}}_r'}{\|\overline{\mathbf{h}}_r'\|}\right\| = 1$ and the elements in $\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'$ have the identical covariance, we can get

$$\mathbb{E}\left[\left|\overline{\mathbf{h}}_r'^H \frac{\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right] = \mathbb{E}\left[\left|\frac{\overline{\mathbf{h}}_r'^H}{\|\overline{\mathbf{h}}_r'\|}(\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t')\frac{\|\overline{\mathbf{h}}_r'\|}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right]$$

$$\overset{(a)}{\le} \frac{\lambda_{\min}^{-2}}{M}\mathbb{E}\left[\|\hat{\mathbf{h}}_r'\|^2\right], \tag{48}$$

where (a) is based on exploiting that $\left\|\hat{\mathbf{h}}_t'\right\|^2 = \left\|\overline{\mathbf{h}}_t'\right\|^2 + \left\|\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'\right\|^2 \ge \left\|\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'\right\|^2$ and $\left\|\hat{\mathbf{h}}_r'\right\|^2 = \|\overline{\mathbf{h}}_r'\|^2 + \left\|\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r'\right\|^2 \ge \left\|\overline{\mathbf{h}}_r'\right\|^2$. Furthermore, $\mathbb{E}[|\overline{\mathbf{h}}_r'^H \frac{\overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}|^2]$ can be derived as

$$\mathbb{E}\left[\left|\overline{\mathbf{h}}_r'^H \frac{\overline{\mathbf{h}}_t'}{\|\hat{\mathbf{h}}_t'\|}\right|^2\right]$$

$$= \left\|\overline{\mathbf{a}}_t^{(AP)H}\mathbf{R}^{-1}\overline{\mathbf{a}}_r^{(AP)}\right\|^2 \cdot \frac{\varrho_{\mathbf{A}_t}\varrho_{\mathbf{g}_t}\kappa_{\mathbf{A}_t}\kappa_{\mathbf{g}_t}\left\|\overline{\mathbf{a}}_t^{(IOS)H}\boldsymbol{\Theta}_t\overline{\mathbf{g}}_t\right\|^2}{(1+\kappa_{\mathbf{A}_t})(1+\kappa_{\mathbf{g}_t})}\cdot$$

$$\frac{\varrho_{\mathbf{A}_r}\varrho_{\mathbf{g}_r}\kappa_{\mathbf{A}_r}\kappa_{\mathbf{g}_r}\left\|\overline{\mathbf{a}}_r^{(IOS)H}\boldsymbol{\Theta}_r\overline{\mathbf{g}}_r\right\|^2}{(1+\kappa_{\mathbf{A}_r})(1+\kappa_{\mathbf{g}_r})}\cdot \mathbb{E}\left[\frac{1}{\|\hat{\mathbf{h}}_t'\|^2}\right]$$

$$= \frac{1}{M^2}\left\|\overline{\mathbf{h}}_r'\right\|^2\left\|\overline{\mathbf{h}}_t'\right\|^2\sum_{m=1}^{M}\lambda_m^{-1}e^{j2\pi d_0 m(\cos\psi_t - \cos\psi_r)}\cdot\mathbb{E}\left[\frac{1}{\|\hat{\mathbf{h}}_t'\|^2}\right]$$

$$\overset{(a)}{\le} \frac{1}{M^2}\sum_{m=1}^{M}\lambda_m^{-1}e^{j2\pi d_0 m(\cos\psi_t - \cos\psi_r)}\cdot\mathbb{E}\left[\|\hat{\mathbf{h}}_r'\|^2\right], \tag{49}$$

where (a) is based on $\left\|\hat{\mathbf{h}}_r'\right\|^2 = \left\|\overline{\mathbf{h}}_r'\right\|^2 + \left\|\hat{\mathbf{h}}_r' - \overline{\mathbf{h}}_r'\right\|^2 \ge \left\|\overline{\mathbf{h}}_r'\right\|^2$, $\left\|\hat{\mathbf{h}}_t'\right\|^2 = \left\|\overline{\mathbf{h}}_t'\right\|^2 + \left\|\hat{\mathbf{h}}_t' - \overline{\mathbf{h}}_t'\right\|^2 \ge \left\|\overline{\mathbf{h}}_t'\right\|^2$.

The denominator in (45) can be further expressed as

$$\mathbb{E}\left[\hat{\mathbf{h}}_r^H \mathbf{R}^{-1}\hat{\mathbf{h}}_r\right] = \mathbb{E}\left[\left\|\hat{\mathbf{h}}_t'\right\|^2\right]. \tag{50}$$

According to (46), (47), (48), (49) and (50), we can get $\frac{\mathbb{E}\left[\frac{\rho_t \varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_r^H\mathbf{R}^{-1}\hat{\mathbf{h}}_t\hat{\mathbf{h}}_t^H\mathbf{R}^{-1}\hat{\mathbf{h}}_r}{1+\rho_t\varepsilon_{\mathbf{v}}\hat{\mathbf{h}}_t^H\mathbf{R}^{-1}\hat{\mathbf{h}}_t}\right]}{\mathbb{E}\left[\hat{\mathbf{h}}_r^H\mathbf{R}^{-1}\hat{\mathbf{h}}_r\right]} < \frac{\lambda_{\min}^{-2}}{M}\left(2 + \frac{\sin^2(\pi d_0 M(\cos\psi_t - \cos\psi_r))}{M\sin^2(\pi d_0(\cos\psi_t - \cos\psi_r))}\right)$, which tends to 0 when $M \to \infty$ and $\cos\psi_t \ne \cos\psi_r$.

APPENDIX E
PROOF OF THEOREM 4

When $\frac{\varepsilon_{u,r}}{1-\varepsilon_{u,r}} \gg 1$, $\frac{\varepsilon_{u,t}}{1-\varepsilon_{u,t}} \gg 1$ and $\frac{\varepsilon_{\mathbf{v}}}{1-\varepsilon_{\mathbf{v}}} \gg 1$, based on (17), $\mathbf{C}_{\mathbf{x}_i\mathbf{x}_i}$ is approximately given by $\mathbf{C}_{\mathbf{x}_i\mathbf{x}_i} = \rho_i\varepsilon_{\mathbf{v}}(1+(K-1)\varepsilon_{u,i})\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i} + \sigma_w^2\mathbf{I}_M$. Therefore, $\mathbf{C}_{\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i}$ and $\mathbf{C}_{\check{\mathbf{h}}_i\check{\mathbf{h}}_i}$ can be approximated as

$$\mathbf{C}_{\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i} = K\rho_i\varepsilon_{\mathbf{v}}\varepsilon_{u,i}\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i}(\rho_i\varepsilon_{\mathbf{v}}(1+(K-1)\varepsilon_{u,i})\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i}$$
$$+\sigma_w^2\mathbf{I}_M)^{-1}\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i}, \tag{51}$$

$$\mathbf{C}_{\check{\mathbf{h}}_i\check{\mathbf{h}}_i} = \mathbf{C}_{\mathbf{h}_i\mathbf{h}_i} - K\rho_i\varepsilon_{\mathbf{v}}\varepsilon_{u,i}\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i}(\rho_i\varepsilon_{\mathbf{v}}(1+(K-1)\varepsilon_{u,i})\cdot$$
$$\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i} + \sigma_w^2\mathbf{I}_M)^{-1}\mathbf{C}_{\mathbf{h}_i\mathbf{h}_i}. \tag{52}$$

According to (11), (51) and (52), we can arrive at $\frac{\partial\mathbf{C}_{\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i}}{\partial\left\|\overline{\mathbf{a}}_i^{(IOS)H}\boldsymbol{\Theta}_i\overline{\mathbf{g}}_i\right\|^2} = \mathbf{O}_{M\times M}$ and $\frac{\partial\mathbf{C}_{\check{\mathbf{h}}_i\check{\mathbf{h}}_i}}{\partial\|\overline{\mathbf{a}}_i^{(IOS)H}\boldsymbol{\Theta}_i\overline{\mathbf{g}}_i\|^2} = \mathbf{O}_{M\times M}$.

The partial derivative of $\ddot{\mathcal{R}}_{erg,r}$ with respect to $\|\overline{\mathbf{a}}_r^{(IOS)H}\boldsymbol{\Theta}_r\overline{\mathbf{g}}_r\|^2$ is given by

$$\frac{\partial\ddot{\mathcal{R}}_{erg,r}}{\partial\left\|\overline{\mathbf{a}}_r^{(IOS)H}\boldsymbol{\Theta}_r\overline{\mathbf{g}}_r\right\|^2}$$

$$= \frac{\log_2 e \cdot \rho_r\varepsilon_{\mathbf{v}}\varepsilon_{u,r}}{\left(1+\rho_r\varepsilon_{\mathbf{v}}\ddot{\zeta}_r\right)\left(1+\rho_r\varepsilon_{\mathbf{v}}(1-\varepsilon_{u,r})\ddot{\zeta}_r\right)}\cdot$$

$$\left(\frac{\varrho_{\mathbf{A}_r}\varrho_{\mathbf{g}_r}\kappa_{\mathbf{A}_r}\kappa_{\mathbf{g}_r}}{(1+\kappa_{\mathbf{A}_r})(1+\kappa_{\mathbf{g}_r})}\cdot\mathrm{Tr}\left[\mathbf{R}^{-1}\overline{\mathbf{a}}_r^{(AP)}\overline{\mathbf{a}}_r^{(AP)H}\right]\right.$$

$$\left.+\frac{\varrho_{\mathbf{A}_r}\varrho_{\mathbf{g}_r}\kappa_{\mathbf{A}_r}\kappa_{\mathbf{g}_r}\left\|\overline{\mathbf{a}}_r^{(IOS)H}\boldsymbol{\Theta}_r\overline{\mathbf{g}}_r\right\|^2}{(1+\kappa_{\mathbf{A}_r})(1+\kappa_{\mathbf{g}_r})}\cdot\frac{\partial\mathrm{Tr}\left[\mathbf{R}^{-1}\overline{\mathbf{a}}_r^{(AP)}\overline{\mathbf{a}}_r^{(AP)H}\right]}{\partial\left\|\overline{\mathbf{a}}_r^{(IOS)H}\boldsymbol{\Theta}_r\overline{\mathbf{g}}_r\right\|^2}\right.$$

$$+\frac{\partial \mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}}\right]}{\partial \left\|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\right\|^2}\Bigg),\tag{53}$$

where $\frac{\partial \mathrm{Tr}\left[\mathbf{R}^{-1}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}}\right]}{\partial \left\|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\right\|^2}$ can be further derived as

$$\frac{\partial \mathrm{Tr}\left[\mathbf{R}^{-1}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}}\right]}{\partial \|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\|^2} = -\frac{\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\kappa_{\mathbf{A}_{\mathrm{r}}}\kappa_{\mathbf{g}_{\mathrm{r}}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)}\cdot$$
$$\mathrm{Tr}\left[\mathbf{R}^{-1}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}}\mathbf{R}^{-1}\right],\tag{54}$$

and $\frac{\partial \mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}}\right]}{\partial \left\|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\right\|^2}$ can be further rewritten as

$$\frac{\partial \mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}}\right]}{\partial \|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\|^2} = -\frac{\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\kappa_{\mathbf{A}_{\mathrm{r}}}\kappa_{\mathbf{g}_{\mathrm{r}}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)}\cdot$$
$$\mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{r}}\hat{\mathbf{h}}_{\mathrm{r}}}\mathbf{R}^{-1}\right].\tag{55}$$

Substituting (54) and (55) into (53), we can get

$$\frac{\partial \ddot{\mathcal{R}}_{\mathrm{erg,r}}}{\partial \left\|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\right\|^2} = \frac{\log_2 e \cdot \rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}\cdot\frac{\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\kappa_{\mathbf{A}_{\mathrm{r}}}\kappa_{\mathbf{g}_{\mathrm{r}}}}{(1+\kappa_{\mathbf{A}_{\mathrm{r}}})(1+\kappa_{\mathbf{g}_{\mathrm{r}}})}}{\left(1+\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\ddot{\zeta}_{\mathrm{r}}\right)\left(1+\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{r}}\right)\ddot{\zeta}_{\mathrm{r}}\right)}\cdot$$
$$\mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{\Xi}\mathbf{R}^{-1}\right],\tag{56}$$

where

$$\mathbf{\Xi}=\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}}\left(\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}\right)\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{r}}\check{\mathbf{h}}_{\mathrm{r}}} + \rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{t}}\right)\cdot\right.$$
$$\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{t}}\check{\mathbf{h}}_{\mathrm{t}}} + \rho_{\mathrm{t}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\frac{\varrho_{\mathbf{A}_{\mathrm{t}}}\varrho_{\mathbf{g}_{\mathrm{t}}}\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)N_{\mathrm{t}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}\right)} + \varrho_{\mathbf{b}_{\mathrm{t}}}\right.$$
$$\left.+\frac{\varrho_{\mathbf{A}_{\mathrm{t}}}\varrho_{\mathbf{g}_{\mathrm{t}}}\kappa_{\mathbf{A}_{\mathrm{t}}}\kappa_{\mathbf{g}_{\mathrm{t}}}\left\|\overline{\mathbf{a}}_{\mathrm{t}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{t}}\overline{\mathbf{g}}_{\mathrm{t}}\right\|^2}{\left(1+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}\right)}\right)\mathbf{I}_M + \sigma_w^2\mathbf{I}_M\right)$$
$$+\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\frac{\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)N_{\mathrm{r}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)} + \varrho_{\mathbf{b}_{\mathrm{r}}}\right)\cdot$$
$$\left(\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}} - \mathbf{I}_M\right).\tag{57}$$

Since $\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{r}}\check{\mathbf{h}}_{\mathrm{r}}}\succ\mathbf{0}$, $\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{t}}\check{\mathbf{h}}_{\mathrm{t}}}\succ\mathbf{0}$, $\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}}\succeq\mathbf{0}$ and $\mathbf{R}\succ\mathbf{0}$, we can write:

$$\mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{\Xi}\mathbf{R}^{-1}\right] > \rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\frac{\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)N_{\mathrm{r}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)} + \varrho_{\mathbf{b}_{\mathrm{r}}}\right)\cdot$$
$$\mathrm{Tr}\left[\mathbf{R}^{-1}\left(\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}} - \mathbf{I}_M\right)\mathbf{R}^{-1}\right].\tag{58}$$

Furthermore, according to (52), we have

$$\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{r}}\right)\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{r}}\check{\mathbf{h}}_{\mathrm{r}}} + \rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\left(1-\varepsilon_{u,\mathrm{t}}\right)\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{t}}\check{\mathbf{h}}_{\mathrm{t}}} + \sigma_w^2\mathbf{I}_M$$
$$=\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}} - K\rho_{\mathrm{r}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}}\left(\rho_{\mathrm{r}}\left(1+\left(K-1\right)\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{r}}\right)\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}}\right.$$
$$\left.+\sigma_w^2\mathbf{I}_M\right)^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}} + \mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{h}_{\mathrm{t}}} - K\rho_{\mathrm{t}}\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{t}}\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{h}_{\mathrm{t}}}\cdot$$
$$\left(\rho_{\mathrm{t}}\left(1+\left(K-1\right)\varepsilon_{\mathbf{v}}\varepsilon_{u,\mathrm{t}}\right)\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{h}_{\mathrm{t}}} + \sigma_w^2\mathbf{I}_M\right)^{-1}\mathbf{C}_{\mathbf{h}_{\mathrm{t}}\mathbf{h}_{\mathrm{t}}} + \sigma_w^2\mathbf{I}_M$$
$$\overset{(a)}{\approx}\sigma_w^2\mathbf{I}_M,\tag{59}$$

where (a) is based on $\frac{\varepsilon_{u,\mathrm{r}}}{1-\varepsilon_{u,\mathrm{r}}}\gg 1$, $\frac{\varepsilon_{u,\mathrm{t}}}{1-\varepsilon_{u,\mathrm{t}}}\gg 1$ and $\frac{\varepsilon_{\mathbf{v}}}{1-\varepsilon_{\mathbf{v}}}\gg 1$.

Therefore, we can write:

$$\mathbf{R}=\left(\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\frac{\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)N_{\mathrm{r}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)} + \varrho_{\mathbf{b}_{\mathrm{r}}}\right.\right.$$
$$\left.+\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\frac{\kappa_{\mathbf{A}_{\mathrm{r}}}\kappa_{\mathbf{g}_{\mathrm{r}}}\left\|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\right\|^2}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)}\right)$$
$$+\rho_{\mathrm{t}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\frac{\varrho_{\mathbf{A}_{\mathrm{t}}}\varrho_{\mathbf{g}_{\mathrm{t}}}\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)N_{\mathrm{t}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}\right)} + \varrho_{\mathbf{b}_{\mathrm{t}}}\right.$$
$$\left.\left.+\varrho_{\mathbf{A}_{\mathrm{t}}}\varrho_{\mathbf{g}_{\mathrm{t}}}\frac{\kappa_{\mathbf{A}_{\mathrm{t}}}\kappa_{\mathbf{g}_{\mathrm{t}}}\left\|\overline{\mathbf{a}}_{\mathrm{t}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{t}}\overline{\mathbf{g}}_{\mathrm{t}}\right\|^2}{\left(1+\kappa_{\mathbf{A}_{\mathrm{t}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{t}}}\right)}\right) + \sigma_w^2\right)\mathbf{I}_M,\tag{60}$$

which is a diagonal matrix. According to (58), we have:

$$\mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{\Xi}\mathbf{R}^{-1}\right] > \rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\left(\frac{\varrho_{\mathbf{A}_{\mathrm{r}}}\varrho_{\mathbf{g}_{\mathrm{r}}}\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)N_{\mathrm{r}}}{\left(1+\kappa_{\mathbf{A}_{\mathrm{r}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{r}}}\right)} + \varrho_{\mathbf{b}_{\mathrm{r}}}\right)\cdot$$
$$\mathrm{Tr}[\mathbf{R}^{-2}]\cdot\mathrm{Tr}\left[\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{AP})\mathrm{H}} - \mathbf{I}_M\right]$$
$$=0.\tag{61}$$

Based on (56) and (61), we can get $\frac{\partial \ddot{\mathcal{R}}_{\mathrm{erg,r}}}{\partial \|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\|^2} > 0$. Similarly, we can get $\frac{\partial \ddot{\mathcal{R}}_{\mathrm{erg,t}}}{\partial \|\overline{\mathbf{a}}_{\mathrm{t}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{t}}\overline{\mathbf{g}}_{\mathrm{t}}\|^2} > 0$. Therefore, $\ddot{\mathcal{R}}_{\mathrm{erg,r}}$ and $\ddot{\mathcal{R}}_{\mathrm{erg,t}}$ are monotonically increasing functions with respect to $\left\|\overline{\mathbf{a}}_{\mathrm{r}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{r}}\overline{\mathbf{g}}_{\mathrm{r}}\right\|^2$ and $\left\|\overline{\mathbf{a}}_{\mathrm{t}}^{(\mathrm{IOS})\mathrm{H}}\mathbf{\Theta}_{\mathrm{t}}\overline{\mathbf{g}}_{\mathrm{t}}\right\|^2$, respectively. Hence, problem (P1) is equivalent to problem (P2).

APPENDIX F
PROOF OF THEOREM 5

In (36), $\ddot{\zeta}_{\mathrm{i}}$ is given by

$$\ddot{\zeta}_{\mathrm{i}} = \frac{\varrho_{\mathbf{A}_{\mathrm{i}}}\varrho_{\mathbf{g}_{\mathrm{i}}}\kappa_{\mathbf{A}_{\mathrm{i}}}\kappa_{\mathbf{g}_{\mathrm{i}}}N_{\mathrm{i}}^2}{\left(1+\kappa_{\mathbf{A}_{\mathrm{i}}}\right)\left(1+\kappa_{\mathbf{g}_{\mathrm{i}}}\right)}\overline{\mathbf{a}}_{\mathrm{i}}^{(\mathrm{AP})\mathrm{H}}\mathbf{R}^{-1}\overline{\mathbf{a}}_{\mathrm{i}}^{(\mathrm{AP})} + \mathrm{Tr}\left[\mathbf{R}^{-1}\mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{i}}\hat{\mathbf{h}}_{\mathrm{i}}}\right]$$
$$\overset{(a)}{\leq}\frac{\frac{\varrho_{\mathbf{A}_{\mathrm{i}}}\varrho_{\mathbf{g}_{\mathrm{i}}}\kappa_{\mathbf{A}_{\mathrm{i}}}\kappa_{\mathbf{g}_{\mathrm{i}}}N_{\mathrm{i}}^2}{(1+\kappa_{\mathbf{A}_{\mathrm{i}}})(1+\kappa_{\mathbf{g}_{\mathrm{i}}})}\mathrm{Tr}\left[\overline{\mathbf{a}}_{\mathrm{i}}^{(\mathrm{AP})}\overline{\mathbf{a}}_{\mathrm{i}}^{(\mathrm{AP})}\right] + \mathrm{Tr}\left[\mathbf{C}_{\mathbf{h}_{\mathrm{i}}\mathbf{h}_{\mathrm{i}}}\right]}{\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\eta_{\mathrm{r}} + \rho_{\mathrm{t}}\left(1-\varepsilon_{\mathbf{v}}\right)\eta_{\mathrm{t}} + \sigma_w^2}$$
$$\overset{(b)}{=}\frac{M\eta_{\mathrm{i}}}{\rho_{\mathrm{r}}\left(1-\varepsilon_{\mathbf{v}}\right)\eta_{\mathrm{r}} + \rho_{\mathrm{t}}\left(1-\varepsilon_{\mathbf{v}}\right)\eta_{\mathrm{t}} + \sigma_w^2},\tag{62}$$

where (a) is based on (60) as well as on $\mathbf{C}_{\mathbf{h}_{\mathrm{i}}\mathbf{h}_{\mathrm{i}}} = \mathbf{C}_{\hat{\mathbf{h}}_{\mathrm{i}}\hat{\mathbf{h}}_{\mathrm{i}}} + \mathbf{C}_{\check{\mathbf{h}}_{\mathrm{i}}\check{\mathbf{h}}_{\mathrm{i}}}$, $\mathbf{C}_{\check{\mathbf{h}}_{\mathrm{i}}\check{\mathbf{h}}_{\mathrm{i}}}\succ\mathbf{0}$ and $\mathbf{R}\succ\mathbf{0}$, while (b) is based on (11). Upon substituting (62) into (31), we arrive at (37).

REFERENCES

[1] Y. Li, F. Wang, M. El-Hajjar, and L. Hanzo, "Analog radio-over-fiber-aided optical-domain MIMO signal processing for high-performance low-cost radio access networks," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 126–132, 2021.

[2] E. Basar and H. V. Poor, "Present and future of reconfigurable intelligent surface-empowered communications [perspectives]," *IEEE Sign. Proces. Mag.*, vol. 38, no. 6, pp. 146–152, 2021.

[3] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst *et al.*, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, 2021.

[4] Q. Li, M. El-Hajjar, I. Hemadeh, D. Jagyasi, A. Shojaeifard, E. Basar, and L. Hanzo, "The reconfigurable intelligent surface-aided multi-node IoT downlink: Beamforming design and performance analysis," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 6400–6414, 2022.

[5] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. Di Renzo, A. L. Swindlehurst, R. Zhang *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Top. Sign. Proces.*, vol. 16, no. 5, pp. 883–917, 2022.

[6] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, A. A. Mourad, and L. Hanzo, "Reconfigurable intelligent surface aided amplitude-and phase-modulated downlink transmission," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8146–8151, 2023.

[7] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, 2019.

[8] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surv. & Tut.*, vol. 22, no. 4, pp. 2283–2314, 2020.

[9] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.

[10] E. Björnson, H. Wymeersch, B. Matthiesen, P. Popovski, L. Sanguinetti, and E. de Carvalho, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *IEEE Sign. Proces. Mag.*, vol. 39, no. 2, pp. 135–158, 2022.

[11] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surv. & Tut.*, vol. 24, no. 2, pp. 1035–1071, 2022.

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

[13] Q. Li, M. El-Hajjar, Y. Sun, I. Hemadeh, A. Shojaeifard, Y. Liu, and L. Hanzo, "Achievable rate analysis of the STAR-RIS aided NOMA uplink in the face of imperfect CSI and hardware impairments," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 6100–6114, 2023.

[14] Y. Jiang, F. Gao, M. Jian, S. Zhang, and W. Zhang, "Reconfigurable intelligent surface for near field communications: Beamforming and sensing," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3447–3459, 2023.

[15] H. Zhang, N. Shlezinger, G. C. Alexandropoulos, A. Shultzman, I. Alamzadeh, M. F. Imani, and Y. C. Eldar, "Channel estimation with hybrid reconfigurable intelligent metasurfaces," *IEEE Trans. Commun.*, vol. 71, no. 4, pp. 2441–2456, 2023.

[16] Q. Li, M. El-Hajjar, I. Hemadeh, A. Shojaeifard, and L. Hanzo, "Low-overhead channel estimation for RIS-aided multi-cell networks in the presence of phase quantization errors," *IEEE Trans. Veh. Technol.*, 2023, Early Access.

[17] K. Xie, G. Cai, G. Kaddoum, and J. He, "Performance analysis and resource allocation of STAR-RIS aided wireless-powered NOMA system," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5740–5755, 2023.

[18] Q. Li, M. El-Hajjar, I. Hemadeh, D. Jagyasi, A. Shojaeifard, and L. Hanzo, "Performance analysis of active RIS-aided systems in the face of imperfect CSI and phase shift noise," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8140–8145, 2023.

[19] L. Yang, Y. Yang, D. B. da Costa, and I. Trigui, "Outage probability and capacity scaling law of multiple RIS-aided networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 256–260, 2020.

[20] L. Yang, W. Zhang, P. S. Bithas, H. Liu, M. O. Hasna, T. A. Tsiftsis, and D. W. K. Ng, "Covert transmission and secrecy analysis of RS-RIS-NOMA-aided 6G wireless communication systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10 659–10 670, 2023.

[21] P. Chen, L. Yang, H. Liu, G. Pan, Y. Li, and Z. Yan, "Active RIS-NOMA-aided covert communication with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 12, no. 12, pp. 2278–2282, 2023.

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

[23] Z. Li, M. Hua, Q. Wang, and Q. Song, "Weighted sum-rate maximization for multi-IRS aided cooperative transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1620–1624, 2020.

[24] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, 2019.

[25] J. Xu, Y. Liu, X. Mu, J. T. Zhou, L. Song, H. V. Poor, and L. Hanzo, "Simultaneously transmitting and reflecting intelligent omni-surfaces: Modeling and implementation," *IEEE Veh. Technol. Mag.*, vol. 17, no. 2, pp. 46–54, 2022.

[26] H. Zhang, S. Zeng, B. Di, Y. Tan, M. Di Renzo, M. Debbah, Z. Han, H. V. Poor, and L. Song, "Intelligent omni-surfaces for full-dimensional wireless communications: Principles, technology, and implementation," *IEEE Commun. Mag.*, vol. 60, no. 2, pp. 39–45, 2022.

[27] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, 2021.

[28] H. Niu, Z. Chu, F. Zhou, P. Xiao, and N. Al-Dhahir, "Weighted sum rate optimization for STAR-RIS-assisted MIMO system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 2122–2127, 2021.

[29] S. Zhang, H. Zhang, B. Di, Y. Tan, Z. Han, and L. Song, "Beyond intelligent reflecting surfaces: Reflective-transmissive metasurface aided communications for full-dimensional coverage extension," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 905–13 909, 2020.

[30] S. Zhang, H. Zhang, B. Di, Y. Tan, M. Di Renzo, Z. Han, H. V. Poor, and L. Song, "Intelligent omni-surfaces: Ubiquitous wireless transmission by reflective-refractive metasurfaces," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 219–233, 2021.

[31] S. Dhok and P. K. Sharma, "Rate-splitting multiple access with STAR RIS over spatially-correlated channels," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6410–6424, 2022.

[32] C. Wu, Y. Liu, X. Mu, X. Gu, and O. A. Dobre, "Coverage characterization of STAR-RIS networks: NOMA and OMA," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3036–3040, 2021.

[33] C. Wu, X. Mu, Y. Liu, X. Gu, and X. Wang, "Resource allocation in STAR-RIS-aided networks: OMA and NOMA," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7653–7667, 2022.

[34] B. Zhao, C. Zhang, W. Yi, and Y. Liu, "Ergodic rate analysis of STAR-RIS aided NOMA systems," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2297–2301, 2022.

[35] A. Papazafeiropoulos, Z. Abdullah, P. Kourtessis, S. Kisseleff, and I. Krikidis, "Coverage probability of STAR-RIS assisted massive MIMO systems with correlation and phase errors," *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1738–1742, 2022.

[36] C. Wu, C. You, Y. Liu, S. Han, and M. Di Renzo, "Two-timescale design for STAR-RIS aided NOMA systems," *IEEE Trans. Commun.*, vol. 72, no. 1, pp. 585–600, 2024.

[37] M. Aldababsa, A. Khaleel, and E. Basar, "Simultaneous transmitting and reflecting reconfigurable intelligent surfaces-empowered NOMA networks," *IEEE Syst. J.*, vol. 17, no. 4, pp. 5441–5451, 2023.

[38] Y. Liu, X. Mu, J. Xu, R. Schober, Y. Hao, H. V. Poor, and L. Hanzo, "STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, 2021.

[39] L. Yang, P. Li, F. Meng, and S. Yu, "Performance analysis of RIS-assisted UAV communication systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 9078–9082, 2022.

[40] J. An, C. Xu, L. Wang, Y. Liu, L. Gan, and L. Hanzo, "Joint training of the superimposed direct and reflected links in reconfigurable intelligent surface assisted multiuser communications," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 2, pp. 739–754, 2022.

[41] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Reconfigurable intelligent surface assisted MIMO symbiotic radio networks," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4832–4846, 2021.

[42] E. Björnson, J. Hoydis, L. Sanguinetti *et al.*, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[43] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[44] X. Kang, H. Lei, L. Yang, G. Pan, T. A. Tsiftsis, and H. Liu, "Active-RIS-aided covert communications in NOMA systems with cooperative jamming," *IEEE Trans. Veh. Technol.*, 2023.

[45] X. Zhang, *Matrix analysis and applications*. Cambridge University Press, 2017.