

Achievable Rate Analysis of the STAR-RIS Aided NOMA Uplink in the Face of Imperfect CSI and Hardware Impairments

Qingchao Li, *Graduate Student Member, IEEE*, Mohammed El-Hajjar, *Senior Member, IEEE*, Yanshi Sun, *Member, IEEE*, Ibrahim Hemadeh, *Member, IEEE*, Arman Shojaeifard, *Senior Member, IEEE*, Yuanwei Liu, *Senior Member, IEEE*, Lajos Hanzo, *Life Fellow, IEEE*

Abstract—Reconfigurable intelligent surfaces (RIS) are capable of beneficially ameliorating the propagation environment by appropriately controlling the passive reflecting elements. To extend the coverage area, the concept of simultaneous transmitting and reflecting reconfigurable intelligent surfaces (STAR-RIS) has been proposed, yielding supporting 360° coverage user equipment (UE) located on both sides of the RIS. In this paper, we theoretically formulate the ergodic sum-rate of the STAR-RIS assisted non-orthogonal multiple access (NOMA) uplink in the face of channel estimation errors and hardware impairments (HWI). Specifically, the STAR-RIS phase shift is configured based on the statistical channel state information (CSI), followed by linear minimum mean square error (LMMSE) channel estimation of the equivalent channel spanning from the UEs to the access point (AP). Afterwards, successive interference cancellation (SIC) is employed at the AP using the estimated instantaneous CSI, and we derive the theoretical ergodic sum-rate upper bound for both perfect and imperfect SIC decoding algorithm. The theoretical analysis and the simulation results show that both the channel estimation and the ergodic sum-rate have performance floor at high transmit power region caused by transceiver hardware impairments.

Index Terms—Reconfigurable intelligent surfaces (RIS), simultaneous transmitting and reflecting (STAR), non-orthogonal multiple access (NOMA), imperfect channel state information (CSI), hardware impairments (HWI).

I. INTRODUCTION

THE fifth-generation (5G) systems is being rolled out across the globe and research is well under way on massive multiple-input-multiple-output (MIMO) techniques [1], [2], millimeter wave (mmWave) communications [3] and ultra-dense networking (UDN) [4]. However, driven by the ever-

increasing thirst for increased data rate, the research turned to the exploration of radical next-generation concept [5]. Reconfigurable intelligent surfaces (RIS) are capable of enhancing the transmission reliability by harnessing passive elements for reconfiguring the phase shift of the impinging signals to construct smart radio environments [6]–[9]. However, a conventional RIS can only support the users located at the same side of the RIS as the access point (AP) heating a 180° coverage. To deal with this issue, recently the simultaneous transmitting and reflecting RIS (STAR-RIS) concept was proposed for providing 360° coverage [10], [11]. In STAR-RIS architectures, three popular STAR-RIS protocols were proposed, including the energy splitting (ES), time switching (TS) and mode switching (MS) protocols [12]. Specifically, in the ES protocol, all the STAR-RIS elements are used for transmitting and reflecting signals simultaneously by splitting the power of the impinging waves. In the TS protocol, a pair of orthogonal time slots are employed, where all the STAR-RIS elements are switched to the transmission mode in one time slot and to the reflection mode in another time slot. By contrast, in the MS protocol, the STAR-RIS elements are divided into two parts, where the STAR-RIS elements are split into transmit and reflect mode.

Since non-orthogonal multiple access (NOMA) techniques are capable of approaching the boundary of capacity region, the beamforming design and performance analysis of STAR-RIS assisted NOMA systems was documented in [13]–[26]. Wu *et al.* [13] maximized the coverage range of the STAR-RIS for both orthogonal multiple access (OMA) and NOMA systems, where the numerical results showed that the coverage range of the STAR-RIS assisted NOMA scheme is higher than that of its OMA counterpart as well as that of the conventional RIS assisted schemes. Ma *et al.* [14] formulated the transmit power minimization problem of the STAR-RIS NOMA uplink, where the transmit power of each user and the passive beamforming weights of the STAR-RIS are jointly optimized by the popular semi-definite relaxation method. A resource allocation strategy is proposed in [15] for multi-carrier STAR-RIS assisted NOMA systems by jointly optimizing the channel assignment, power allocation and passive beamforming at the STAR-RIS. In [16], Zuo *et al.* employed a two-layer iterative algorithm for maximizing the sum-rate by jointly optimizing the decoding order, power allocation, and the beamformer design at the AP and the STAR-RIS. Hou *et al.* [17] extended

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1 and EP/X01228X/1 as well as of the European Research Councils Advanced Fellow Grant QuantCom (Grant No. 789028) (*Corresponding author: Lajos Hanzo.*)

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

Yanshi Sun is with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, 230009, China. (e-mail:sys@hfut.edu.cn).

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

Yuanwei Liu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: Yuanwei.Liu@qmul.ac.uk).

the STAR-RIS assisted NOMA scheme to coordinated multi-point transmission (CoMP), where both inter-cell interference cancellation and signals enhancement are attained by the passive beamforming weights of the STAR-RIS. Aldababsa *et al.* [18] derived the bit error rate (BER) expression of STAR-RIS NOMA networks relying on successive interference cancellation (SIC) by exploiting the central limit theorem (CLT). The ergodic rate of STAR-RIS assisted NOMA systems was theoretically analysed in [19]–[21]. The closed-form expressions of the outage probability (OP) and diversity gain of STAR-RIS NOMA systems were derived in [22], [23], respectively. Additionally, Liu *et al.* [24] characterized the effective capacity of STAR-RIS assisted NOMA schemes in support of ultra-reliable low-latency communications. Wang *et al.* [25] employed the moment-matching method for deriving the outage probability of STAR-RIS assisted NOMA over spatially correlated channels. Furthermore, the moment-matching method was employed in [26] by Xie *et al.* for analyzing the coverage probability of STAR-RIS aided multicell NOMA systems.

The above treatises on the STAR-RIS aided NOMA scheme have the following limitations. Firstly, the STAR-RIS aided NOMA scheme is operated based on the assumption of perfect channel state information (CSI), ignoring the channel estimation errors. Secondly, perfect transceiver hardware is assumed, i.e. the practical signal impairments of the transceivers are ignored. However, the effect of realistic channel estimation errors and hardware impairments (HWI) may impose a performance floor at high transmit powers, both in terms of the achievable rate, outage probability and bit error ratio (BER) [27]–[31]. This means that the performance improvement of RIS-aided wireless communication systems is limited by the CSI imperfection and HWI, which cannot be compensated upon increasing the transmit power. Furthermore, the above treatises on the STAR-RIS aided NOMA scheme assume that the RIS phase shifts can be perfectly configured. However, the RIS phase noise is inevitable, which imposes a potentially significant performance erosion on RIS-aided systems [32]–[36].

The above-mentioned idealized simplifying assumptions are unrealistic in practical STAR-RIS aided systems. Hence, in order to deal with the above aspects, our contributions in this paper are as follows:

- We conceive a two-timescale beamforming design to reduce the CSI acquisition overhead of the STAR-RIS aided systems. Specifically, the passive beamforming pattern of the STAR-RIS is designed based on the statistical CSI and then based on the resultant STAR-RIS beamformer, the equivalent channels spanning from the user equipment (UE) to the AP are estimated in each channel coherence interval using the linear minimum mean square error (LMMSE) technique. Then, the information is transmitted using the estimated instantaneous UE-AP CSI.
- We theoretically derive the ergodic sum-rate of the STAR-RIS assisted NOMA uplink relying on both perfect and imperfect SIC decoding algorithms, in the face of realistic RIS phase noise, imperfect CSI and transceiver HWIs. More specifically, we exploit the statistical knowledge of

the RIS phase noise distribution, of the channel estimation error variance, and of the transceiver signal distortion distribution in our theoretical derivations.

- The theoretical analysis and the simulation results show that the ergodic sum-rate of the STAR-RIS assisted NOMA uplink is independent of the decoding order when perfect SIC is assumed. By contrast, in the imperfect SIC decoding case, the user having lower channel gain should have decoding priority to maximize the ergodic sum-rate.
- The theoretical analysis and the simulation results show that the achievable sum-rate performance degradation of the STAR-RIS assisted NOMA uplink caused by the RIS phase noise can be compensated upon increasing the transmit power or deploying more RIS elements. We also show that the CSI accuracy can be improved by increasing the pilot sequence length. However, the transceiver HWI can cause an error floor at high SNRs. Furthermore, the transceiver HWI has a non-negligible effect on the rate-fairness in the STAR-RIS aided NOMA uplink.

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

The rest of this paper is organized as follows. In Section II, we present the system model, while the associated channel estimation is described in Section III. Section IV presents the performance analysis and beamforming design of the STAR-RIS assisted NOMA uplink. Our simulation results are presented in Section V, while we conclude in Section VI.

Notations: $j = \sqrt{-1}$, while $(\cdot)^*$ and $(\cdot)^H$ represent the operations of conjugation and Hermitian transpose, respectively, $|a|$ represents the amplitude of the complex scalar a , $\mathbb{C}^{m \times n}$ denotes the space of $m \times n$ complex-valued matrices, \mathbf{I}_N represents the $N \times N$ identity matrix, $\text{diag}\{\mathbf{a}\}$ denotes a diagonal matrix with the diagonal elements being the elements of \mathbf{a} 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 $\boldsymbol{\Sigma}$, $\mathbb{E}[x]$ represents the mean of the random variable x , the covariance between the random variables x and y is denoted by C_{xy} .

II. SYSTEM MODEL

The STAR-RIS-aided wireless communication system model of [19], [22] is shown in Fig. 1, including a single-antenna AP¹, a STAR-RIS and a pair of single-antenna UEs distributed at the different sides of the STAR-RIS². We denote the UE at the different side of the AP as UE-T and the UE at the same side of the AP as UE-R. Since the direct links between the UEs and the AP are blocked, the STAR-RIS creates additional links to support information transfer. We focus our

¹To unveil the theoretically achievable rate limit of the STAR-RIS aided NOMA uplink, we assume that the AP is equipped with a single-antenna. Note that the alternating optimization (AO) algorithm is also applicable to the case of multiple antennas at the AP [37].

²To support multiple UEs at each sides of the STAR-RIS, we can employ the general hybrid time division multiple access (TDMA)-NOMA technique associated with device grouping [38]. Specifically, the UEs are divided into multiple groups, where the signals of the UEs in the same group are transmitted simultaneously via NOMA, while that in different groups occupy orthogonal time slots.

TABLE I: Novelty comparison with the literatures related to STAR-RIS.

	Our paper	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]
STAR-RIS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NOMA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Downlink		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Uplink	✓		✓												
Achievable rate	✓				✓	✓		✓	✓	✓			✓		
Channel estimation	✓														
Imperfect CSI	✓														
Transceiver hardware impairments	✓														
RIS phase noise	✓														

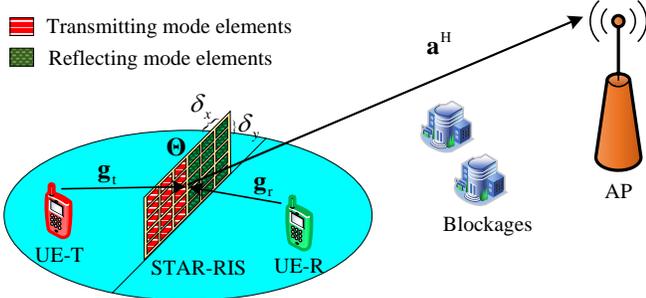


Fig. 1: System model of the STAR-RIS aided uplink NOMA wireless communication system.

attention on the power-domain NOMA uplink architecture, where the transmit power at the UE-T and UE-R are ρ_t and ρ_r , respectively. Specifically, the popular SIC algorithm is employed for information recovery in NOMA [39], where the AP first detects the signal of one user, followed by remodulating it and subtracting the interference imposed by it on the composite NOMA signal. As a result, this operation leaves behind the uncontaminated signal for another user.

A. STAR-RIS Architecture

In our work, we focus on the MS STAR-RIS architecture. Recall that for the ES protocol power splitting circuits are required and having energy leakage is inevitable, the TS protocol requires mode switching circuits for each elements. Hence, the MS protocol is deemed simplest implementation [12]. We assume that a total of $N_t + N_r$ STAR-RIS elements are used in a uniform rectangular planar array (URPA) containing $N_t = N_{t,x} \times N_{t,y}$ elements operating in the transmit mode to support the UE-T, and a URPA containing $N_r = N_{r,x} \times N_{r,y}$ element in the reflecting mode to support the UE-R. These two types of RIS elements are placed block-wise, as shown in Fig. 1. Furthermore, the distances between the adjacent STAR-RIS elements in the horizontal and vertical direction are represented by δ_x and δ_y , respectively. We denote the phase shift of the n_t th element in the transmit mode as $\theta_{n_t}^t$ and that of the n_r th element in the reflect mode as $\theta_{n_r}^r$, where we have $0 \leq \theta_{n_t}^t < 2\pi$ for $n_t = 1, 2, \dots, N_t$,

and $0 \leq \theta_{n_r}^r < 2\pi$ for $n_r = 1, 2, \dots, N_r$. Therefore, the phase shift matrix of the STAR-RIS can be characterized as $\Theta = \text{diag}\{\Theta^t, \Theta^r\}$, with $\Theta^t = \text{diag}\{e^{j\theta_1^t}, e^{j\theta_2^t}, \dots, e^{j\theta_{N_t}^t}\}$ and $\Theta^r = \text{diag}\{e^{j\theta_1^r}, e^{j\theta_2^r}, \dots, e^{j\theta_{N_r}^r}\}$.

In [13]–[26] it was assumed that the phase shift can be perfectly configured without phase noise. However, due to the realistic RIS hardware impairments, the phase shift of each reflecting element is practically modelled as [32]

$$\theta_{n_i}^i = \bar{\theta}_{n_i}^i + \tilde{\theta}_{n_i}^i, \quad i \in \{t, r\}, \quad (1)$$

where $\bar{\theta}_{n_i}^i$ represents the expected phase shift, while $\tilde{\theta}_{n_i}^i$ is the phase noise of the n_i -th element. The phase noise $\tilde{\theta}_{n_i}^i$ is modelled by identically and independently distributed (i.i.d.) random variables having a mean of 0, and following either the von Mises distribution or the uniform distribution [32]. These may be represented as $\tilde{\theta}_{n_i}^i \sim \mathcal{VM}(0, \varsigma_p)$ and $\tilde{\theta}_{n_i}^i \sim \mathcal{UF}(-\iota_p, \iota_p)$, where ς_p is the concentration parameter of the von Mises distributed variables and $(-\iota_p, \iota_p)$ is the support interval of the uniformly distributed variables. We denote the RIS phase noise power as σ_p^2 . Hence, $\varsigma_p = \frac{1}{\sigma_p^2}$ and $\iota_p = \sqrt{3\sigma_p^2}$.

B. Channel Model

As shown in Fig. 1, we denote the links spanning from the UE-T to the STAR-RIS transmitting elements as $\mathbf{g}_t \in \mathbb{C}^{N_t \times 1}$, and the channel from the UE-R to the STAR-RIS reflecting elements as $\mathbf{g}_r \in \mathbb{C}^{N_r \times 1}$. The channel impinging from the STAR-RIS upon the AP is denoted as $\mathbf{a}^H = [\mathbf{a}_t^H, \mathbf{a}_r^H]$, where $\mathbf{a}_t^H \in \mathbb{C}^{1 \times N_r}$ and $\mathbf{a}_r^H \in \mathbb{C}^{1 \times N_t}$ are the links spanning from the transmitting elements to the AP and those from the reflecting elements to the AP, respectively. Then, \mathbf{g}_t , \mathbf{g}_r , \mathbf{a}_t and \mathbf{a}_r are assumed to obey Rician fading, given by [17], [19]

$$\mathbf{g}_i \sim \mathcal{CN}\left(\frac{\sqrt{\kappa_i} \bar{\mathbf{g}}_i}{\sqrt{1 + \kappa_i}}, \frac{\mathbf{I}_{N_i}}{1 + \kappa_i}\right), \quad (2)$$

$$\mathbf{a}_i \sim \mathcal{CN}\left(\frac{\sqrt{\kappa_a} \bar{\mathbf{a}}_i}{\sqrt{1 + \kappa_a}}, \frac{\mathbf{I}_{N_i}}{1 + \kappa_a}\right), \quad (3)$$

where $i \in \{t, r\}$, κ_t , κ_r and κ_a denote the Rician factors of the corresponding link, $\bar{\mathbf{g}}_t$, $\bar{\mathbf{g}}_r$, $\bar{\mathbf{a}}_t$ and $\bar{\mathbf{a}}_r$ represent the LoS

component vectors, given by

$$\bar{\mathbf{g}}_i = [1, \dots, e^{-j\frac{2\pi}{\lambda}(\delta_x n_{i,x} \sin \phi_i \cos \varphi_i + \delta_y n_{i,y} \cos \phi_i)}, \dots, e^{-j\frac{2\pi}{\lambda}(\delta_x (N_{i,x}-1) \sin \phi_i \cos \varphi_i + \delta_y (N_{i,y}-1) \cos \phi_i)}]_{\text{H}}, \quad (4)$$

$$\bar{\mathbf{a}}_i = [1, \dots, e^{-j\frac{2\pi}{\lambda}(\delta_x n_{i,x} \sin \omega_i \cos \varpi_i + \delta_y n_{i,y} \cos \omega_i)}, \dots, e^{-j\frac{2\pi}{\lambda}(\delta_x (N_{i,x}-1) \sin \omega_i \cos \varpi_i + \delta_y (N_{i,y}-1) \cos \omega_i)}]_{\text{H}}, \quad (5)$$

where $i \in \{t, r\}$, λ is the wavelength [17], [19]. ϕ_t and φ_t are the elevation and azimuth angle of arrival (AoA) to the STAR-RIS transmitting elements, respectively, while ϕ_r and φ_r are the elevation and azimuth AoA to the STAR-RIS reflecting elements, respectively. Furthermore, ω_t and ϖ_t are the elevation and azimuth angle of departure (AoD) from the STAR-RIS transmitting elements, respectively, while ω_r and ϖ_r are the elevation and azimuth AoD from the reflecting STAR-RIS elements, respectively.

We denote the path loss of the signals from the UE-T to the STAR-RIS and that from the UE-R to the STAR-RIS by ϱ_t and ϱ_r , respectively. While ϱ_a denotes the path loss of the signals from the STAR-RIS to the AP. The distance-dependent path loss model is employed [7], which is given by $\varrho_t = \varrho_0 d_t^{-\alpha_t}$, $\varrho_r = \varrho_0 d_r^{-\alpha_r}$ and $\varrho_a = \varrho_0 d_a^{-\alpha_a}$, where d_t , d_r and d_a represent the length of the corresponding links, α_t , α_r and α_a represent the path loss exponent of the corresponding links, and ϱ_0 is the path loss at the reference distance of 1 meter. The equivalent channel spanning from the UE-T to the AP is the function of Θ^t , given by $h_t(\Theta^t) = \sqrt{\varrho_t \varrho_a} \mathbf{a}_t^{\text{H}} \Theta^t \mathbf{g}_t$. Similarly, the equivalent channel departing from the UE-R to the AP is the function of Θ^r , given by $h_r(\Theta^r) = \sqrt{\varrho_r \varrho_a} \mathbf{a}_r^{\text{H}} \Theta^r \mathbf{g}_r$.

III. CHANNEL ESTIMATION

Channel estimation is essential for RIS-aided wireless communication systems. To mitigate the CSI acquisition overhead, we employ the two-timescale beamforming protocol of [40]. Specifically, the passive beamforming at the STAR-RIS is designed based on the statistical CSI, i.e. $\bar{\mathbf{g}}_t$, $\bar{\mathbf{g}}_r$, $\bar{\mathbf{a}}_t$ and $\bar{\mathbf{a}}_r$, which remain constant for numerous coherence intervals. Based on the passive beamforming at the RIS, the equivalent channels, i.e. $h_t(\Theta^t)$ and $h_r(\Theta^r)$, can be estimated for each coherence time. Based on the estimated equivalent channels, the SIC detection may be used at the AP. The two-timescale channel estimation and information transfer protocol conceived is illustrated in Fig. 2. In each quasi-stationary block, the statistical CSI is fixed, which is estimated at the beginning of the block. The rest of the quasi-stationary block includes Q coherence intervals. Since the statistical CSI changes slowly, the pilot overhead of the statistical CSI acquisition remains modest [41]. Each coherence interval is composed of several symbol slots, which are assumed to have the same instantaneous CSI. The STAR-RIS beamforming is designed based on the statistical CSI. Thus, the STAR-RIS phase shift is fixed within each quasi-stationary block as seen in Fig. 2, and the equivalent instantaneous UE-AP channel is estimated at the beginning of each coherence interval. Then, the information delivery and recovery are performed based on the corresponding instantaneous UE-AP channel.

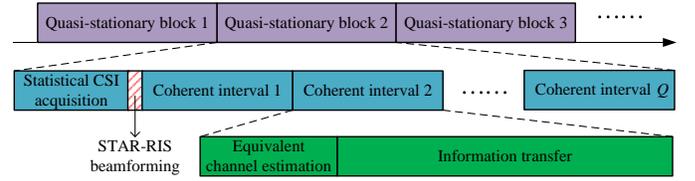


Fig. 2: Illustration of the two-timescale channel estimation and information transfer protocol.

To realize the LMMSE estimator, both the mean and the second moment of the equivalent channels are required.

Theorem 1. Based on a specific STAR-RIS phase shift, the mean and the second moment of the equivalent channels $h_t(\Theta^t)$ and $h_r(\Theta^r)$ are given by

$$\mathbb{E}[h_t(\Theta^t)] = \sqrt{\frac{\varrho_t \varrho_a \kappa_t \kappa_a}{(1 + \kappa_t)(1 + \kappa_a)}} \xi \sum_{n=1}^{N_t} \bar{a}_{t,n}^* \bar{g}_{t,n} e^{j\bar{\theta}_n^t}, \quad (6)$$

$$\mathbb{E}[h_r(\Theta^r)] = \sqrt{\frac{\varrho_r \varrho_a \kappa_r \kappa_a}{(1 + \kappa_r)(1 + \kappa_a)}} \xi \sum_{n=1}^{N_r} \bar{a}_{r,n}^* \bar{g}_{r,n} e^{j\bar{\theta}_n^r}, \quad (7)$$

$$\mathbb{E}[|h_t(\Theta^t)|^2] = \frac{\varrho_t \varrho_a}{(1 + \kappa_t)(1 + \kappa_a)} \left(\kappa_t \kappa_a \xi^2 \left| \sum_{n=1}^{N_t} \bar{a}_{t,n}^* \bar{g}_{t,n} e^{j\bar{\theta}_n^t} \right|^2 + (\kappa_t + \kappa_a + 1) N_t \right), \quad (8)$$

$$\mathbb{E}[|h_r(\Theta^r)|^2] = \frac{\varrho_r \varrho_a}{(1 + \kappa_r)(1 + \kappa_a)} \left(\kappa_r \kappa_a \xi^2 \left| \sum_{n=1}^{N_r} \bar{a}_{r,n}^* \bar{g}_{r,n} e^{j\bar{\theta}_n^r} \right|^2 + (\kappa_r + \kappa_a + 1) N_r \right), \quad (9)$$

where $\xi = \frac{I_1(\varrho_p)}{I_0(\varrho_p)}$ when $\tilde{\theta}_{n_i}^i \sim \mathcal{VM}(0, \varrho_p)$ or $\xi = \frac{\sin(\varrho_p)}{\varrho_p}$ when $\tilde{\theta}_{n_i}^i \sim \mathcal{UF}(-\varrho_p, \varrho_p)$, with $I_m(\cdot)$ representing the modified Bessel functions of the first kind of order m . Furthermore, the variance of the equivalent channels $h_t(\Theta^t)$ and $h_r(\Theta^r)$ are

$$C_{h_t(\Theta^t)h_t(\Theta^t)} = \frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)}, \quad (10)$$

$$C_{h_r(\Theta^r)h_r(\Theta^r)} = \frac{\varrho_r \varrho_a (\kappa_r + \kappa_a + 1) N_r}{(1 + \kappa_r)(1 + \kappa_a)}. \quad (11)$$

Proof: See Appendix A. ■

To eliminate the pilot contamination between the UE-T and UE-R, a pair of orthogonal pilot sequences of length $K \geq 2$ are deployed, denoted as $[\tau_t^{(1)}; \tau_t^{(2)}; \dots; \tau_t^{(K)}]$ and $[\tau_r^{(1)}; \tau_r^{(2)}; \dots; \tau_r^{(K)}]$. In the k th ($k = 1, 2, \dots, K$) symbol slot, the pilot $\tau_t^{(k)}$ at the UE-T and the pilot $\tau_r^{(k)}$ at the UE-R are transmitted simultaneously, leading to the signal received at the AP formulated as

$$\begin{aligned} & \mathbf{y}^{(k)}(\Theta) \\ &= \underbrace{\sqrt{P_t \varepsilon_v \varepsilon_{u_t}} h_t(\Theta^t) \tau_t^{(k)}}_{\text{Desired UE-T pilot}} + \underbrace{\sqrt{P_r \varepsilon_v \varepsilon_{u_r}} h_r(\Theta^r) \tau_r^{(k)}}_{\text{Desired UE-R pilot}} \end{aligned}$$

$$\begin{aligned}
& + \underbrace{\sqrt{P_t(1-\varepsilon_v)}h_t(\Theta^t)v_t^{(k)} + \sqrt{P_r(1-\varepsilon_v)}h_r(\Theta^r)v_r^{(k)}}_{\text{HWI distortion at AP}} \\
& + \underbrace{\sqrt{P_t\varepsilon_v(1-\varepsilon_{u_t})}h_t(\Theta^t)u_t^{(k)}}_{\text{HWI distortion at UE-T}} \\
& + \underbrace{\sqrt{P_r\varepsilon_v(1-\varepsilon_{u_r})}h_r(\Theta^r)u_r^{(k)}}_{\text{HWI distortion at UE-R}} + \underbrace{w^{(k)}}_{\text{Noise}}, \quad (12)
\end{aligned}$$

where P_t and P_r are the power of pilot sequences at UE-T and UE-R respectively, ε_{u_t} , ε_{u_r} and ε_v represent the hardware quality factor of the UE-T, UE-R and AP, respectively, satisfying $0 \leq \varepsilon_v \leq 1$, $0 \leq \varepsilon_{u_t} \leq 1$ and $0 \leq \varepsilon_{u_r} \leq 1$ [42]. Explicitly, a hardware quality factor of 1 indicates that the hardware is ideal, while 0 means the hardware is completely inadequate, where $u_t^{(k)} \sim \mathcal{CN}(0, 1)$, $u_r^{(k)} \sim \mathcal{CN}(0, 1)$, $v_t^{(k)} \sim \mathcal{CN}(0, 1)$, and $v_r^{(k)} \sim \mathcal{CN}(0, 1)$ are the HWIs of the k th pilot. Furthermore, $w^{(k)} \sim \mathcal{CN}(0, \sigma_w^2)$ is the additive noise at the k th symbol interval with σ_w^2 being the power spectral density.

Firstly, to estimate $h_t(\Theta^t)$, the conjugate transpose of the UE-T pilot sequence, i.e. $[\tau_t^{(1)*}, \tau_t^{(2)*}, \dots, \tau_t^{(K)*}]$, is employed to combine the AP observations $x^{(1)}(\Theta), x^{(2)}(\Theta), \dots, x^{(K)}(\Theta)$. Then we arrive at the processed received observation formulated as

$$\begin{aligned}
x_t(\Theta) &= \frac{1}{\sqrt{K}} \sum_{k=1}^K x^{(k)}(\Theta) \tau_t^{(k)*} \\
&= \sqrt{K P_t \varepsilon_v \varepsilon_{u_t}} h_t(\Theta^t) + \sqrt{P_t \varepsilon_v (1 - \varepsilon_{u_t})} h_t(\Theta^t) u'_t \\
&\quad + \sqrt{P_r \varepsilon_v (1 - \varepsilon_{u_r})} h_r(\Theta^r) u'_r \\
&\quad + \sqrt{P_t (1 - \varepsilon_v)} h_t(\Theta^t) v'_t + \sqrt{P_r (1 - \varepsilon_v)} h_r(\Theta^r) v'_r \\
&\quad + w', \quad (13)
\end{aligned}$$

where $u'_t = \frac{1}{\sqrt{K}} \sum_{k=1}^K u_t^{(k)} \tau_t^{(k)*}$, $u'_r = \frac{1}{\sqrt{K}} \sum_{k=1}^K u_r^{(k)} \tau_t^{(k)*}$, $v'_t = \frac{1}{\sqrt{K}} \sum_{k=1}^K v_t^{(k)} \tau_t^{(k)*}$, $v'_r = \frac{1}{\sqrt{K}} \sum_{k=1}^K v_r^{(k)} \tau_t^{(k)*}$ and $w' = \frac{1}{\sqrt{K}} \sum_{k=1}^K w^{(k)} \tau_t^{(k)*}$. Due to the independence of $u_t^{(k)}$, $u_r^{(k)}$, $v_t^{(k)}$, $v_r^{(k)}$ and $w^{(k)}$ for $k = 1, 2, \dots, K$, it satisfies $u'_t \sim \mathcal{CN}(0, 1)$, $u'_r \sim \mathcal{CN}(0, 1)$, $v'_t \sim \mathcal{CN}(0, 1)$, $v'_r \sim \mathcal{CN}(0, 1)$ and $w' \sim \mathcal{CN}(0, \sigma_w^2)$.

Based on (13), we can express the mean of $x_t(\Theta)$, the covariance of $h_t(\Theta^t)$ and $x_t(\Theta)$, and the variance of $x_t(\Theta)$ as

$$\begin{aligned}
& \mathbb{E}[x_t(\Theta)] \\
&= \sqrt{K P_t \varepsilon_v \varepsilon_{u_t}} \mathbb{E}[h_t(\Theta^t)] \\
&= \sqrt{K P_t \varepsilon_v \varepsilon_{u_t}} \sqrt{\frac{\varrho_t \varrho_a \kappa_t \kappa_a}{(1 + \kappa_t)(1 + \kappa_a)}} \xi \sum_{n=1}^{N_t} \bar{a}_{i,n}^* \bar{g}_{i,n} e^{j\theta_n^t}, \quad (14)
\end{aligned}$$

$$\begin{aligned}
C_{h_t(\Theta^t)x_t(\Theta)} &= \sqrt{K P_t \varepsilon_v \varepsilon_{u_t}} C_{h_t(\Theta^t)h_t(\Theta^t)} \\
&= \sqrt{K P_t \varepsilon_v \varepsilon_{u_t}} \frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)}, \quad (15)
\end{aligned}$$

$$\begin{aligned}
C_{x_t(\Theta)x_t(\Theta)} &= K P_t \varepsilon_v \varepsilon_{u_t} C_{h_t(\Theta^t)h_t(\Theta^t)} + P_t \varepsilon_v (1 - \varepsilon_{u_t}) \mathbb{E}[|h_t(\Theta^t)|^2] \\
&\quad + P_r \varepsilon_v (1 - \varepsilon_{u_r}) \mathbb{E}[|h_r(\Theta^r)|^2] + P_t (1 - \varepsilon_v) \mathbb{E}[|h_t(\Theta^t)|^2] \\
&\quad + P_r (1 - \varepsilon_v) \mathbb{E}[|h_r(\Theta^r)|^2] + \sigma_w^2
\end{aligned}$$

$$= K P_t \varepsilon_v \varepsilon_{u_t} \frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)} + \zeta + \sigma_w^2. \quad (16)$$

where ζ is given by

$$\begin{aligned}
\zeta &= \sum_{i \in \{t, r\}} \frac{P_i (1 - \varepsilon_v \varepsilon_{u_i}) \varrho_i \varrho_a}{(1 + \kappa_i)(1 + \kappa_a)} (\kappa_i \kappa_a \xi^2 \left| \sum_{n=1}^{N_i} \bar{a}_{i,n}^* \bar{g}_{i,n} e^{j\theta_n^i} \right|^2 \\
&\quad + (\kappa_i + \kappa_a + 1) N_i). \quad (17)
\end{aligned}$$

According to (6), (13), (14), (15) and (16), the LMMSE estimator of $h_t(\Theta^t)$, denoted as $\hat{h}_t(\Theta^t)$, is given by

$$\begin{aligned}
\hat{h}_t(\Theta^t) &= \mathbb{E}[h_t(\Theta^t)] + C_{h_t(\Theta^t)x_t(\Theta)} C_{x_t(\Theta)x_t(\Theta)}^{-1} \\
&\quad (x_t(\Theta) - \mathbb{E}[x_t(\Theta)]), \quad (18)
\end{aligned}$$

and its corresponding variance is

$$\begin{aligned}
C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^t)} &= C_{h_t(\Theta^t)x_t(\Theta)} C_{x_t(\Theta)x_t(\Theta)}^{-1} C_{h_t(\Theta^t)x_t(\Theta)}^H \\
&= \frac{K P_t \varepsilon_v \varepsilon_{u_t} \left(\frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)} \right)^2}{K P_t \varepsilon_v \varepsilon_{u_t} \frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)} + \zeta + \sigma_w^2}. \quad (19)
\end{aligned}$$

Thus, the estimation error is $\check{h}_t(\Theta^t) = h_t(\Theta^t) - \hat{h}_t(\Theta^t)$, and the estimation error variance is

$$\begin{aligned}
C_{\check{h}_t(\Theta^t)\check{h}_t(\Theta^t)} &= C_{h_t(\Theta^t)h_t(\Theta^t)} - C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^t)} \\
&= \frac{\frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)} (\zeta + \sigma_w^2)}{K P_t \varepsilon_v \varepsilon_{u_t} \frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)} + \zeta + \sigma_w^2}. \quad (20)
\end{aligned}$$

Similarly, to estimate $h_r(\Theta^r)$, the conjugate transpose of the UE-R pilot sequence, i.e. $[\tau_r^{(1)*}, \tau_r^{(2)*}, \dots, \tau_r^{(K)*}]$, is employed for combining the AP observations $x^{(1)}(\Theta), x^{(2)}(\Theta), \dots, x^{(K)}(\Theta)$, and then the LMMSE estimator of $h_r(\Theta^r)$, denoted as $\hat{h}_r(\Theta^r)$, is given by

$$\begin{aligned}
\hat{h}_r(\Theta^r) &= \mathbb{E}[h_r(\Theta^r)] + C_{h_r(\Theta^r)x_r(\Theta)} C_{x_r(\Theta)x_r(\Theta)}^{-1} \\
&\quad (x_r(\Theta) - \mathbb{E}[x_r(\Theta)]), \quad (21)
\end{aligned}$$

and the corresponding estimation variance becomes

$$\begin{aligned}
C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^r)} &= C_{h_r(\Theta^r)x_r(\Theta)} C_{x_r(\Theta)x_r(\Theta)}^{-1} C_{h_r(\Theta^r)x_r(\Theta)}^H \\
&= \frac{K P_r \varepsilon_v \varepsilon_{u_r} \left(\frac{\varrho_r \varrho_a (\kappa_r + \kappa_a + 1) N_r}{(1 + \kappa_r)(1 + \kappa_a)} \right)^2}{K P_r \varepsilon_v \varepsilon_{u_r} \frac{\varrho_r \varrho_a (\kappa_r + \kappa_a + 1) N_r}{(1 + \kappa_r)(1 + \kappa_a)} + \zeta + \sigma_w^2}. \quad (22)
\end{aligned}$$

Thus, the estimation error is $\check{h}_r(\Theta^r) = h_r(\Theta^r) - \hat{h}_r(\Theta^r)$, and the estimation error variance is

$$\begin{aligned}
C_{\check{h}_r(\Theta^r)\check{h}_r(\Theta^r)} &= C_{h_r(\Theta^r)h_r(\Theta^r)} - C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^r)} \\
&= \frac{\frac{\varrho_r \varrho_a (\kappa_r + \kappa_a + 1) N_r}{(1 + \kappa_r)(1 + \kappa_a)} (\zeta + \sigma_w^2)}{K P_r \varepsilon_v \varepsilon_{u_r} \frac{\varrho_r \varrho_a (\kappa_r + \kappa_a + 1) N_r}{(1 + \kappa_r)(1 + \kappa_a)} + \zeta + \sigma_w^2}. \quad (23)
\end{aligned}$$

Therefore, the normalized mean square error (N-MSE) of the estimated channel links, denoted as $\Xi_{\text{LMMSE}}(\Theta)$, is given by

$$\Xi_{\text{LMMSE}}(\Theta) = \frac{1}{2} \left(\frac{C_{\check{h}_t(\Theta^t)\check{h}_t(\Theta^t)}}{C_{h_t(\Theta^t)h_t(\Theta^t)}} + \frac{C_{\check{h}_r(\Theta^r)\check{h}_r(\Theta^r)}}{C_{h_r(\Theta^r)h_r(\Theta^r)}} \right). \quad (24)$$

Based on (10), (11), (20), (23) and (24), when the power of the pilot sequences obeys $P_t \rightarrow \infty$ and $P_r \rightarrow \infty$, the N-MSE

tends to a constant, denoted as $\ddot{\Xi}_{\text{LMMSE}}(\Theta)$, which is given by

$$\ddot{\Xi}_{\text{LMMSE}}(\Theta) = \frac{1}{2} \left(\frac{\zeta}{K P_t \varepsilon_v \varepsilon_{u_t} \frac{\rho_t \rho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)} + \zeta} + \frac{\zeta}{K P_r \varepsilon_v \varepsilon_{u_r} \frac{\rho_r \rho_a (\kappa_r + \kappa_a + 1) N_r}{(1 + \kappa_r)(1 + \kappa_a)} + \zeta} \right). \quad (25)$$

It shows that when the transceiver hardware is ideal, i.e. $\varepsilon_v = \varepsilon_{u_t} = \varepsilon_{u_r} = 1$, the N-MSE tends to 0 when $P_t = P_r \rightarrow \infty$. By contrast, when the transceiver hardware is non-ideal, i.e. $\varepsilon_v \varepsilon_{u_t} \varepsilon_{u_r} < 1$, the N-MSE has a performance floor as $P_t = P_r \rightarrow \infty$, given by

$$\ddot{\Xi}_{\text{LMMSE}}(\Theta) \rightarrow \frac{1}{2} \left(\frac{\zeta'}{K \varepsilon_v \varepsilon_{u_t} \frac{\rho_t \rho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)} + \zeta'} + \frac{\zeta'}{K \varepsilon_v \varepsilon_{u_r} \frac{\rho_r \rho_a (\kappa_r + \kappa_a + 1) N_r}{(1 + \kappa_r)(1 + \kappa_a)} + \zeta'} \right), \quad (26)$$

$P_t = P_r \rightarrow \infty,$

where ζ' is

$$\zeta' = \sum_{i \in \{t, r\}} \frac{(1 - \varepsilon_v \varepsilon_{u_i}) \rho_i \rho_a}{(1 + \kappa_i)(1 + \kappa_a)} \left(\kappa_i \kappa_a \zeta^2 \left| \sum_{n=1}^{N_i} \bar{a}_{i,n}^* \bar{g}_{i,n} e^{j\bar{\theta}_i^n} \right|^2 + (\kappa_i + \kappa_a + 1) N_i \right). \quad (27)$$

Furthermore, when the pilot sequence length is large enough, i.e. $K \gg 1$, the performance floor of the N-MSE is given by

$$\ddot{\Xi}_{\text{LMMSE}}(\Theta) \rightarrow \frac{1}{2} \left(\frac{\zeta'}{K \varepsilon_v \varepsilon_{u_t} \frac{\rho_t \rho_a (\kappa_t + \kappa_a + 1) N_t}{1 + \kappa_t}} + \frac{\zeta'}{K \varepsilon_v \varepsilon_{u_r} \frac{\rho_r \rho_a (\kappa_r + \kappa_a + 1) N_r}{1 + \kappa_r}} \right), \quad (28)$$

$P_t = P_r \rightarrow \infty, K \gg 1.$

When the statistical information of the links, i.e. $\mathbb{E}[h_t(\Theta^t)]$, $\mathbb{E}[x_t(\Theta)]$, $C_{h_t(\Theta^t)x_t(\Theta)}$, $C_{x_t(\Theta)x_t(\Theta)}$, $\mathbb{E}[h_r(\Theta^r)]$, $\mathbb{E}[x_r(\Theta)]$, $C_{h_r(\Theta^r)x_r(\Theta)}$ and $C_{x_r(\Theta)x_r(\Theta)}$, is not available, the least square (LS) estimator can be employed. Specifically, the LS estimators of $h_t(\Theta^t)$ and $h_r(\Theta^r)$ are given by $\frac{x_t(\Theta)}{\sqrt{K P_t \varepsilon_v \varepsilon_{u_t}}}$ and $\frac{x_r(\Theta)}{\sqrt{K P_r \varepsilon_v \varepsilon_{u_r}}}$, respectively.

IV. PERFORMANCE ANALYSIS AND BEAMFORMING DESIGN

Based on the estimated CSI, we derive the ergodic sum-rate of the STAR-RIS aided NOMA systems with perfect SIC decoding algorithms considering both the channel estimation error and the HWI of the transceivers, compared to that of the OMA systems. Furthermore, we also present the optimal phase shift design of the STAR-RIS elements for maximizing the ergodic sum-rate. Finally, we discuss the effect of the imperfect SIC decoding algorithms.

A. Ergodic Sum-rate Analysis

We denote the information symbols at the UE-T and UE-R as $s_t \in \mathbb{C}^{1 \times 1}$ and $s_r \in \mathbb{C}^{1 \times 1}$ respectively, with a mean of 0

and variance of 1. The observation $y(\Theta)$ at the AP is given by

$$y(\Theta) = \underbrace{\sqrt{\rho_t \varepsilon_v \varepsilon_{u_t}} \hat{h}_t(\Theta^t) s_t + \sqrt{\rho_r \varepsilon_v \varepsilon_{u_r}} \hat{h}_r(\Theta^r) s_r}_{\text{Desired signal of UE-T and UE-R over estimated channel}} + \underbrace{\sqrt{\rho_t \varepsilon_v \varepsilon_{u_t}} \check{h}_t(\Theta^t) s_t + \sqrt{\rho_r \varepsilon_v \varepsilon_{u_r}} \check{h}_r(\Theta^r) s_r}_{\text{Desired signal of UE-T and UE-R over unknown channel}} + \underbrace{\sqrt{\rho_t (1 - \varepsilon_v)} h_t(\Theta^t) v_t + \sqrt{\rho_r (1 - \varepsilon_v)} h_r(\Theta^r) v_r}_{\text{AP HWI distortion}} + \underbrace{\sqrt{\rho_t \varepsilon_v (1 - \varepsilon_{u_t})} h_t(\Theta^t) u_t}_{\text{UE-T HWI distortion}} + \underbrace{\sqrt{\rho_r \varepsilon_v (1 - \varepsilon_{u_r})} h_r(\Theta^r) u_r}_{\text{UE-R HWI distortion}} + \underbrace{w}_{\text{Noise}}, \quad (29)$$

where $w \sim \mathcal{CN}(0, \sigma_w^2)$ is the additive noise at the AP, $u_t \sim \mathcal{CN}(0, 1)$ and $v_t \sim \mathcal{CN}(0, 1)$ represent the distortion of the information symbol s_t at the UE-T and the AP, respectively. Furthermore, $u_r \sim \mathcal{CN}(0, 1)$ and $v_r \sim \mathcal{CN}(0, 1)$ represent the distortion of the information symbol s_r at the UE-R and the AP, respectively.

When we have a perfect SIC algorithm for information recovery [39], and the instantaneous achievable rate is presented in the following.

Theorem 2. Based on the perfect SIC algorithm used for the NOMA information recovery, when the AP detects the UE-T signal first, followed by the detection of the UE-R signal after removing the interference imposed by the UE-T signal from the composite received observation, let us denote this as the $t \rightarrow r$ SIC order. The instantaneous achievable rate of the UE-T and that of the UE-R, denoted as $R_t^{t \rightarrow r}(\Theta)$ and $R_r^{t \rightarrow r}(\Theta)$ respectively, are expressed as

$$R_t^{t \rightarrow r}(\Theta) = \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \mathcal{E}} \right), \quad (30)$$

$$R_r^{t \rightarrow r}(\Theta) = \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\mathcal{E}} \right), \quad (31)$$

with

$$\mathcal{E} = \rho_t \varepsilon_t(\Theta^t) + \rho_r \varepsilon_r(\Theta^r) + \sigma_w^2, \quad (32)$$

where $\varepsilon_i(\Theta^i) = (1 - \varepsilon_v \varepsilon_{u_i}) |\hat{h}_i(\Theta^i)|^2 + C_{\check{h}_i(\Theta^i) \check{h}_i(\Theta^i)}$ represents the equivalent noise caused by the transceiver HWI and the channel estimation error. By contrast, when the AP detects the UE-R signal and then removes its effect from the composite received observation for the detection of the signal from the UE-T, let us denote this as the $r \rightarrow t$ SIC order. Then, the instantaneous achievable rate of the UE-T and that of the UE-R, denoted as $R_t^{r \rightarrow t}(\Theta)$ and $R_r^{r \rightarrow t}(\Theta)$, are given by

$$R_t^{r \rightarrow t}(\Theta) = \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\mathcal{E}} \right), \quad (33)$$

$$R_r^{r \rightarrow t}(\Theta) = \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \mathcal{E}} \right). \quad (34)$$

Proof: See Appendix B. ■

By employing the time-sharing strategy, where the time

duration of β is for the $t \rightarrow r$ SIC order and that of $(1 - \beta)$ is for the $r \rightarrow t$ SIC order, the instantaneous achievable rate of the UE-T and that of the UE-R, denoted as $R_t^{\text{NOMA}}(\Theta)$ and $R_r^{\text{NOMA}}(\Theta)$ respectively, are

$$R_t^{\text{NOMA}}(\Theta) = \beta R_t^{t \rightarrow r}(\Theta) + (1 - \beta) R_t^{r \rightarrow t}(\Theta), \quad (35)$$

$$R_r^{\text{NOMA}}(\Theta) = \beta R_r^{t \rightarrow r}(\Theta) + (1 - \beta) R_r^{r \rightarrow t}(\Theta). \quad (36)$$

According to (30), (31), (33), (34), (35) and (36), the instantaneous sum-rate of the STAR-RIS aided NOMA uplink is given by

$$\begin{aligned} & R_{\text{sum}}^{\text{NOMA}}(\Theta) \\ &= R_t^{\text{NOMA}}(\Theta) + R_r^{\text{NOMA}}(\Theta) \\ &= \beta (R_t^{t \rightarrow r}(\Theta) + R_r^{t \rightarrow r}(\Theta)) + (1 - \beta) (R_t^{r \rightarrow t}(\Theta) + R_r^{r \rightarrow t}(\Theta)) \\ &= \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\mathcal{E}} \right). \end{aligned} \quad (37)$$

This shows that the sum-rate is independent of the time-sharing factor β . However, the time-sharing factor β has an effect on the fairness of the achievable rate pair of the UE-T and UE-R.

According to (37), the ergodic sum-rate of the STAR-RIS aided uplink NOMA system, denoted as $R_{\text{sum,erg}}^{\text{NOMA}}(\Theta)$, is given by

$$\begin{aligned} & R_{\text{sum,erg}}^{\text{NOMA}}(\Theta) \\ &= \mathbb{E}[R_{\text{sum}}^{\text{NOMA}}(\Theta)] \\ &= \mathbb{E} \left[\log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\mathcal{E}} \right) \right]. \end{aligned} \quad (38)$$

Since $R_{\text{sum,erg}}^{\text{NOMA}}(\Theta)$ is a concave function with respect to $|\hat{h}_t(\Theta^t)|^2$ and $|\hat{h}_r(\Theta^r)|^2$, we can formulate the upper bound of $R_{\text{sum,erg}}^{\text{NOMA}}(\Theta)$, denoted as $\check{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta)$, as follows:

$$\begin{aligned} & \check{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta) \\ &= \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} \mathbb{E}[|\hat{h}_t(\Theta^t)|^2] + \rho_r \varepsilon_v \varepsilon_{u_r} \mathbb{E}[|\hat{h}_r(\Theta^r)|^2]}{\rho_t \check{\epsilon}_t(\Theta^t) + \rho_r \check{\epsilon}_r(\Theta^r) + \sigma_w^2} \right) \\ &= \log_2 \left(1 + \frac{1}{\rho_t \check{\epsilon}_t(\Theta^t) + \rho_r \check{\epsilon}_r(\Theta^r) + \sigma_w^2} \right. \\ & \quad \left. \left(\rho_t \varepsilon_v \varepsilon_{u_t} (\mathbb{E}[|h_t(\Theta^t)|^2] + C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^t)}) \right. \right. \\ & \quad \left. \left. + \rho_r \varepsilon_v \varepsilon_{u_r} (\mathbb{E}[|h_r(\Theta^r)|^2] + C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^r)}) \right) \right), \end{aligned} \quad (39)$$

where $\check{\epsilon}_t(\Theta^t) = (1 - \varepsilon_v \varepsilon_{u_t}) (\mathbb{E}[|h_t(\Theta^t)|^2] + C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^t)}) + C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^t)}$ and $\check{\epsilon}_r(\Theta^r) = (1 - \varepsilon_v \varepsilon_{u_r}) (\mathbb{E}[|h_r(\Theta^r)|^2] + C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^r)}) + C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^r)}$.

For comparison, we consider an OMA scheme having a fraction B of the time/frequency assigned to the UE-T and the remaining fraction $(1 - B)$ of the time/frequency assigned to the UE-R. Then, the instantaneous achievable rate of the UE-T and that of the UE-R, denoted as $R_t^{\text{OMA}}(\Theta)$ and $R_r^{\text{OMA}}(\Theta)$, are given by [43]

$$R_t^{\text{OMA}}(\Theta) = B \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\rho_t \epsilon_t(\Theta^t) + B \sigma_w^2} \right), \quad (40)$$

$$R_r^{\text{OMA}}(\Theta) = (1 - B) \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\rho_r \epsilon_r(\Theta^r) + (1 - B) \sigma_w^2} \right). \quad (41)$$

Observe from (40) and (41) that varying B from 0 to 1 finds all rate pairs that can be achieved by the OMA scheme. We denote the sum-rate of the OMA system as $R_{\text{sum}}^{\text{OMA}}(\Theta)$, which is given by

$$\begin{aligned} & R_{\text{sum}}^{\text{OMA}}(\Theta) = R_t^{\text{OMA}}(\Theta) + R_r^{\text{OMA}}(\Theta) \\ &= B \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\rho_t \epsilon_t(\Theta^t) + B \sigma_w^2} \right) \\ & \quad + (1 - B) \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\rho_r \epsilon_r(\Theta^r) + (1 - B) \sigma_w^2} \right). \end{aligned} \quad (42)$$

Theorem 3. When the number of pilots is high enough, and the UE-T as well as the UE-R have the same hardware quality factor, i.e. $K \rightarrow \infty$ and $\varepsilon_{u_t} = \varepsilon_{u_r}$, the achievable sum-rate of the OMA scheme is not higher than that of the NOMA scheme, i.e.

$$R_{\text{sum}}^{\text{NOMA}}(\Theta) \geq R_{\text{sum}}^{\text{OMA}}(\Theta), \quad (43)$$

where the equality is established only when the time/frequency fraction in the OMA scheme is $B = \frac{\rho_t |\hat{h}_t(\Theta^t)|^2}{\rho_t |\hat{h}_t(\Theta^t)|^2 + \rho_r |\hat{h}_r(\Theta^r)|^2}$.

Proof: See Appendix C. ■

B. STAR-RIS Passive Beamforming Design

Our objective is to design the STAR-RIS phase shift Θ^t and Θ^r for maximizing the ergodic sum-rate upper bound of the NOMA scheme. Thus, the optimization problem can be formulated as

$$\begin{aligned} & \text{(P1)} \quad \max_{\bar{\theta}_1^t, \bar{\theta}_2^t, \dots, \bar{\theta}_{N_t}^t, \bar{\theta}_1^r, \bar{\theta}_2^r, \dots, \bar{\theta}_{N_r}^r} \check{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta) \\ & \text{s.t.} \quad \bar{\theta}_{n_t}^t \in [0, 2\pi), \quad n_t = 1, 2, \dots, N_t, \\ & \quad \quad \bar{\theta}_{n_r}^r \in [0, 2\pi), \quad n_r = 1, 2, \dots, N_r. \end{aligned} \quad (44)$$

Since the STAR-RIS phase shift is designed based on the statistical CSI instead of the instantaneous CSI, it has considerably lower channel estimation overhead than the estimation of instantaneous CSI [40]. Thus, the number of pilot sequences K can be high enough to mitigate the channel estimation error effect on the system performance. As shown in (20) and (23), we have $C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^t)} = 0$ and $C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^r)} = 0$ when $K \rightarrow \infty$. Thus, the optimization problem (P1) can be formulated as follows.

Theorem 4. When the number of pilot sequence is large enough, i.e. $K \rightarrow \infty$, the STAR-RIS phase shifts in the problem (P1) maximizing the ergodic sum-rate upper bound of the NOMA scheme are designed as

$$\begin{aligned} \bar{\theta}_{n_t}^t &= \frac{2\pi}{\lambda} (\delta_x n_{t,x} (\sin \phi_t \cos \varphi_t - \sin \omega_t \cos \varpi_t) \\ & \quad + \delta_y n_{t,y} (\cos \phi_t - \cos \omega_t)), \end{aligned} \quad (45)$$

$$\begin{aligned} \bar{\theta}_{n_r}^r &= \frac{2\pi}{\lambda} (\delta_x n_{r,x} (\sin \phi_r \cos \varphi_r - \sin \omega_r \cos \varpi_r) \\ & \quad + \delta_y n_{r,y} (\cos \phi_r - \cos \omega_r)), \end{aligned} \quad (46)$$

based on which $\sum_{n=1}^{N_t} \bar{a}_{t,n}^* \bar{g}_{t,n} e^{j\bar{\theta}_n^t} = N_t$ in $\mathbb{E}[h_t(\Theta^t)]$, $\mathbb{E}[|h_t(\Theta^t)|^2]$, $C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^t)}$, and $C_{\hat{h}_t(\Theta^t)\hat{h}_t(\Theta^r)}$, while $\sum_{n=1}^{N_r} \bar{a}_{r,n}^* \bar{g}_{r,n} e^{j\bar{\theta}_n^r} = N_r$ in $\mathbb{E}[h_r(\Theta^r)]$, $\mathbb{E}[|h_r(\Theta^r)|^2]$, $C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^r)}$, and $C_{\hat{h}_r(\Theta^r)\hat{h}_r(\Theta^t)}$.

Proof: See Appendix D. ■

C. Effect of Imperfect SIC Decoding Algorithms

The above analysis is based on the ideal SIC decoding algorithm. However, in practical NOMA system, the interference from the user which is decoded first cannot be completely removed for the user which is decoded later. Considering the effect of imperfect SIC decoding algorithm [44], the instantaneous achievable rates of $R_t^{\rightarrow r}(\Theta)$ and $R_r^{\rightarrow t}(\Theta)$ are expressed as

$$R_t^{\rightarrow r}(\Theta) = \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \mathcal{E}} \right), \quad (47)$$

$$R_r^{\rightarrow t}(\Theta) = \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\eta \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \mathcal{E}} \right), \quad (48)$$

and the instantaneous achievable rates of $R_t^{\rightarrow t}(\Theta)$ and $R_r^{\rightarrow r}(\Theta)$ are expressed as

$$R_t^{\rightarrow t}(\Theta) = \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\eta \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \mathcal{E}} \right), \quad (49)$$

$$R_r^{\rightarrow r}(\Theta) = \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \mathcal{E}} \right), \quad (50)$$

where $\eta \in [0, 1]$ represents the SIC imperfection coefficient resulting from implementation issues such as complexity scaling and error propagation [45]. A value of $\eta = 0$ implies that the interference can be completely removed for the following user information recovery, i.e. perfect SIC.

According to (47), (48), (49) and (50), the instantaneous sum-rate of the STAR-RIS aided NOMA uplink is given by

$$\begin{aligned} & R_{\text{sum}}^{\text{NOMA}}(\Theta) \\ &= R_t^{\text{NOMA}}(\Theta) + R_r^{\text{NOMA}}(\Theta) \\ &= \beta (R_t^{\rightarrow r}(\Theta) + R_r^{\rightarrow t}(\Theta)) + (1 - \beta) (R_t^{\rightarrow t}(\Theta) + R_r^{\rightarrow r}(\Theta)) \\ &= \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\mathcal{E}} \right) \\ &\quad - \Delta R(\Theta), \end{aligned} \quad (51)$$

where $\Delta R(\Theta)$ is the sum-rate degradation imposed by the imperfect SIC algorithm compared to that of the perfect SIC algorithm in (37), given by

$$\begin{aligned} \Delta R(\Theta) &= \beta \left(\log_2 \left(1 + \frac{\eta \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\mathcal{E}} \right) \right. \\ &\quad \left. - \log_2 \left(1 + \frac{\eta \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \mathcal{E}} \right) \right) \\ &\quad + (1 - \beta) \left(\log_2 \left(1 + \frac{\eta \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\mathcal{E}} \right) \right. \\ &\quad \left. - \log_2 \left(1 + \frac{\eta \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \mathcal{E}} \right) \right). \end{aligned} \quad (52)$$

According to (49) and (50), it can be shown that in contrast to the perfect SIC algorithm, the achievable sum-rate of the imperfect SIC algorithm is determined by the decoding order fraction β .

Theorem 5. To maximize the achievable sum-rate, the decoding order fraction β is designed as follows:

$$\beta = \begin{cases} 0, & \rho_t \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 > \rho_r \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 \\ 1, & \rho_t \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 < \rho_r \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 \end{cases}. \quad (53)$$

Proof: See Appendix E. ■

V. NUMERICAL AND SIMULATION RESULTS

In this section, the theoretical and simulation results of the N-MSE and the achievable rate are presented. Unless otherwise specified, the simulation parameters are given in Table II, the transceiver hardware quality factors satisfies $\varepsilon = \varepsilon_v = \varepsilon_t = \varepsilon_r$, the number of RIS elements satisfies $N = N_t = N_r$, the pilot transmit power and the UE transmit power follow $P_t = P_r = \rho_t = \rho_r$. For channel estimation, the average received signal-to-noise-ratio (SNR) is defined as $\gamma = \frac{P_t \varrho_t \varrho_t}{\sigma_w^2} = \frac{P_r \varrho_r \varrho_r}{\sigma_w^2}$ since $P_t = P_r$ and $\varrho_t = \varrho_r$. For achievable rate, the average received SNR is defined as $\gamma = \frac{\rho_t \varrho_t \varrho_t}{\sigma_w^2} = \frac{\rho_r \varrho_r \varrho_r}{\sigma_w^2}$ since $\rho_t = \rho_r$ and $\varrho_t = \varrho_r$. We utilize lines, e.g. ‘—’ and ‘- · -’, to represent theoretical analysis, and utilize markers, e.g. ‘□’, and ‘◆’, to represent simulation results.

A. Channel Estimation Performance

Fig. 3 compares the normalized mean square error versus the average received SNR γ in the NOMA system, while employing the LMMSE estimator based on (24) and the LS estimator for different transceiver hardware quality factors ε and different pilot sequence lengths K , where the RIS phase noise follows the von Mises distribution and uniform distribution respectively, with the same power of $\sigma_p^2 = 0.1$. It shows that when the hardware quality factor $\varepsilon < 1$, the N-MSE tends to a constant value even when $\gamma \rightarrow \infty$, which

TABLE II: Simulation Parameters.

Parameters	Values
Cartesian coordinates of the AP	(0m, -80m, 20m)
Cartesian coordinates of the UE-T	(0m, 20m, 5m)
Cartesian coordinates of the UE-R	(0m, -20m, 5m)
Cartesian coordinates of the RIS	(0m, 0m, 15m)
Rician factors between the RIS and the AP	$\kappa_a = 0\text{dB}$
Rician factors between the UEs and the RIS	$\kappa_t = \kappa_r = 0\text{dB}$
Path loss at the reference distance of 1 meter	$\varrho_0 = -30\text{dB}$
Path loss exponent between the RIS and the AP	$\alpha_a = 2.4$
Path loss exponent between the UEs and the RIS	$\alpha_t = \alpha_r = 2.542$
Noise power	$\sigma_w^2 = -100\text{dBm}$
Distance between adjacent RIS elements	$\delta_x = \delta_y = \frac{\lambda}{2}$
Number of RIS elements	$N_t = N_r = 20 \times 20$

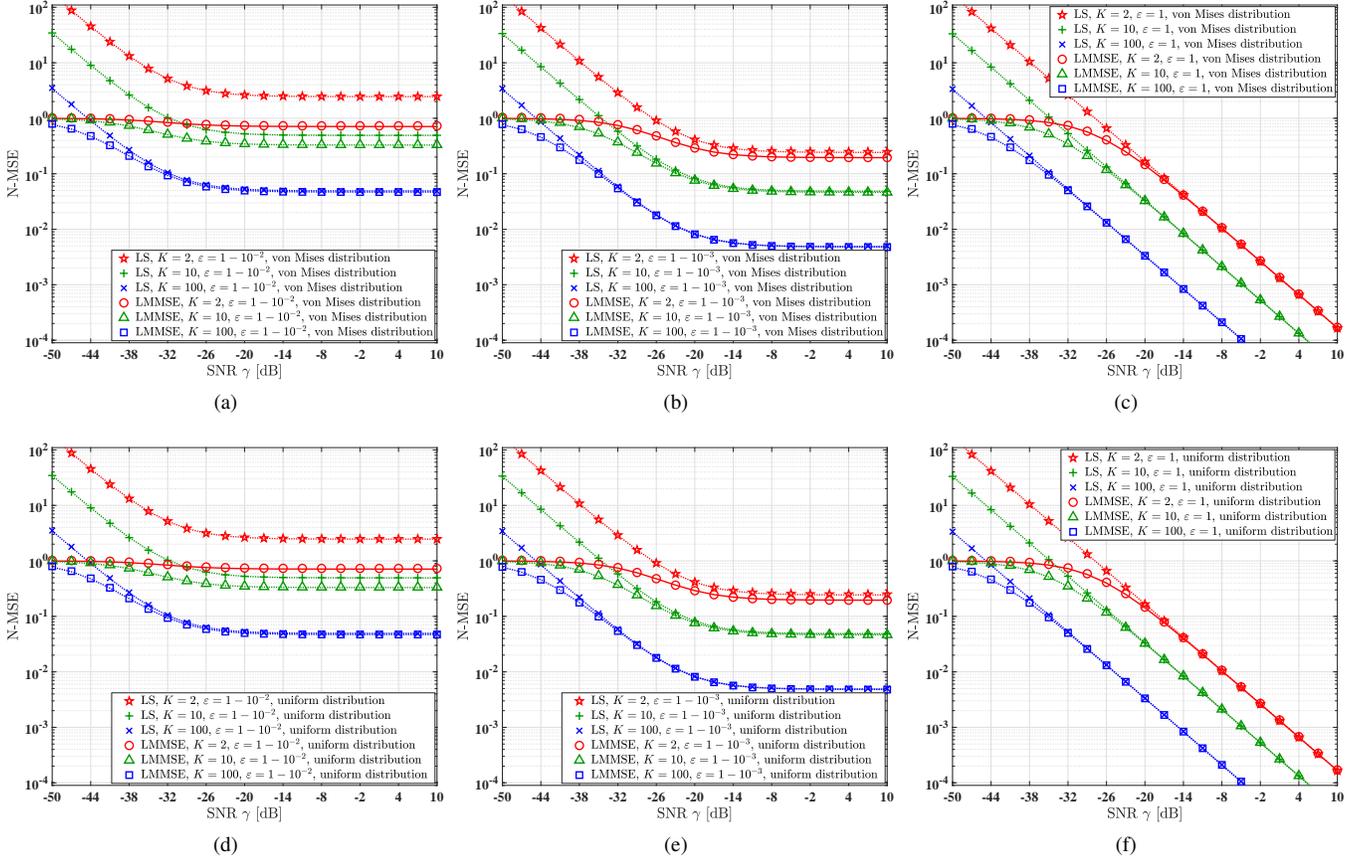


Fig. 3: Comparison of the normalized mean square error versus the average received SNR γ in the NOMA system with different transceiver hardware quality factors ε and different pilot sequence length K , where the RIS phase noise follows the von Mises distribution and uniform distribution respectively, with the same power of $\sigma_p^2 = 0.1$.

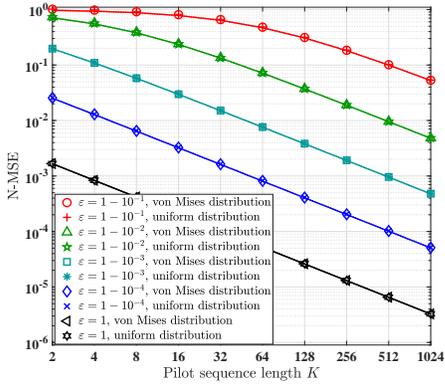


Fig. 4: Comparison of the normalized mean square error based on the LMMSE scheme versus the pilot sequence length K in the NOMA system with different transceiver hardware quality factors ε .

can be illustrated in (26). Furthermore, increasing the pilot sequence length K effectively improves the N-MSE performance. Almost the same N-MSE performance can be achieved, when the RIS phase noise follows the von Mises distribution and uniform distribution with the identical phase noise power. Furthermore, for any values of hardware quality factors, the

LMMSE estimator achieves better N-MSE performance than the LS estimator in the low SNR region due to the employment of the first-order and second-order statistical information of the links.

Fig. 4 shows the normalized mean square error performance of the STAR-RIS assisted NOMA uplink, when employing the LMMSE scheme versus the pilot sequence length K , while considering different transceiver hardware quality factors ε . In these simulation results, we consider the average received SNR $\gamma = 0$ dB, and the RIS phase noise power $\sigma_p^2 = 0.1$ following both the von Mises distribution and the uniform distribution. Our results show that the N-MSE tends to 0 with the increase of the pilot sequence length K . Furthermore, the N-MSE performance degradation caused by hardware impairments can be compensated upon increasing the pilot sequence length.

B. Ergodic Rate Performance

In Fig. 5 to 9 we present the ergodic sum-rate, including our simulation results based on (38) and the theoretical upper bound based on (39), for the NOMA system considered based on a perfect SIC decoding algorithm, where the CSI is inferred by the LMMSE estimator.

Fig. 5 shows the effect of the RIS phase noise on the ergodic sum-rate performance, when considering ideal transceiver

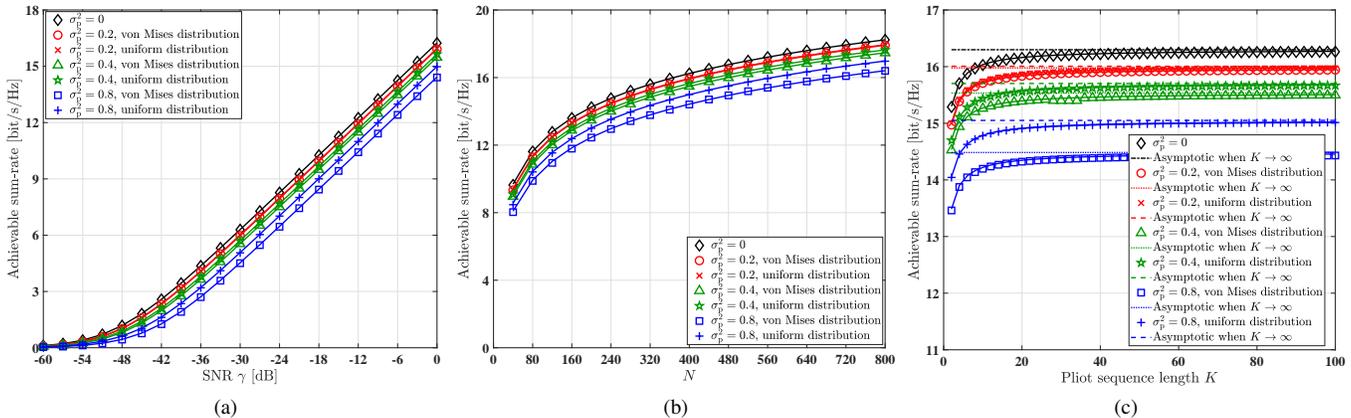


Fig. 5: Theoretical analysis and simulation results of the ergodic sum-rate in the NOMA system with different RIS phase noise power σ_p^2 .

hardware, and the RIS phase shift is optimized based on the statistical CSI according to (45) and (46). Fig. 5(a) compares the ergodic sum-rate versus the average received SNR γ in the NOMA system having different RIS phase noise powers σ_p^2 , when the pilot sequence length is $K = 50$. Fig. 5(a) shows that there is some ergodic sum-rate performance loss upon increasing the RIS phase noise power σ_p^2 . Furthermore, the von Mises-distributed RIS phase noise has higher performance degradation than its uniformly-distributed counterpart at the same phase noise power. Fig. 5(b) compares the ergodic sum-rate versus the number of RIS elements N in the NOMA system, where the pilot sequence length is $K = 50$, and the average received SNR is $\gamma = 0$ dB. It can be seen in Fig. 5(b) that the sum-rate performance degradation caused by the RIS phase noise can be compensated upon deploying more RIS elements. For example, the NOMA system with the RIS phase noise power of $\sigma_p^2 = 0.8$ following the von Mises distribution can achieve the same sum-rate performance as that having a RIS phase noise power of $\sigma_p^2 = 0$ by an approximately twice the number of RIS elements. Fig. 5(c) compares the ergodic sum-rate, as well as its asymptotic trend for $K \rightarrow \infty$, versus the pilot sequence length K in the NOMA system, with the average received SNR is $\gamma = 0$ dB. Fig. 5(c) shows that the sum-rate degradation caused by channel estimation can be eliminated upon increasing the pilot sequence length. For example, the ergodic sum-rate can approach its upper bound when the pilot sequence length is approximately 30.

Fig. 6 compares the ergodic sum-rate versus the average received SNR γ in the NOMA system for different transceiver hardware quality factors and for diverse RIS configurations, where the number of pilots is $K = 50$ and the RIS phase noise power is $\sigma_p^2 = 0$. It can be seen that upon doubling the number of RIS elements, 6dB performance improvement can be achieved when the RIS phase shift is optimally designed. This reduces to 3dB performance improvement, when the RIS phase shift is randomly designed. However, when the transceiver hardware quality factors obey $\varepsilon < 1$ and the average SNR is $\gamma \rightarrow \infty$, the same ergodic sum-rate can be achieved in the case of optimal RIS phase shift design

and random RIS phase shift design. This is seemed to be due to the fact that the achievable rate is limited by the transceiver hardware quality in the high-SNR region, which can be explained with the aid of (37).

In Fig. 7 to Fig. 9, the achievable rate performance, including the sum-rate and rate pair fairness of the NOMA system and the OMA system is compared.

Fig. 7 compares the achievable ergodic rate of UE-T R_t and that of UE-R R_r , as well as their sum-rate $R_t + R_r$, in both the NOMA and OMA system for different transceiver hardware quality factors associated with different received SNR, where the pilot sequence length is $K = 2$, the transceiver hardware quality is ideal and the RIS phase noise follows a uniform distribution having the power of $\sigma_p^2 = 0.1$. On the x-axis, the time-frequency fraction B is for the OMA system and the SIC decoding order ratio β is for the NOMA system. Specifically, in Fig. 7(a), the received SNR of UE-T is $\gamma_t = 0$ dB and that of UE-R is $\gamma_r = 0$ dB, which means that the UE-T and UE-R have the same channel gain. By contrast, in Fig. 7(b), the received SNR of UE-T is $\gamma_t = 10$ dB, while in Fig. 7(c), the received SNR of UE-T is $\gamma_t = 20$ dB and that of UE-R is $\gamma_r = -20$ dB. It can be seen in Fig. 7 that the achievable sum-rate of the NOMA system remains constant under different decoding order fraction β . By contrast, the optimal sum-rate of the OMA system is achieved when the time/frequency fraction $B = 0.5$ in the case of identical channel gain for the UE-T and the UE-R, and when $B \approx 1$ in the case of higher channel gain for the UE-T compared to the UE-R. Furthermore, Fig. 7(a) shows that when $\gamma_t = \gamma_r$, the optimal sum-rate of the OMA system is the same as that of the NOMA. By contrast, Fig. 7(b) and Fig. 7(c) show that when $\gamma_t > \gamma_r$, the optimal sum-rate of the OMA system can be slightly higher than that of the NOMA system when the time/frequency fraction is $B \approx 1$. However, in terms of the rate pair fairness, when $\gamma_t = \gamma_r$, the OMA system and the NOMA system have the same performance for the time/frequency fraction of $B = 0.5$ in the OMA system and the decoding order fraction of $\beta = 0.5$ in the NOMA system. By contrast, when $\gamma_t > \gamma_r$, the NOMA system has

better fairness than that of the OMA system. Specifically, when $\gamma_t = 10\text{dB}$ and $\gamma_r = -10\text{dB}$, the two users achieve the same rate of $R_t^{\text{OMA}} = R_r^{\text{OMA}} \approx 7.2$ bit/s/Hz when $B \approx 0.39$ in the OMA system, while they achieve the same rate of $R_t^{\text{NOMA}} = R_r^{\text{NOMA}} \approx 8.7$ bit/s/Hz when $\beta \approx 0.8$ in the NOMA system. When $\gamma_t = 20\text{dB}$ and $\gamma_r = -20\text{dB}$, the two users achieve the same rate of $R_t^{\text{OMA}} = R_r^{\text{OMA}} \approx 6.1$ bit/s/Hz when $B \approx 0.27$ in the OMA system, while the optimal rate pair in the NOMA system is achieved when $\beta = 1$ in which $R_t^{\text{NOMA}} \approx 13.3$ bit/s/Hz and $R_r^{\text{NOMA}} \approx 7.6$ bit/s/Hz.

Fig. 8 and Fig. 9 compare the achievable ergodic rate of UE-T R_t and that of UE-R R_r , as well as their sum-rate $R_t + R_r$, in both the NOMA and OMA system for different transceiver hardware qualities, while considering perfect channel knowledge. The RIS phase noise follows uniform distribution having the power of $\sigma_p^2 = 0.1$. In Fig. 8, the UE-T and the UE-R have identical channel gain, i.e. $\gamma_t = \gamma_r = 0\text{dB}$, where it can be seen that the OMA system and the NOMA system can get the same performance in terms of both the achievable sum-rate and the rate pair fairness. In Fig. 9, the UE-T has higher channel gain than the UE-R associated with $\gamma_t = 20\text{dB}$ and $\gamma_r = -20\text{dB}$. Fig. 9 shows that in terms of the achievable sum-rate, the OMA system can get the same performance as the NOMA system, when the time/frequency fraction is $B \approx 1$, which can be theoretically derived as $B = \frac{\gamma_t}{\gamma_t + \gamma_r} = \frac{10000}{10001}$, when considering all transceiver hardware qualities. By contrast, the performance of the rate pair fairness depends on the transceiver hardware qualities. Specifically, when the transceiver hardware quality is ideal, i.e. $\varepsilon = 1$, the NOMA system outperforms the OMA systems, since both users achieve the same rate of $R_t^{\text{OMA}} = R_r^{\text{OMA}} \approx 6.5$ bit/s/Hz when $B \approx 0.28$ in the OMA system, while the UE-T and the UE-R achieve the rate of $R_t^{\text{NOMA}} \approx 13.3$ bit/s/Hz and $R_r^{\text{NOMA}} \approx 8.5$ bit/s/Hz when $\beta = 1$ in the NOMA system. However, when the transceiver hardware quality is non-ideal, the OMA system achieves better rate pair fairness than the NOMA system, albeit at the cost of some sum-rate performance degradation. For example, in Fig. 9(b), when the transceiver hardware quality factor is $\varepsilon = 1 - 10^{-4}$, the OMA system has better rate pair fairness than the NOMA system, since the optimal rate pair for the OMA system is $R_t^{\text{OMA}} = R_r^{\text{OMA}} \approx 5.2$ bit/s/Hz when $B \approx 0.43$, while that for the NOMA system is $R_t^{\text{NOMA}} \approx 11.7$ bit/s/Hz and $R_r^{\text{NOMA}} \approx 0.58$ bit/s/Hz when $\beta = 1$. However, in this case, the sum-rate of the OMA system, i.e. $R_t^{\text{OMA}} + R_r^{\text{OMA}} \approx 10.4$ bit/s/Hz, is lower than that of the NOMA system, i.e. $R_t^{\text{NOMA}} + R_r^{\text{NOMA}} \approx 12.3$ bit/s/Hz.

Fig. 10 presents the simulation results of the achievable sum-rate versus the average received SNR γ in the NOMA system based on imperfect SIC decoding for different transceiver hardware quality ε and different SIC imperfection coefficients η , where the CSI is inferred by the LMMSE estimators for a pilot sequence length of $K = 2$, and the RIS phase noise power $\sigma_p^2 = 0.1$ following the uniform distribution. It shows that when the transceiver hardware quality is low, i.e. $\varepsilon = 1 - 10^{-2}$, the effect of the imperfect SIC on the sum-rate performance is limited. By contrast, when the transceiver hardware quality is high, i.e. $\varepsilon = 1 - 10^{-3}$ or $\varepsilon = 1$, imperfect SIC decoding has a substantial effect on the achievable sum

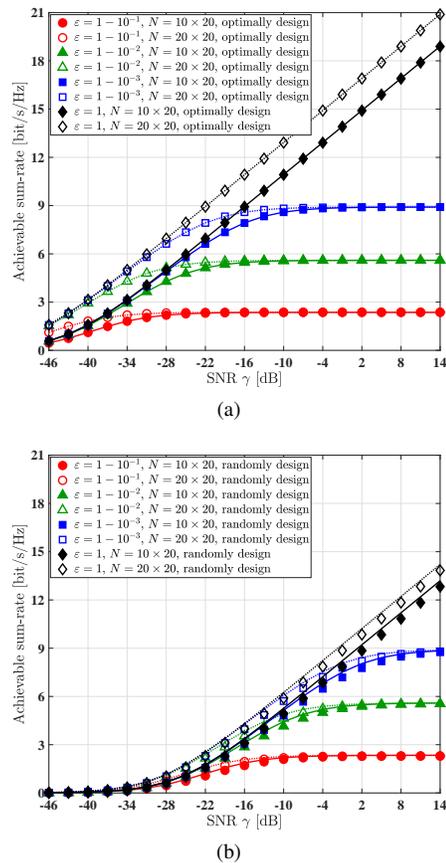


Fig. 6: Comparison of the ergodic sum-rate versus the average received SNR γ in the NOMA system with different transceiver hardware quality factors ε and different number of RIS elements with $N = N_t = N_r$, where the RIS phase shift is optimally designed and randomly designed, respectively.

rate. Furthermore, when the SIC imperfection coefficients obey $\eta > 0$, the achievable sum-rate tends to a constant value as $\gamma \rightarrow \infty$, even though the transceiver hardware is perfect.

VI. CONCLUSIONS

We theoretically analyzed the ergodic sum-rate of the STAR-RIS assisted uplink NOMA scheme relying on SIC detection in the face of imperfect CSI, as well as in the presence of RIS and transceiver HWI. Our theoretical analysis and simulation results showed that the CSI accuracy can be improved by increasing the pilot sequence length, and deploying more RIS elements can compensate the achievable sum-rate performance degradation caused the RIS phase noise. However, the transceiver HWI results in performance floor at high transmit power region in terms of both the channel estimation and the achievable sum-rate. Besides, transceiver HWIs also have significant side effect on the rate pair fairness in the STAR-RIS aided uplink NOMA systems.

APPENDIX A

PROOF OF THEOREM 1

Firstly, when the RIS phase noise obeys $\tilde{\theta}_{n_t}^t \sim \mathcal{VM}(0, \varsigma_p)$, upon referring to [46], we have $\mathbb{E}[e^{j\tilde{\theta}_{n_t}^t}] = \frac{I_1(\varsigma_p)}{I_0(\varsigma_p)}$. Secondly,

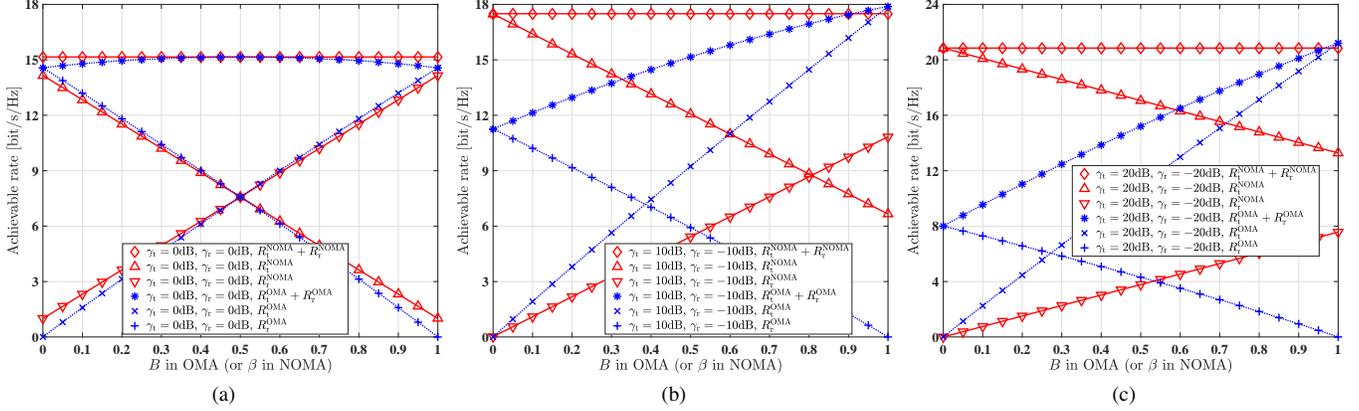


Fig. 7: Comparison of the achievable rate versus the time/frequency fraction B in the OMA scheme or the decoding order fraction β in the NOMA system with different received SNR.

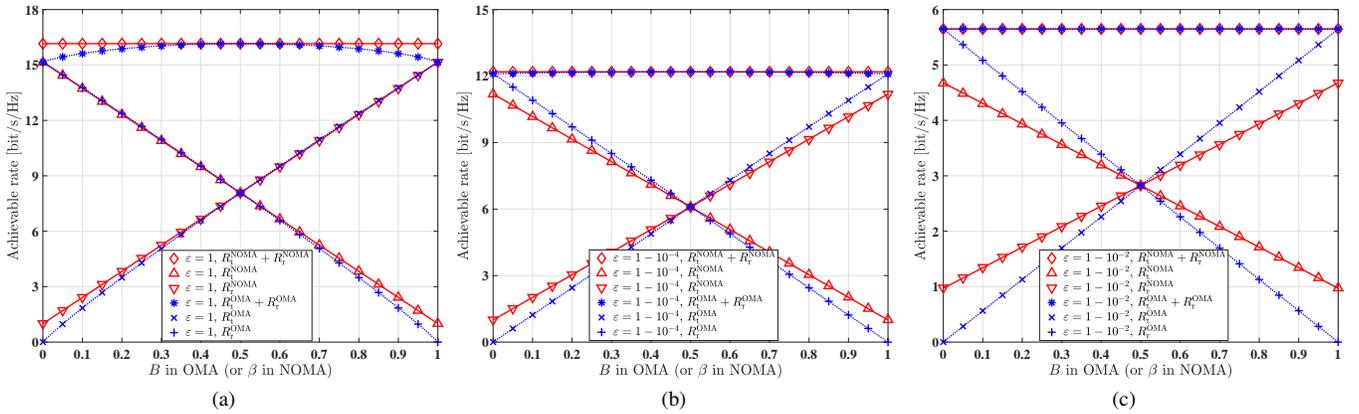


Fig. 8: Comparison of the achievable rate versus the time/frequency fraction B in the OMA scheme or the decoding order fraction β in the NOMA system with different transceiver hardware qualities, where the received SNR of UE-T and UE-R are $\gamma_t = 0\text{dB}$ and $\gamma_r = 0\text{dB}$, respectively.

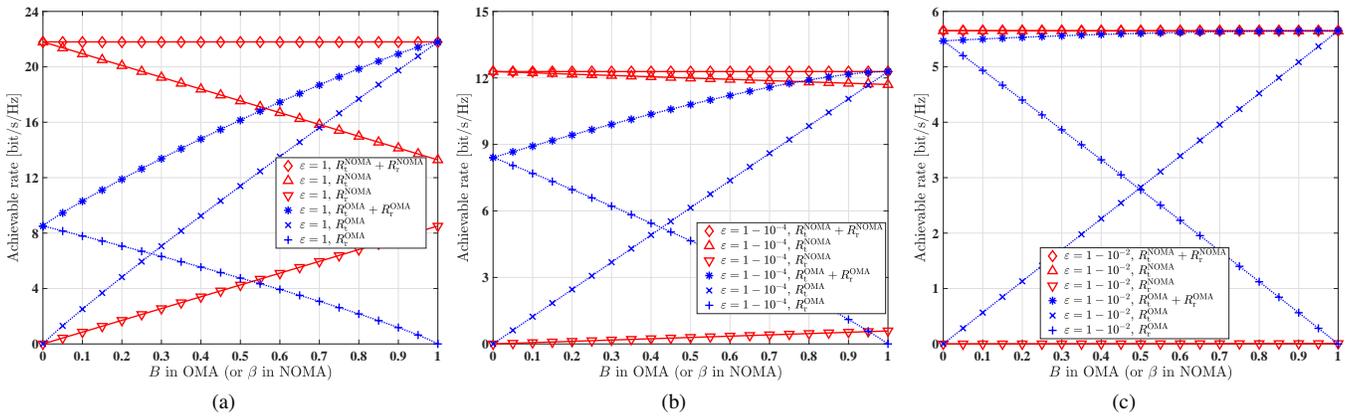


Fig. 9: Comparison of the achievable rate versus the time/frequency fraction B in the OMA scheme or the decoding order fraction β in the NOMA system with different transceiver hardware qualities, where the received SNR of UE-T and UE-R are $\gamma_t = 20\text{dB}$ and $\gamma_r = -20\text{dB}$, respectively.

when the RIS phase noise obeys $\tilde{\theta}_{n_t}^t \sim \mathcal{UF}(-\iota_p, \iota_p)$, the m th-order moment of $\tilde{\theta}_{n_t}^t$, namely $\mathbb{E}[(\tilde{\theta}_{n_t}^t)^m]$, is equal to 0 when m is odd and it is equal to $\frac{1}{m+1}\iota_p^m$ when m is even.

Thus, we can show that $\mathbb{E}[e^{j\tilde{\theta}_{n_t}^t}] = \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \mathbb{E}[(\tilde{\theta}_{n_t}^t)^{2m}] = \sum_{m=0}^{\infty} \frac{(-1)^m \iota_p^{2m}}{(2m+1)!} = \frac{\sin(\iota_p)}{\iota_p}$. Hence, we have $\mathbb{E}[e^{j\tilde{\theta}_{n_t}^t}] = \xi$,

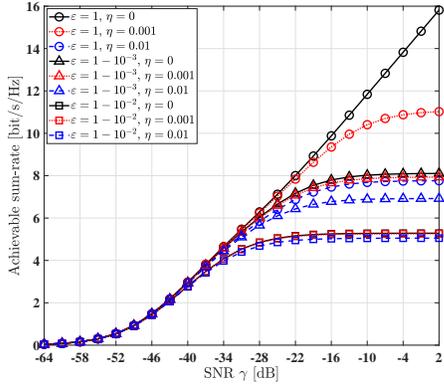


Fig. 10: Simulation results of the achievable sum-rate versus the average received SNR γ in the NOMA system based on imperfect SIC decoding algorithm with different transceiver hardware qualities ε and different SIC imperfection coefficients η .

with $\xi = \frac{I_1(c_p)}{I_0(c_p)}$ when $\tilde{\theta}_{n_t}^t \sim \mathcal{VM}(0, c_p)$ or $\xi = \frac{\sin(\ell_p)}{\ell_p}$ when $\tilde{\theta}_{n_t}^t \sim \mathcal{UF}(-\ell_p, \ell_p)$.

The mean and the second moment of the equivalent channel $h_t(\Theta^t)$ is given by

$$\begin{aligned} \mathbb{E}[h_t(\Theta^t)] &= \sqrt{\varrho_t \varrho_a} \mathbb{E}[\mathbf{a}_t^H \Theta^t \mathbf{g}_t] \\ &= \sqrt{\frac{\varrho_t \varrho_a \kappa_t \kappa_a}{(1 + \kappa_t)(1 + \kappa_a)}} \sum_{n=1}^{N_t} \mathbb{E}[\tilde{a}_{t,n}^* \tilde{g}_{t,n} e^{j(\tilde{\theta}_n^t + \tilde{\theta}_n^t)}] \\ &\stackrel{(a)}{=} \sqrt{\frac{\varrho_t \varrho_a \kappa_t \kappa_a}{(1 + \kappa_t)(1 + \kappa_a)}} \xi \sum_{n=1}^{N_t} \tilde{a}_{t,n}^* \tilde{g}_{t,n} e^{j\tilde{\theta}_n^t}, \quad (54) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[|h_t(\Theta^t)|^2] &= \varrho_t \varrho_a \mathbb{E}[|\mathbf{a}_t^H \Theta^t \mathbf{g}_t|^2] \\ &= \frac{\varrho_t \varrho_a (\kappa_t \kappa_a \xi^2 \sum_{n=1}^{N_t} |\tilde{a}_{t,n}^* \tilde{g}_{t,n} e^{j\tilde{\theta}_n^t}|^2 + (\kappa_t + \kappa_a + 1) N_t)}{(1 + \kappa_t)(1 + \kappa_a)}, \quad (55) \end{aligned}$$

where (a) is based on $\mathbb{E}[e^{j\tilde{\theta}_n^t}] = \xi$. According to (54) and (55), we can express the variance of the equivalent channels $h_t(\Theta^t)$ as

$$\begin{aligned} C_{h_t(\Theta^t)h_t(\Theta^t)} &= \mathbb{E}[|h_t(\Theta^t)|^2] - |\mathbb{E}[h_t(\Theta^t)]|^2 \\ &= \frac{\varrho_t \varrho_a (\kappa_t + \kappa_a + 1) N_t}{(1 + \kappa_t)(1 + \kappa_a)}. \quad (56) \end{aligned}$$

Similarly, we can formulate the mean, the second moment and the variance of the equivalent channel $h_r(\Theta^r)$ in (7), (9) and (11), respectively.

APPENDIX B PROOF OF THEOREM 2

Based on the SIC algorithm, in the $t \rightarrow r$ SIC order, $R_t^{\text{NOMA},t}$ and $R_r^{\text{NOMA},t}$ are given by

$$R_t(\Theta^t) = \log_2 \left(\frac{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\mathbb{E}[|y(\Theta)|^2] - \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2} \right), \quad (57)$$

$$\begin{aligned} R_r(\Theta^r) &= \log_2 \left(\frac{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\mathbb{E}[|y(\Theta)|^2] - \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 - \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2} \right), \quad (58) \end{aligned}$$

where $\mathbb{E}[|y(\Theta)|^2]$ can be derived as

$$\begin{aligned} \mathbb{E}[|y(\Theta)|^2] &= \rho_t (|\hat{h}_t(\Theta^t)|^2 + C_{\tilde{h}_t(\Theta^t)\tilde{h}_t(\Theta^t)}) \\ &\quad + \rho_r (|\hat{h}_r(\Theta^r)|^2 + C_{\tilde{h}_r(\Theta^r)\tilde{h}_r(\Theta^r)}) + \sigma_w^2. \quad (59) \end{aligned}$$

According to (57), (58), (59) and some further manipulations, we arrive at (30) and (31). In the $r \rightarrow t$ SIC order, $R_t^{r \rightarrow t}(\Theta)$ and $R_r^{r \rightarrow t}(\Theta)$ in (33) and (34) can be similarly derived.

APPENDIX C PROOF OF THEOREM 3

As shown in (20) and (23), we have $C_{\tilde{h}_t(\Theta^t)\tilde{h}_t(\Theta^t)} = 0$ and $C_{\tilde{h}_r(\Theta^r)\tilde{h}_r(\Theta^r)} = 0$ when $K \rightarrow \infty$. When the UE-T and the UE-R have the same hardware quality factor, we can set $\varepsilon_{u_t} = \varepsilon_{u_r} = \varepsilon_u$. Thus, we have $\epsilon_t(\Theta^t) = (1 - \varepsilon_v \varepsilon_u) |\hat{h}_t(\Theta^t)|^2$ and $\epsilon_r(\Theta^r) = (1 - \varepsilon_v \varepsilon_u) |\hat{h}_r(\Theta^r)|^2$. In this case, according to (42), we can get the achievable sum-rate of the OMA scheme as

$$\begin{aligned} R_{\text{sum}}^{\text{OMA}}(\Theta) &= B \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_u |\hat{h}_t(\Theta^t)|^2}{\rho_t (1 - \varepsilon_v \varepsilon_u) |\hat{h}_t(\Theta^t)|^2 + B \sigma_w^2} \right) \\ &\quad + (1 - B) \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_u |\hat{h}_r(\Theta^r)|^2}{\rho_r (1 - \varepsilon_v \varepsilon_u) |\hat{h}_r(\Theta^r)|^2 + (1 - B) \sigma_w^2} \right). \quad (60) \end{aligned}$$

To get the maximum value of $R_{\text{sum}}^{\text{OMA}}(\Theta)$, we firstly derive the partial derivative of $R_{\text{sum}}^{\text{OMA}}(\Theta)$ with respect to B as shown in (61). According to (61), we can get $\frac{\partial R_{\text{sum}}^{\text{OMA}}(\Theta)}{\partial B} = 0$ when $B = \frac{\rho_t |\hat{h}_t(\Theta^t)|^2}{\rho_t |\hat{h}_t(\Theta^t)|^2 + \rho_r |\hat{h}_r(\Theta^r)|^2}$. This means that $R_{\text{sum}}^{\text{OMA}}(\Theta)$ can be maximized when $B = \frac{\rho_t |\hat{h}_t(\Theta^t)|^2}{\rho_t |\hat{h}_t(\Theta^t)|^2 + \rho_r |\hat{h}_r(\Theta^r)|^2}$. Then, by substituting $B = \frac{\rho_t |\hat{h}_t(\Theta^t)|^2}{\rho_t |\hat{h}_t(\Theta^t)|^2 + \rho_r |\hat{h}_r(\Theta^r)|^2}$ into (60), we can get the maximum value of $R_{\text{sum}}^{\text{OMA}}(\Theta)$, denoted by $\ddot{R}_{\text{sum}}^{\text{OMA}}(\Theta)$, as

$$\begin{aligned} \ddot{R}_{\text{sum}}^{\text{OMA}}(\Theta) &= \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_u |\hat{h}_t(\Theta^t)|^2 + \rho_r \varepsilon_v \varepsilon_u |\hat{h}_r(\Theta^r)|^2}{\rho_t (1 - \varepsilon_v \varepsilon_u) |\hat{h}_t(\Theta^t)|^2 + \rho_r (1 - \varepsilon_v \varepsilon_u) |\hat{h}_r(\Theta^r)|^2 + \sigma_w^2} \right) \\ &\stackrel{(a)}{=} \log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_u |\hat{h}_t(\Theta^t)|^2 + \rho_r \varepsilon_v \varepsilon_u |\hat{h}_r(\Theta^r)|^2}{\mathcal{E}} \right) \\ &= R_{\text{sum}}^{\text{NOMA}}(\Theta), \quad (62) \end{aligned}$$

where (a) is true for $K \rightarrow \infty$ and $\varepsilon_u = \varepsilon_{u_t} = \varepsilon_{u_r}$. Based on (62), we can have $R_{\text{sum}}^{\text{NOMA}}(\Theta) \geq R_{\text{sum}}^{\text{OMA}}(\Theta)$, where the equality is established only when the time/frequency fraction in the OMA scheme is $B = \frac{\rho_t |\hat{h}_t(\Theta^t)|^2}{\rho_t |\hat{h}_t(\Theta^t)|^2 + \rho_r |\hat{h}_r(\Theta^r)|^2}$.

$$\begin{aligned} \frac{\partial R_{\text{sum}}^{\text{OMA}}(\Theta)}{\partial B} &= \log_2 e \cdot \left(\log_2 \left(1 + \frac{\rho_t \varepsilon_v \varepsilon_u |\hat{h}_t(\Theta^t)|^2}{\rho_t(1 - \varepsilon_v \varepsilon_u) |\hat{h}_t(\Theta^t)|^2 + B \sigma_w^2} \right) - \log_2 \left(1 + \frac{\rho_r \varepsilon_v \varepsilon_u |\hat{h}_r(\Theta^r)|^2}{\rho_r(1 - \varepsilon_v \varepsilon_u) |\hat{h}_r(\Theta^r)|^2 + (1 - B) \sigma_w^2} \right) \right) \\ &\quad - \log_2 e \cdot \left(\frac{\rho_t \varepsilon_v \varepsilon_u |\hat{h}_t(\Theta^t)|^2 B \sigma_w^2}{(\rho_t |\hat{h}_t(\Theta^t)|^2 + B \sigma_w^2)(\rho_t(1 - \varepsilon_v \varepsilon_u) |\hat{h}_t(\Theta^t)|^2 + B \sigma_w^2)} \right. \\ &\quad \left. - \frac{\rho_r \varepsilon_v \varepsilon_u |\hat{h}_r(\Theta^r)|^2 (1 - B) \sigma_w^2}{(\rho_r |\hat{h}_r(\Theta^r)|^2 + (1 - B) \sigma_w^2)(\rho_r(1 - \varepsilon_v \varepsilon_u) |\hat{h}_r(\Theta^r)|^2 + (1 - B) \sigma_w^2)} \right). \end{aligned} \quad (61)$$

APPENDIX D PROOF OF THEOREM 4

According to (6), (8), (19), (20), (22), (23) and (39), when $K \rightarrow \infty$, we can get

$$\begin{aligned} \ddot{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta) \\ = \log_2 \left(1 + \frac{\sum_{i \in \{t,r\}} \rho_i \varepsilon_v \varepsilon_{u_i} \mathbb{E}[|h_i(\Theta^i)|^2]}{\sum_{i \in \{t,r\}} \rho_i (1 - \varepsilon_v \varepsilon_{u_i}) \mathbb{E}[|h_i(\Theta^i)|^2] + \sigma_w^2} \right). \end{aligned} \quad (63)$$

According to (8) and (63), we can formulate the derivative of $\ddot{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta)$ with respect to $\left| \sum_{n=1}^{N_t} \bar{a}_{t,n}^* \bar{g}_{t,n} e^{j\bar{\theta}_n^t} \right|^2$ as

$$\begin{aligned} \frac{\partial \ddot{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta)}{\partial \left| \sum_{n=1}^{N_t} \bar{a}_{t,n}^* \bar{g}_{t,n} e^{j\bar{\theta}_n^t} \right|^2} &= \frac{\log_2 e \cdot \rho_t \varepsilon_v \varepsilon_{u_t} \frac{\partial \mathbb{E}[|h_t(\Theta^t)|^2]}{\partial \left| \sum_{n=1}^{N_t} \bar{a}_{t,n}^* \bar{g}_{t,n} e^{j\bar{\theta}_n^t} \right|^2}}{\sum_{i \in \{t,r\}} \rho_i (1 - \varepsilon_v \varepsilon_{u_i}) \mathbb{E}[|h_i(\Theta^i)|^2] + \sigma_w^2} \\ &> 0. \end{aligned} \quad (64)$$

Similarly, we can formulate the derivative of $\ddot{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta)$ with respect to $\left| \sum_{n=1}^{N_r} \bar{a}_{r,n}^* \bar{g}_{r,n} e^{j\bar{\theta}_n^r} \right|^2$ as

$$\begin{aligned} \frac{\partial \ddot{R}_{\text{sum,erg}}^{\text{NOMA}}(\Theta)}{\partial \left| \sum_{n=1}^{N_r} \bar{a}_{r,n}^* \bar{g}_{r,n} e^{j\bar{\theta}_n^r} \right|^2} &= \frac{\log_2 e \cdot \rho_r \varepsilon_v \varepsilon_{u_r} \frac{\partial \mathbb{E}[|h_r(\Theta^r)|^2]}{\partial \left| \sum_{n=1}^{N_r} \bar{a}_{r,n}^* \bar{g}_{r,n} e^{j\bar{\theta}_n^r} \right|^2}}{\sum_{i \in \{t,r\}} \rho_i (1 - \varepsilon_v \varepsilon_{u_i}) \mathbb{E}[|h_i(\Theta^i)|^2] + \sigma_w^2} \\ &> 0. \end{aligned} \quad (65)$$

Therefore, the optimization problem (P1) is equivalent to the following two sub-problems.

$$\begin{aligned} \text{(P2a)} \quad & \max_{\bar{\theta}_1^t, \bar{\theta}_2^t, \dots, \bar{\theta}_{N_t}^t} \mathbb{E}[|h_t(\Theta^t)|^2] \\ \text{s.t.} \quad & \bar{\theta}_{n_t}^t \in [0, 2\pi), \quad n_t = 1, 2, \dots, N_t. \end{aligned} \quad (66)$$

$$\begin{aligned} \text{(P2b)} \quad & \max_{\bar{\theta}_1^r, \bar{\theta}_2^r, \dots, \bar{\theta}_{N_r}^r} \mathbb{E}[|h_r(\Theta^r)|^2] \\ \text{s.t.} \quad & \bar{\theta}_{n_r}^r \in [0, 2\pi), \quad n_r = 1, 2, \dots, N_r. \end{aligned} \quad (67)$$

Based on (4), (5), (66) and (67), the optimal STAR-RIS phase shift is designed according to (45) and (46).

APPENDIX E PROOF OF THEOREM 5

According to (51) and (52), when $\rho_t \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 > \rho_r \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2$, we can get $\frac{\partial R_{\text{sum}}^{\text{NOMA}}(\Theta)}{\partial \beta} >$

$$\frac{\partial R_{\text{sum}}^{\text{NOMA}}(\Theta)}{\partial \beta}$$

$$\begin{aligned} &= \log_2 \left(1 + \frac{\eta \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\mathcal{E}} \right) \\ &\quad + \log_2 \left(1 + \frac{\eta \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \mathcal{E}} \right) \\ &\quad - \log_2 \left(1 + \frac{\eta \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2}{\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \mathcal{E}} \right) \\ &\quad - \log_2 \left(1 + \frac{\eta \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2}{\mathcal{E}} \right) \\ &= \log_2 \left(\eta (\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2) (\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2) \right. \\ &\quad \left. + \mathcal{E} (\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \eta \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \mathcal{E}) \right) \\ &\quad - \log_2 \left(\eta (\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2) (\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2) \right. \\ &\quad \left. + \mathcal{E} (\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \eta \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \mathcal{E}) \right) \\ &\quad + \log_2 \left(\rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \eta \rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \mathcal{E} \right) \\ &\quad - \log_2 \left(\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 + \eta \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2 + \mathcal{E} \right) \\ &\stackrel{(a)}{<} 0. \end{aligned} \quad (68)$$

where (a) is based on $\rho_t \varepsilon_v \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 > \rho_r \varepsilon_v \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2$. Upon exploiting (68), we can show that $\frac{\partial R_{\text{sum}}^{\text{NOMA}}(\Theta)}{\partial \beta} < 0$. Thus, the achievable sum-rate is maximized, when $\beta = 0$. Similarly, when $\rho_t \varepsilon_{u_t} |\hat{h}_t(\Theta^t)|^2 < \rho_r \varepsilon_{u_r} |\hat{h}_r(\Theta^r)|^2$, we can get $\frac{\partial R_{\text{sum}}^{\text{NOMA}}(\Theta)}{\partial \beta} > 0$, which means that the achievable sum-rate is maximized when $\beta = 1$.

REFERENCES

- [1] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.
- [2] P. Yang, Y. Xiao, Y. L. Guan, S. Li, and L. Hanzo, "Transmit antenna selection for multiple-input multiple-output spatial modulation systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2035–2048, 2016.
- [3] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4980–4995, 2017.
- [4] M. Jafri, A. Anand, S. Srivastava, A. K. Jagannatham, and L. Hanzo, "Robust distributed hybrid beamforming in coordinated multi-user multi-cell mmWave MIMO systems relying on imperfect CSI," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8123–8137, 2022.
- [5] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G Internet of things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, 2022.
- [6] 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.
- [7] Q. Li, M. El-Hajjar, I. A. Hemadeh, A. Shojafifard, 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.

- [8] 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, 2023.
- [9] 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.*, 2023, Early Access.
- [10] J. Zhao, Y. Zhu, X. Mu, K. Cai, Y. Liu, and L. Hanzo, "Simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) assisted UAV communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 3041–3056, 2022.
- [11] Z. Zhang, J. Chen, Y. Liu, Q. Wu, B. He, and L. Yang, "On the secrecy design of STAR-RIS assisted uplink NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 11 207–11 221, 2022.
- [12] 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.
- [13] 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.
- [14] H. Ma, H. Wang, H. Li, and Y. Feng, "Transmit power minimization for STAR-RIS-empowered uplink NOMA system," *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2430–2434, 2022.
- [15] 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.
- [16] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint design for simultaneously transmitting and reflecting (STAR) RIS assisted NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 611–626, 2022.
- [17] T. Hou, J. Wang, Y. Liu, X. Sun, A. Li, and B. Ai, "A joint design for STAR-RIS enhanced NOMA-CoMP networks: A simultaneous-signal-enhancement-and-cancellation-based (SSECB) design," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 1043–1048, 2021.
- [18] M. Aldababsa, A. Khaleel, and E. Basar, "STAR-RIS-NOMA networks: An error performance perspective," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1784–1788, 2022.
- [19] X. Yue, J. Xie, Y. Liu, Z. Han, R. Liu, and Z. Ding, "Simultaneously transmitting and reflecting reconfigurable intelligent surface assisted NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 189–204, 2023.
- [20] 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.
- [21] J. Chen and X. Yu, "Ergodic rate analysis and phase design of STAR-RIS aided NOMA with statistical CSI," *IEEE Commun. Lett.*, vol. 26, no. 12, pp. 2889–2893, 2022.
- [22] S. Yang, J. Zhang, W. Xia, H. Gao, and H. Zhu, "Joint power and discrete amplitude allocation for STAR-RIS-aided NOMA system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13 382–13 386, 2022.
- [23] C. Zhang, W. Yi, Y. Liu, Z. Ding, and L. Song, "STAR-IOA aided NOMA networks: Channel model approximation and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 6861–6876, 2022.
- [24] H. Liu, G. Li, X. Li, Y. Liu, G. Huang, and Z. Ding, "Effective capacity analysis of STAR-RIS assisted NOMA networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 9, pp. 1930–1934, 2022.
- [25] T. Wang, M.-A. Badiu, G. Chen, and J. P. Coon, "Outage probability analysis of STAR-RIS assisted NOMA network with correlated channels," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1774–1778, 2022.
- [26] Z. Xie, W. Yi, X. Wu, Y. Liu, and A. Nallanathan, "STAR-RIS aided NOMA in multi-cell networks: A general analytical framework with gamma distributed channel modeling," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5629–5644, 2022.
- [27] P. Yang, L. Yang, and S. Wang, "Performance analysis for RIS-aided wireless systems with imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 588–592, 2021.
- [28] Z. Peng, T. Li, C. Pan, H. Ren, and J. Wang, "RIS-aided D2D communications relying on statistical CSI with imperfect hardware," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 473–477, 2021.
- [29] T.-H. Vu and S. Kim, "Performance analysis of full-duplex two-way RIS-based systems with imperfect CSI and discrete phase-shift design," *IEEE Commun. Lett.*, 2022.
- [30] 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.*, 2023, Early Access.
- [31] J. Wang, S. Gong, Q. Wu, and S. Ma, "RIS-aided MIMO systems with hardware impairments: Robust beamforming design and analysis," *IEEE Trans. Wireless Commun.*, 2023, Early Access.
- [32] M.-A. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 184–188, 2019.
- [33] X. Qian, M. Di Renzo, J. Liu, A. Kammoun, and M.-S. Alouini, "Beamforming through reconfigurable intelligent surfaces in single-user MIMO systems: SNR distribution and scaling laws in the presence of channel fading and phase noise," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 77–81, 2020.
- [34] K. Zhi, C. Pan, H. Ren, and K. Wang, "Uplink achievable rate of intelligent reflecting surface-aided millimeter-wave communications with low-resolution ADC and phase noise," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 654–658, 2020.
- [35] A. Papazafeiropoulos, C. Pan, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Intelligent reflecting surface-assisted MU-MISO systems with imperfect hardware: Channel estimation and beamforming design," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 2077–2092, 2021.
- [36] J. Dai, F. Zhu, C. Pan, H. Ren, and K. Wang, "Statistical CSI-based transmission design for reconfigurable intelligent surface-aided massive MIMO systems with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 38–42, 2021.
- [37] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735–1748, 2020.
- [38] G. Chen, Q. Wu, C. He, W. Chen, J. Tang, and S. Jin, "Active IRS aided multiple access for energy-constrained IoT systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1677–1694, 2023.
- [39] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 2018.
- [40] 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. Topics Signal Process.*, vol. 16, no. 5, pp. 883–917, 2022.
- [41] K. Zhi, C. Pan, H. Ren, and K. Wang, "Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3558–3574, 2022.
- [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] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [44] N. S. Mouni, A. Kumar, and P. K. Upadhyay, "Adaptive user pairing for NOMA systems with imperfect SIC," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1547–1551, 2021.
- [45] D.-T. Do and T.-T. T. Nguyen, "Impacts of imperfect SIC and imperfect hardware in performance analysis on AF non-orthogonal multiple access network," *Telecommunication Systems*, vol. 72, no. 4, pp. 579–593, 2019.
- [46] T. Hillen, K. J. Painter, A. C. Swan, and A. D. Murtha, "Moments of von Mises and Fisher distributions and applications," *Mathematical Biosciences & Engineering*, vol. 14, no. 3, p. 673, 2017.