Two-Timescale Design for STAR-RIS-Aided NOMA Systems

Chenyu Wu, Changsheng You, Member, IEEE, Yuanwei Liu, Senior Member, IEEE, Shuai Han, Senior Member, IEEE, and Marco Di Renzo, Fellow, IEEE

Abstract—Simultaneously transmitting and reflecting reconfigurable intelligent surfaces (STAR-RISs) have emerged as a promising technology to reconfigure the radio propagation environment in the full space. Prior works on STAR-RISs have mostly considered the energy splitting operation protocol, which has high hardware complexity in practice. Moreover, the full and instantaneous channel state information (CSI) is always assumed available for designing the STAR-RIS nearly passive beamforming, which, however, is practically difficult to obtain due to the large number of STAR-RIS elements. To address these issues, we study the mode switching design in STAR-RIS aided non-orthogonal multiple access (NOMA) communication systems. Moreover, two efficient two-timescale (TTS) transmission protocols are proposed for different channel setups to maximize the respective average achievable sum-rate. Specifically, 1) for the case of line-of-sight (LoS) dominant channels, we propose the beamforming-then-estimate (BTE) protocol, where the long-term STAR-RIS transmission and reflection coefficients are optimized based on the statistical CSI only, while the short-term power allocation at the base station (BS) is designed based on the estimated effective fading channels of all the users; 2) for the case of rich scattering environments, we propose an alternative partition-then-estimate (PTE) protocol, where the BS first determines the long-term STAR-RIS surface-partition strategy based on the path-loss information only, with each subsurface being assigned to one user; and then the BS estimates the instantaneous subsurface channels associated with the users and designs its power allocation and STAR-RIS phase-shifts accordingly. For the two proposed transmission protocols, we further propose efficient algorithms to solve the respective long-term and short-term optimization problems. Moreover, we show that both proposed transmission protocols substantially reduce the channel estimation overhead as compared to the existing schemes based on full instantaneous CSI. Last, simulation results validate the superiority of our proposed transmission protocols as compared to various benchmarks. It is shown that the BTE protocol outperforms the PTE protocol when the number of STAR-RIS elements is large and/or the LoS channel components are dominant, and vice versa.

Index Terms—Reconfigurable intelligent surface (RIS), simultaneous transmission and reflection (STAR), non-orthogonal multiple access (NOMA), statistical CSI, two-timescale optimization.

I. INTRODUCTION

RECONFIGURABLE intelligent surface (RIS) [2], [3], also known as intelligent reflecting surface (IRS) [4], has been envisioned as a promising technology to enhance the spectral efficiency of wireless communication systems. Specifically, RIS is a planar array comprising a massive number of tunable and low-cost elements. By smartly controlling the amplitude/phase-shift of individual elements, RIS can program the wireless propagation environment for achieving various functions, such as enhancing the desirable signal strength [5] or suppressing the interference [6]. Nevertheless, the conventional reflecting-type RISs can only achieve half-space reflection as specified by their pointing directions, thereby limiting the network coverage. To tackle this issue, a new type of RIS, called simultaneous transmitting and reflecting reconfigurable intelligent surfaces (STAR-RISs) [7], or intelligent omni-surface (IOS) [8], has been recently proposed. In contrast to conventional RISs [2], [3], [4], [5], [6], STAR-RISs enable simultaneously reflecting and/or transmitting the signals incident on both sides of the surface, thus achieving 360° wireless coverage. Particularly, the amplitude and phase responses of the transmitted and reflected signals can be jointly and appropriately optimized, which provides an extra degree of freedom for enhancing the communication performance.

On the other hand, non-orthogonal multiple access (NOMA) is a promising technology to satisfy the stringent requirements of next-generation wireless networks such as extremely high spectral efficiency and massive connectivity [9]. Different from the conventional orthogonal multiple-access (OMA) that serves users over orthogonal time/frequency resource blocks (RBs), NOMA allows multiple users to share the...
same RB in the code or power domain. Specifically, for power-domain NOMA, the users’ signals are superposed at the transmitter and recovered at the receiver by using the successive interference cancellation (SIC) technique. Existing works have revealed that NOMA is able to achieve a much higher rate than OMA when the users’ channels are more distinctive [10]. This thus motivates the study of STAR-RIS aided NOMA systems, as STAR-RISs can flexibly manipulate the wireless channels to achieve enhanced NOMA gain.

A. Related Works

1) RIS-Aided Wireless Systems: In the existing literature, RIS has been extensively investigated in various setups, such as orthogonal frequency division multiplexing (OFDM) [11], millimeter wave (mmWave) communications [12], unmanned aerial vehicle (UAV) communications [13], NOMA transmissions [14], [15], [16], and so on. These works have mostly assumed the availability of instantaneous channel state information (CSI) for reaping the RIS passive beamforming gain. However, obtaining the full and accurate CSI is practically difficult due to the large number of reflecting elements and their nearly passive nature. As a result, various approaches have been proposed to efficiently estimate the RIS-reflected channels. For example, under the single-user setup, the authors of [17] and [18] designed efficient RIS reflection training patterns to estimate the cascaded channels under the ideal case with continuous phase-shifts and the practical case with discrete phase-shifts, respectively. Particularly, it was shown that the length of the required training sequence for channel estimation scales with the number of RIS elements [17]. For the multiuser scenario, various channel properties such as the common base station (BS) to RIS channel [19], channel spatial correlation [20], and channel sparsity [21] were leveraged for reducing the channel estimation overhead. A fast multi-beam training method for RIS-assisted mmWave system was proposed in [22]. Moreover, since the statistical CSI (S-CSI) varies slowly and hence can be more easily obtained, a novel two-timescale (TTS) beamforming design was first proposed in [23], whose main idea is to exploit the S-CSI for designing RIS reflections and the instantaneous CSI (I-CSI) for designing the BS transmit beamforming. INSPIRED by this, efficient TTS designs have been proposed in RIS-aided multi-user systems leveraging various channel statistics such as users’ locations and channel covariance matrices [24], [25], [26], [27], [28]. Despite the prominent passive beamforming gain, the performance of wireless systems aided by an RIS is fundamentally constrained by its serving region. This thus motivates the design of a new STAR-RIS that is capable of serving the users on both of its sides.

2) STAR-RIS Assisted Wireless Systems: Recently, growing research attention has been devoted to investigating the benefits of deploying STAR-RISs in wireless communication systems. Specifically, the authors of [29] presented an hardware model for STAR-RIS and its channel models in the near-field and far-field regions. Besides, in [30], three STAR-RIS operation protocols have been proposed, namely energy splitting (ES), mode switching (MS), and time switching (TS). Based on these protocols, the joint active and nearly passive beamforming design was studied to minimize the total power consumption at the access point [30]. For multiuser radio access, the integration of STAR-RISs and NOMA has drawn great attention. For example, the authors of [31] revealed that the coverage range can be extended when employing STAR-RISs in NOMA networks. Besides, a STAR-RIS-assisted multi-cluster NOMA system was studied in [32], where the decoding order, power allocation, active beamforming, and STAR-RIS coefficients were jointly optimized for maximizing the achievable sum-rate. Moreover, the authors of [33] proposed three channel models for a STAR-RIS aided NOMA network, based on which the diversity gain was analyzed. STAR-RISs have also been studied to integrate with other promising technologies such as physical layer security [34], unmanned aerial vehicle [35], etc.

B. Motivation and Contributions

Although the benefits of incorporating NOMA and STAR-RIS have been intensively investigated, there are two limitations that have not been well addressed. First, most works have considered the ES operation protocol for STAR-RIS. However, this protocol generally incurs a relatively high hardware complexity, since it necessitates the simultaneous manipulation of both the transmitted and reflected signals for each STAR-RIS element. Moreover, the ES protocol requires the joint optimization of the energy splitting ratio and transmission/reflection phase-shifts, which is generally difficult to solve. Second, existing works on STAR-RIS mainly focused on its nearly passive beamforming design for data transmission, under the assumption of full I-CSI available at the transmitter. Nevertheless, as recently reported in [36] and [37], acquiring the I-CSI associated with all STAR-RIS elements and all users usually incurs a prohibitively high channel estimation overhead, which, in general, is proportional to the number of users and STAR-RIS elements. In practical systems, a large channel training overhead may result in less time remaining for data transmission and may hence degrade the communication performance. This issue becomes more severe when the wireless channels are highly dynamic due to the short channel coherence time.

To overcome these issues, we study the MS design for a STAR-RIS aided NOMA system. Compared with ES, MS is much easier to implement, since each STAR-RIS element can be flexibly tuned to operate in either the transmission or reflection mode, thus alleviating the intrinsic coupling between them. Moreover, to reduce the channel estimation overhead for acquiring instantaneous and full CSI, we propose two efficient TTS transmission protocols for two channel setups. This work presents the first attempt to develop the TTS design for the new STAR-RIS NOMA system and makes a comprehensive study of different channel fading conditions. The main contributions of this paper are summarized as follows.

- First, we consider the channel setup where the STAR-RIS associated channels are line-of-sight (LoS) dominant. For this case, we propose an efficient TTS transmission protocol called beamforming-then-estimate (BTE). Specifically, the BS exploits the S-CSI for designing
the STAR-RIS beamforming coefficients (including the mode selection and phase-shift of each element) in the long term and keeps them fixed as long as the S-CSI remains unchanged. Meanwhile, in each time slot, the BS estimates the instantaneous effective CSI for each user and designs its short-term power allocation accordingly. To maximize the average sum-rate, we formulate an optimization problem that jointly optimizes the long-term STAR-RIS beamforming coefficients and the short-term BS power allocation. To tackle this challenging problem, we first solve the short-term optimization problem to obtain the BS power allocation based on the effective CSI. Then, an efficient penalty-based method is developed to suboptimally solve the long-term optimization problem for obtaining the STAR-RIS transmission- and reflection-coefficient vectors based on the S-CSI. Particularly, for the BTE protocol, the channel estimation overhead is equal to the number of users \( K \).

- Next, as the BTE protocol is ineffective in rich scattering environments, we propose another dedicated TTS transmission protocol for this channel setup, which is called partition-then-estimate (PTE).\(^1\) Specifically, based on the information of large-scale path-loss, the STAR-RIS is partitioned into several transmission/reflection subsurfaces, each associated with one specific user. Then, in each time slot, the BS estimates the instantaneous subsurface channels associated with each user (referred to as the partial I-CSI), based on which it jointly designs the short-term BS power allocation and STAR-RIS phase-shifts. Similarly, the short-term optimization problem is optimally solved based on the partial I-CSI. For the long-term optimization problem, on the other hand, we propose a low-complexity yet efficient algorithm to determine the proper number of elements assigned to each user. The channel estimation overhead in each time slot for this protocol is \( K + M \), where \( M \) denotes the number of STAR-RIS elements.

- Last, simulation results are presented under various system setups to demonstrate the effectiveness of our proposed TTS transmission protocols as compared to several benchmarks. It is shown that the channel estimation overheads for the two proposed protocols are much smaller than that of the full I-CSI based design, i.e. \( K + M \). Moreover, the BTE protocol is more effective when \( M \) is large and the LoS channel components are dominant, whereas the PTE protocol is preferred when \( M \) is small and/or the channel coherence time is sufficiently long.

C. Organization and Notations

The rest of this paper is organized as follows. In Section II, we present the system model for the considered STAR-RIS aided NOMA communication system. In Sections III and IV, we propose two practical TTS transmission protocols for LoS-dominant channels and rich scattering channels, respectively. Efficient algorithms for solving the corresponding average sum-rate maximization problem are also provided. Section V presents numerical results to validate the effectiveness of the proposed designs. Finally, Section VI concludes this paper.

Notations: \( \mathbb{C}^{M \times 1} \) denotes the space of \( M \times 1 \) complex-valued vectors. \( \mathbf{a}^H \) denotes the conjugate transpose of vector \( \mathbf{a} \). diag(\( \mathbf{a} \)) denotes a diagonal matrix with the diagonal elements given by \( \mathbf{a} \). diag(\( \mathbf{A} \)) denotes a vector extracted from the diagonal elements of matrix \( \mathbf{A} \). The distribution of a circularly symmetric complex Gaussian variable (CSG) with mean \( \mu \) and variance \( \sigma^2 \) is denoted by \( \mathcal{CN} \sim (\mu, \sigma^2) \). The \( m \)-th element of a vector \( \mathbf{a} \) is denoted by \( [\mathbf{a}]_m \). For a complex scalar \( x \), \( |x| \), \( \angle x \) and Re(\( x \)) denote its modulus, angle and real part, respectively. \( \triangleq \) stands for definition.

II. System Model

As depicted in Fig. 1, we consider a narrow-band wireless communication system, where a STAR-RIS comprising \( M \) elements is deployed to assist the downlink communication from a single-antenna BS to \( K \) single-antenna users, denoted by the set \( \mathcal{K} = \{1, \ldots, K\} \). The STAR-RIS divides the full space of signal propagation into two regions, namely, the transmission region and the reflection region. Accordingly, the users are divided into two groups, namely T users and R users, which are denoted by \( \mathcal{K}_t = \{1, \ldots, K_t\} \) and \( \mathcal{K}_r = \{1, \ldots, K_r\} \), respectively.

Let \( h_k \in \mathbb{C} \) denote the baseband equivalent channel from the BS to user \( k \in \mathcal{K}_s \), with \( s \in \{t, r\} \) specifying the transmission/reflection region where the user is located. Further, let \( \mathbf{r}_k^H \in \mathbb{C}^{1 \times M} \) and \( \mathbf{g} \in \mathbb{C}^{M \times 1} \) denote the channel from the STAR-RIS to user \( k \) and from the BS to the STAR-RIS, respectively. We assume the Rician fading channel model for the links associated with the STAR-RIS, i.e., \( \{\mathbf{r}_k^H\} \) and \( \mathbf{g} \); while the BS-user links \( \{h_k\} \) are modeled by Rayleigh fading. As such, the channel coefficients are given by

\[
\begin{align}
\mathbf{g} &= \sqrt{\delta_{BS}} \left( \frac{\kappa_1}{\kappa_1 + 1} \mathbf{g} + \sqrt{\frac{1}{\kappa_1 + 1}} \mathbf{g} \right), \\
\mathbf{r}_k &= \sqrt{\delta_{SK}} \left( \frac{\kappa_2}{\kappa_2 + 1} \mathbf{r}_k + \sqrt{\frac{1}{\kappa_2 + 1}} \mathbf{r}_k \right), \forall k \in \mathcal{K}, \\
h_k &= \sqrt{\delta_k h_k}, \forall k \in \mathcal{K},
\end{align}
\]

where \( \delta_{BS}, \delta_{SK}, \delta_k \) are the distance-dependent path-loss of the BS to STAR-RIS, STAR-RIS to user \( k \), BS to user

---

\(^1\)The BTE protocol has been presented in conference version of this paper [1], while a new PTE protocol is proposed here.
k links, respectively, \(\tilde{g}\) and \(\tilde{r}_k\) denote the normalized LoS components, \(g\) and \(r_k\) denote the normalized non-LoS (NLoS) components, which follow the CSCG distribution with zero mean and unit variance, and \(\kappa_1, \kappa_2\) are the Rician factors.

Remark 1: First, for the general multi-antenna BS scenario, a stochastic optimization framework can be leveraged to deal with the corresponding optimization problem [24]. This is, however, more challenging to tackle as left as our future work. Next, consider a special case where the direct links between the BS and users are blocked. In this case, zero-forcing precoding can be employed at the BS for the short-term transmit beamforming [38]. Moreover, when the channel from the BS to the STAR-RIS is LoS, the BS can align its transmit beamforming with the BS-STAR-RIS LoS channel and thus can be treated as having an equivalent single antenna. Based on the these considerations, the solutions proposed in this paper can be extended to the case with a multi-antenna BS. To derive some closed-form expressions for gaining useful insights, we consider the single-antenna BS.

A. STAR-RIS Operation Protocol: Mode Switching

We consider the mode switching operation protocol for the studied STAR-RIS aided communication system. According to [30], the STAR-RIS coefficients can be characterized by the transmission and reflection matrices \(\Theta_s = \text{diag}(\theta_s)\text{diag}(\beta_s)\), where \(\theta_s = [e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_M}]^T\), \(\beta_s = [\sqrt{\beta_1}, \ldots, \sqrt{\beta_M}]^T\), \(\beta_m \in [0, 1]\) and \(\theta_m \in [0, 2\pi]\), \(m \in M = \{1, \ldots, M\}\) denoting the operation mode and phase-shift of the \(m\)-th element, respectively. Further, we define \(v_s \triangleq \text{diag}(\Theta_s), s \in \{t, r\}\) as the transmission- and reflection-coefficient vectors. According to the law of energy conservation, we have \(\beta_m + \beta_m^* \leq 1\) or equivalently \(||v_t||^2 + ||v_r||^2\leq 1\) needs to be satisfied. Therefore, each element of STAR-RIS operates in either the full transmission or full reflection mode, depending on the value of \(\beta_m^*\). As such, the STAR-RIS-assisted effective channel between the BS and user \(k\) can be expressed as

\[
c_{k}(v_s) = h_k + r_k^H \Theta_s g = h_k + w_k^H v_s,
\]

where \(w_k^H = r_k^H \text{diag}(g)\).

It is worth noting that the MS protocol can be regarded as a special case of the ES protocol where the amplitude coefficients for transmission and reflection are set as continuous variables, i.e., \(\beta_m^* \in [0, 1]\). Compared with the ES protocol, which provides high STAR-RIS reflection/transmission design flexibility, the MS protocol is particularly attractive for practical use due to the following reasons. Firstly, the switch between the transmission and reflection modes of each element is easy to realize from the hardware implementation perspective [30]. Secondly, under the ES protocol, more STAR-RIS coefficients need to be optimized, which requires higher computational complexity than the MS protocol. Thirdly, as reported in [39], the MS protocol is not restricted by the constraint of transmission-reflection phase correlation which may cause performance degradation to the ES protocol.

B. NOMA Transmission

We consider NOMA transmission for all users. By adopting superposition coding, the received signal at user \(k\) is given by

\[
y_k = c_k(v_s) \sum_{k=1}^{K} \sqrt{p_k} s_k + n_k,
\]

where \(s_k \sim \mathcal{CN}(0, 1)\) denotes the transmitted data symbol for user \(k\), \(n_k \sim \mathcal{CN}(0, \sigma^2)\) is the additive white Gaussian noise, \(p_k\) is the transmit power allocated to user \(k\). For NOMA transmission, by invoking the SIC technique, the user with stronger channel power gain is able to first decode the signals of the other users, and then cancel their contributions before decoding its own signal. However, since the effective channel power gain of each user is determined by the STAR-RIS coefficients according to (2), the optimal decoding order has \(K!\) combinations. To reduce the computational complexity for the exhaustive search, we propose a simple yet efficient scheme for the decoding order. Specifically, we first define \(l_k \triangleq \delta_k + M^2 \delta_{BS} \delta_{SK}\), as the optimal BS-user \(k\) equivalent channel gain by aligning the STAR-RIS-associated channels with the direct link. Without loss of generality, we assume that the users are arranged in descending order according to \(l_k\), i.e., \(l_1 > l_2 > \ldots > l_K\). Then, we assume that the NOMA decoding order also follows this permutation, i.e., user 1 is always the strongest user whose signals are last decoded. As such, the achievable rate of user \(k\) in bits/second/Hz (bps/Hz) is given by

\[
R_k = \log_2 \left( 1 + \frac{p_k |c_k(v_s)|^2}{\sum_{j<k} p_j |c_j(v_s)|^2 + \sigma^2} \right).
\]

Remark 2: In practice, the instantaneous channel gain of different users may not follow this order due to the channel fading. However, it is worth pointing out that once the STAR-RIS coefficients are fixed in the long-term (which is introduced next), the short-term design can be regarded as the case of conventional (i.e., in the absence of RIS) power-domain NOMA. Therefore, the decoding orders of different users can be dynamically adjusted according to the instantaneous channel gain in the short-term. The design in our paper can be readily extended to this case since the power allocation strategy is similar.

C. Conventional Transmission Protocol: Full I-CSI Based Design

In the existing literature, the conventional transmission protocol for STAR-RIS is as follows. In each channel coherence block, the BS first estimates the full I-CSI of all element-wise channels (i.e., the cascaded channels \(w_k\) as well as the direct channels \(h_k\). Then, the BS determines the STAR-RIS transmission/reflection coefficients based on the estimated full I-CSI. However, this approach entails at least \(KM + K\) training symbols by using existing channel estimation methods [37], which incurs prohibitively high channel estimation overhead.
and hence reduces the average achievable rate due to a short time for data transmission.

To address this issue, in the next sections, we propose two practical transmission protocols for balancing the tradeoff between system performance and channel estimation overhead.

III. TTS TRANSMISSION PROTOCOL FOR LOs-DOMINANT CHANNELS

In this section, we consider the general scenario where the STAR-RIS associated channels are LoS-dominant. For this scenario, we propose an efficient TTS transmission protocol, called BTE protocol; according to which we optimize the long-term STAR-RIS transmission/reflection coefficient vectors based on the S-CSI, and the short-term BS power allocation based on the effective I-CSI.

A. Proposed BTE Transmission Protocol

As shown in Fig. 2, the proposed BTE protocol consists of the following three main phases. In the first phase, the long-term S-CSI (including the information of the path-loss, Rician factors and the LoS components of all channels) is first estimated at the STAR-RIS and BS. Then, based on the S-CSI, the operation modes and phase-shifts of the STAR-RIS elements are designed and kept fixed for the subsequent coherence blocks for which the S-CSI remains unchanged. In the second phase, the BS estimates the instantaneous effective channels of all the users in each time block, based on which the transmit power allocation is optimized for maximizing the achievable sum-rate. In the third phase, after effective CSI estimation, the BS transmits data to the users.

The benefits of this protocol are as follows. First, the S-CSI mainly depends on the locations and channel statistics which vary slowly and remain largely invariant for a number of coherence blocks. As long as the S-CSI remains unchanged, the STAR-RIS coefficients are kept fixed regardless of the instantaneous channel variations. Therefore, compared to the full I-CSI based design which requires the instantaneous configuration of the STAR-RIS, the proposed BTE protocol greatly reduces the computational complexity and feedback overhead. Second, in each coherence block, the BS only needs to estimate the effective CSI (i.e., $|c_k(v_s)|^2$, $\forall k$), which only requires one pilot symbol for each user. Hence, the overall channel training overhead for the BTE protocol is equal to $K$, which is much smaller than that of the conventional full I-CSI based transmission protocol (i.e., $KM + K$).

For simplicity, we assume that the S-CSI and the effective I-CSI are perfectly acquired; while the extension to a robust design under imperfect CSI is left for future work.

B. Problem Formulation

By considering the channel estimation overhead, the achievable rate of user $k$ is given by $R_k^a = (1 - \frac{T_s}{T_c})R_k$, where $T_c$ denotes the channel coherence time (normalized by the symbol duration). For the proposed BTE protocol, we aim to maximize the average sum-rate by optimizing the long-term STAR-RIS transmission- and reflection-coefficient vectors $\{v_s\}$ and the short-term BS power allocation $\{p_k\}$, subject to constraints on the individual user’s rate given by $\gamma_k$.

Mathematically, this optimization problem can be formulated as

$$(P1) : \max_{\{v_s\}} \max_{\{p_k\}} \mathbb{E} \left\{ \sum_{k=1}^{K} R_k^a \right\}$$

subject to $R_k^a \geq \gamma_k, \forall k \in K$, $|c_k(v_s)|^2 > |c_j(v_s)|^2, \forall k < j$, $||v_m||^2 + ||v_r||^2 \leq 1, \forall m \in M$, $||v_s||_m \in \{0,1\}, \forall m \in M, \forall s \in \{t,r\}$,

$$\sum_{k=1}^{K} p_k \leq P_{\text{max}},$$

where the expectation in (5a) is taken over the random channel realizations. Note that the inner rate maximization problem is to optimize the short-term BS power allocation in each time slot given the STAR-RIS coefficient vectors $\{v_s\}$, while the outer rate maximization problem is to optimize $\{v_s\}$. Problem (P1) is challenging to solve since the design variables are highly coupled in the objective function and constraints. Moreover, it is difficult to obtain a closed-form expression for $\mathbb{E} \left\{ \sum_{k=1}^{K} R_k^a \right\}$. To tackle these challenges, we propose an efficient algorithm to obtain a high-quality suboptimal solution in the next subsection.

C. Proposed Solution to Problem (P1)

1) Short-Term BS Power Allocation Optimization: First, from problem (P1), the short-term optimization problem reduces to

$$(P1.1) : \max_{\{p_k\}} \sum_{k=1}^{K} R_k^a$$

subject to (5b), (5f).

The optimal solution to (P1.1) is given as follows.

---

3These components mainly depend on location/angle information, which can be estimated via the method in [40]. Dedicated sensors can also be deployed to facilitate efficient S-CSI estimation [24].
Proposition 1: Given the effective channel power gains \(|c_k(v_s)|^2\), the optimal BS power allocation is given by
\[
\begin{align*}
p_k^* &= \frac{2\gamma_k - 1}{2\gamma_k} \left( P_{\text{max}} + \frac{\sigma^2}{|c_k(v_s)|^2} \right), \\
p_2^* &= \frac{2\gamma_2 - 1}{2\gamma_2} \left( P_{\text{max}} - \sum_{k=3}^{K} p_k^* + \frac{\sigma^2}{|c_2(v_s)|^2} \right), \\
p_1^* &= P_{\text{max}} - \sum_{k=2}^{K} p_k^*.
\end{align*}
\]
(7)

Proof: See Appendix A. \(\square\)

ii) Long-Term STAR-RIS Transmission- and Reflection-Coefficient Vectors Optimization:

For optimizing the STAR-RIS coefficients by exploiting the S-CSI, we first provide a closed-form expression for the expectation of the channel power gain \(|c_k(v_s)|^2\).

Lemma 1: For the BTE protocol, the expected effective channel power gain of user \(k\), i.e., \(\mathbb{E}\{|c_k(v_s)|^2\}\), is given by
\[
\mathbb{E}\{|c_k(v_s)|^2\} = \mathbb{E}\{|h_k + \bar{r}_k^H \Theta |g|^2\} = \delta_k + \epsilon_k \bar{w}_k^H v_s |^{2} + \zeta_k \sum_{m=1}^{M} \beta_{m},
\]
(8)
where \(\epsilon_k = \frac{\kappa_1 \delta_1 \delta_2}{(\kappa_1 + 1)(\kappa_2 + 1)}\), \(\zeta_k = \frac{(\kappa_1 + \kappa_2 + 1) \delta_1 \delta_2}{(\kappa_1 + 1)(\kappa_2 + 1)}\), and \(\bar{r}_k^H \Theta |g| \in \mathbb{C}^{1 \times M}\) is defined as the cascaded LoS channel.

Proof: See Appendix B. \(\square\)

Next, we focus on the average achievable rate. Since it is difficult to obtain a closed-form expression for \(\mathbb{E}\{R_k^a\}\), we approximate it in the following lemma.

Lemma 2: For the BTE protocol, the expected achievable communication rate of user \(k\) can be approximated as
\[
\mathbb{E}\{R_k^a\} \approx \left(1 - \frac{K}{T_0}\right) \log_2 \left(1 + \frac{p_k \mathbb{E}\{|c_k(v_s)|^2\}}{\sum_{j<k} p_j \mathbb{E}\{|c_j(v_s)|^2\} + \sigma^2}\right)
\]
(9)

\[\triangleq \tilde{R}_k.\]

Proof: The proof is similar to that of [Theorem 1, [13]] and hence is omitted for brevity. \(\square\)

For the long-term optimization problem, directly substituting the optimal power allocation in (7) into (9) will result in complicated rate expressions. To address this difficulty, we introduce a set of auxiliary variables \(\{\tilde{p}_k\}\) that denote the reference power allocation policy based on the S-CSI. Note that \(\{\tilde{p}_k\}\) are used for designing the STAR-RIS coefficient vectors in the long-term optimization only, while the actual BS power allocation policy in each coherence block is given by (7).

Based on the above approximations and introduced auxiliary variables, the long-term optimization problem can be approximated as follows:

(P2) : \(\max_{\{v_s, \tilde{p}_k\}}\ \sum_{k=1}^{K} \tilde{R}_k\) \(\text{s.t.} \ \tilde{R}_k \geq \gamma_k, \forall k \in K, \ \mathbb{E}\{|c_k(v_s)|^2\} > \mathbb{E}\{|c_j(v_s)|^2\}, \forall k < j,\)
(10a)

(10b)

(10c)

Note that to make problem (P2) tractable, we have replaced the instantaneous rates and channel power gain in (5b) and (5c) with their approximations in (10b) and (10c), respectively. The main challenges for solving (P2) lie in its non-convex property and the binary constraint (5e) for MS. To tackle these challenges, we propose an efficient two-layer penalty-based algorithm. Specifically, in the inner layer, we solve a penalized optimization problem by applying the alternating optimization (AO) method, while in the outer layer, the penalty coefficient is updated until convergence is achieved.

1) Inner Layer Iteration: First, in the inner layer, the original problem (P2) is decomposed into two subproblems: BS transmit power optimization and transmission- and reflection-coefficient vectors optimization.

Optimize \(\tilde{p}_k\) for given \(v_s\): For given \(v_s\), the optimal BS power allocation can be obtained via (7) by substituting \(|c_k(v_s)|^2\) with \(|c_k(v_s)|^2\).

Optimize \(v_s\) for given \(\tilde{p}_k\): To tackle the binary constraint (5e), we first transform it equivalently into the following equality constraint:
\[
||v_s||_m |(1 - ||v_s||_m)| = 0, \forall s \in \{t, r\}, m \in M.
\]
(11)
As \(||v_s||_m \leq 1\), we can readily observe that \(||v_s||_m - ||v_s||_m^2 \geq 0\) and the equality holds only when \(||v_s||_m\) is either 0 or 1. Hence, the constraint (11) is satisfied only for binary variables. We then add the left-hand-side (LHS) of (11) as a penalty term into the objective function, yielding the following optimization problem:

(P2.1) : \(\max_{\{v_s\}}\ \sum_{k=1}^{K} \tilde{R}_k - \eta \sum_{m=1}^{M} \sum_{s \in \{t, r\}} (||v_s||_m - ||v_s||_m^2)\)
\(\text{s.t.} \ (5d), (10b), (10c),\)
(12a)

where \(\eta > 0\) is the penalty factor used to penalize the objective function if \(||v_s||_m\) \(\in (0, 1)\). According to [Appendix B, [41]], the results obtained from the penalized problem (P2.1) satisfy the Karush-Kuhn-Tucker (KKT) conditions of the original problem (P2). However, the added penalty term in the objective function is non-convex. To address it, we approximate this term by employing the first-order Taylor series (FTS). Specifically, an upper-bound for the penalty term is obtained as follows:
\[
||v_s||_m - ||v_s||_m^2 \leq ||v_s||_m - \frac{2\text{Re}(|v_s||_m|v_s||_m) - ||v_s||_m^2}{\tilde{f}_0(v_s)} \triangleq f_0(v_s).
\]
(13)

By replacing the penalty term with \(f_0(v_s)\), problem (P2.1) can be transformed into

(P2.2) : \(\max_{\{v_s\}}\ \sum_{k=1}^{K} \tilde{R}_k - \eta \sum_{m=1}^{M} \sum_{s \in \{t, r\}} f_0(v_s)\)
\(\text{s.t.} \ (5d), (10b), (10c).\)
(14a)

Next, to tackle the non-convex objective function and constraint (10b), we first introduce auxiliary variables \(\{\chi_k\}\) and
substitute them into (P2.2). By doing so, the optimization problem can be equivalently expressed as

\[
\text{(P2.3)} : \max_{\{\mathbf{v}_s, \chi_k\}} \sum_{k=1}^{K} \log_2(1 + \chi_k) - \sum_{m=1}^{M} \sum_{s \in \{t, r\}} f_0(\mathbf{v}_s)
\]

\[
\text{s.t. } \log_2(1 + \chi_k) \geq \gamma_k, \forall k, \\
\chi_k \leq \sum_{j < k} \bar{p}_j \mathbb{E}[c_k(\mathbf{v}_s)] + \sigma^2, \forall k,
\]

(5d), (10c).

However, the constraints (15c) and (10c) are still non-convex. To address it, we first rewrite constraint (15c) as

\[
\bar{p}_k (l_k + \epsilon_k) \mathbf{w}_k^H \mathbf{v}_s^2 + \zeta_k \sum_{m=1}^{M} |\mathbf{v}_m|^2 \\
\chi_k \geq \sum_{j < k} \bar{p}_j \mathbb{E}[c_k(\mathbf{v}_s)] + \sigma^2. 
\]

Notice that in the LHS of (16), \( \mathbf{w}_k^H \mathbf{v}_s^2 \) and \( |\mathbf{v}_m|^2 \) are quadratic-over-affine functions, and hence are jointly convex over the involved variables. Besides, \( \frac{1}{\chi_k} \) is convex with respect to \( \chi_k \). Thus, we approximate the non-convex terms in the LHS of (16) by the successive convex approximation (SCA) method. Specifically, according to the FTS, these terms can be approximated at given local points \( \chi_k \), \( \tilde{\chi}_k \) as

\[
\chi_k \geq \frac{2}{\chi_k} - \frac{\chi_k}{(\chi_k)^2} \triangleq f_1(\chi_k), \\
\frac{\mathbf{w}_k^H \mathbf{v}_s^2}{\chi_k} \triangleq f_2(\mathbf{v}_s, \chi_k), \\
\sum_{m=1}^{M} \frac{|\mathbf{v}_m|^2}{\chi_k} \geq \frac{2 \mathbb{E}[\mathbf{v}_s^* \mathbf{v}_m]}{\chi_k} - \frac{(\mathbf{w}_k^H \mathbf{v}_s^2)^2}{\chi_k} \triangleq f_3(\mathbf{v}_s, \chi_k).
\]

Finally, for handling the non-convex constraint (10c), the LHS can also be replaced by its lower-bound, which is given by

\[
\mathbb{E}[c_k(\mathbf{v}_s)] \geq l_k + \epsilon_k \left[ 2 \mathbb{E}[\mathbf{v}_s^* \mathbf{w}_k^H \mathbf{v}_s] - \mathbf{w}_k^H \mathbf{v}_s^2 \right] \\
+ \zeta_k \sum_{m=1}^{M} \left[ 2 \mathbb{E}[\mathbf{v}_s^* \mathbf{v}_m] - |\mathbf{v}_m|^2 \right] \triangleq f_{4,k}(\mathbf{v}_s).
\]

Based on the above approximations, (P2.3) can be transformed into the following problem:

\[
\text{(P2.4)} : \max_{\{\mathbf{v}_s, \chi_k\}} \sum_{k=1}^{K} \log_2(1 + \chi_k) - \sum_{m=1}^{M} \sum_{s \in \{t, r\}} f_0(\mathbf{v}_s)
\]

\[
\text{s.t. } \log_2(1 + \chi_k) \geq \gamma_k, \forall k, \\
\bar{p}_k [l_k f_1(\chi_k) + \epsilon_k f_2(\mathbf{v}_s, \chi_k) + \zeta_k f_{3}(\mathbf{v}_s, \chi_k)] \\
\geq \sum_{j < k} \bar{p}_j \mathbb{E}[c_j(\mathbf{v}_s)] + \sigma^2, \forall k, \\
f_{4,k}(\mathbf{v}_s) \geq \mathbb{E}[c_j(\mathbf{v}_s)], \forall k < j, \\
(5d).
\]

It is easy to verify that problem (P2.4) is a convex optimization problem, which thus can be efficiently solved via standard solvers such as CVX [42].

2) Outer Layer Iteration: In the outer layer, we gradually increase the value of \( \eta \) as \( \eta = \omega \eta \) with \( \omega > 1 \). The iteration of the outer loop terminates when the violation of the equality constraint is within a predefined accuracy, i.e.,

\[
\max \{ ||\mathbf{v}_m|| - ||\mathbf{v}_m||^2, \forall m \in \mathcal{M}, s \in \{t, r\} \} \leq \epsilon.
\]

Algorithm 1 Proposed Penalty-Based Iterative Algorithm for Solving Problem (P2)

1: Initialize feasible points \( \{\mathbf{v}_m^{(n)}, \zeta_k^{(n)}\} \) with \( n = 0 \), the penalty factors \( \eta \).
2: repeat: outer loop
3: \hspace{1em} repeat: inner loop
4: \hspace{2em} For given \( \{\mathbf{v}_m^{(n)}, \chi_k^{(n)}\} \), obtain the power allocation \( \bar{p}_k^{(n+1)} \) via (7).
5: \hspace{2em} For given \( \bar{p}_k^{(n+1)} \), update \( \{\mathbf{v}_s^{(n+1)}, \zeta_k^{(n+1)}\} \) by solving (19).
6: \hspace{1em} until the maximum number of inner iterations \( n_{\text{max}} \) is reached.
7: Update the penalty factor as \( \eta = \omega \eta \).
8: until the constraint violation is below a predefined threshold \( \epsilon > 0 \).

The overall algorithm for solving (P2) is summarized in Algorithm 1. Since the value of the objective function is improved and bounded in each iteration of the inner loop, the proposed penalty-based iterative algorithm is guaranteed to convergence as \( \eta \) grows sufficiently large. Next, we analyze its computational complexity. Specifically, let \( I_{\text{out}} \) and \( I_{\text{in}} \) denote the number of outer and inner iterations. As solving the inner problem requires the computation complexity order of \( O((KM)^{3.5}) \), the whole computational complexity order of Algorithm 1 is given by \( O(I_{\text{out}}I_{\text{in}}(KM)^{3.5}) \) [43].

Remark 3: It is worth mentioning that the performance of the proposed BTE protocol is affected by the Rician factors, as explained below. From (8), it is observed that the expected effective channel power gains (i.e., \( \mathbb{E}[|c_k(\mathbf{v}_s)|^2] \)) are determined by \( \{\mathbf{v}_s\} \) and the Rician factors \( \kappa_1, \kappa_2 \). Specifically, when \( \kappa_1, \kappa_2 \) are large (corresponding to the case of LoS-dominant channels), \( \mathbb{E}[|c_k(\mathbf{v}_s)|^2] \) is proportional to \( \mathbf{w}_k^H \mathbf{v}_s^2 \). On the other hand, when \( \kappa_1, \kappa_2 \rightarrow 0 \) (corresponding to Rayleigh fading), the second term in (8) (i.e., \( \epsilon_k \mathbf{w}_k^H \mathbf{v}_s^2 \)) will have little impact on the channel power gains since \( \epsilon_k \rightarrow 0 \). In this case, the effective channel power gains are mainly determined by the direct channels and the number of transmission/reflection elements, whereas the STAR-RIS phase-shifts do not affect the system performance. In other words, the STAR-RIS serves as a random scatterer, whose phase-shifts can be set arbitrarily. Therefore, for the rich scattering environment, the BTE protocol may be ineffective since the system performance gain brought by STAR-RIS is limited by the dominant NLoS channel components. This motivates
us to devise an alternative scheme in the next section to fully exploit the aperture gain of STAR-RIS when the Rician factors are small.

IV. TTS TRANSMISSION PROTOCOL FOR RICH SCATTERING ENVIRONMENTS

In this section, we consider a rich scattering environment with small Rician fading factors. For this scenario, we propose a customized TTS transmission protocol and devise an efficient algorithm to design the STAR-RIS coefficients and BS power allocation.

A. Proposed PTE Transmission Protocol

One key observation is that if each STAR-RIS element is allocated for serving one user, the instantaneous channel estimation overhead will be greatly reduced. Leveraging this observation, we propose a new PTE transmission protocol as follows. In the first phase, the BS acquires the large-scale path-loss information (i.e., \( \delta_{BS}, \delta_{Sk}, \delta_k \)). Based on this S-CSI, the BS determines the STAR-RIS surface-partition strategy where the STAR-RIS elements are divided into several transmission/reflection subsurfaces, each of which is responsible to assist the data transmission of one specific T/R user.

As introduced before, we target to partition the STAR-RIS elements into \( K \) subsurfaces for sum-rate maximization. Assume that the \( k \)-th subsurface containing \( M_k \) adjacent elements is allocated to user \( k \). Accordingly, we have \( \sum_{k=1}^{K} M_k = M \) and the indices of the STAR-RIS elements associated with user \( k \) are denoted by \( M_k = \{1, \ldots, M_k\} \).

Let \( \mathbf{q}_k^H \in \mathbb{C}^{1 \times M_k} \) denote the cascaded BS to user \( k \) channel through the \( k \)-th subsurface. For each channel coherence block, the BS estimates the direct and cascaded subsurface channels (referred to as the partial I-CSI) for each user instead of the full I-CSI and then optimizes its power allocation as well as the STAR-RIS phase-shifts. In the third phase, the BS proceeds with data transmission.

B. Problem Formulation

As introduced before, we target to partition the STAR-RIS elements into \( K \) subsurfaces for sum-rate maximization. Assume that the \( k \)-th subsurface containing \( M_k \) adjacent elements is allocated to user \( k \). Accordingly, we have \( \sum_{k=1}^{K} M_k = M \) and the indices of the STAR-RIS elements associated with user \( k \) are denoted by \( M_k = \{1, \ldots, M_k\} \).

Let \( \mathbf{q}_k^H \in \mathbb{C}^{1 \times M_k} \) denote the cascaded BS to user \( k \) channel through the \( k \)-th subsurface. For each channel coherence block, the BS estimates the direct and cascaded subsurface channels (referred to as the partial I-CSI) for each user instead of the full I-CSI and then optimizes its power allocation as well as the STAR-RIS phase-shifts. In the third phase, the BS proceeds with data transmission.

C. Proposed Solution to Problem (P3)

i) Short-Term BS Power Allocation and STAR-RIS Phase-Shifts Design:

Phase-Shifts: For the short-term optimization problem, we first propose a suboptimal yet low-complexity phase-shift design that is to align the \( k \)-th cascaded subsurface channel with the direct link of user \( k \). Specifically, the \( n \)-th phase-shift of subsurface \( k \) is set as

\[
\theta_n^k = \angle h_k - \angle [\mathbf{q}_k^H]_n, \forall n \in M_k,
\]

For NOMA transmission, the received superimposed signal at user \( k \) is given by

\[
y_k = c_k \sum_{k=1}^{K} \sqrt{p_k} s_k + n_k,
\]

with

\[
c_k = |h_k| + \sum_{n=1}^{M_k} [\mathbf{q}_k^H]_n + \sum_{j=k, j \neq k}^{M} \sum_{n=1}^{M_j} [\mathbf{q}_j^H]_n \approx |h_k| + \sum_{n=1}^{M_k} [\mathbf{q}_k^H]_n,
\]

where \( \mathbf{q}_j^H \in \mathbb{C}^{1 \times M_j} \) denotes the cascaded channel between the \( j \)-th subsurface and user \( k \). The approximation in (a) follows by neglecting the signals transmitted/reflect by other subsurfaces that are not assigned to user \( k \) (i.e., the third term in \( c_k \)). In practical scenarios, the T/R users are exposed to all transmission/reflection elements. For example, for user \( k \in K_t \), the subsurfaces allocated to other users \( j \in K_t \) (\( j \neq k \)) will also scatter signals, which can be combined either constructively or destructively with user \( k \)'s desired signals. Fortunately, our approximation is practically accurate since the effect of the scattered signals is insignificant, especially when \( \kappa_1 \) and \( \kappa_2 \) are small, which will be numerically verified in Section V.

Power Allocation: Next, regarding \( c_k \) as the effective channel of each user, the BS determines the short-term power allocation based on Proposition 1 after estimating the partial I-CSI.

ii) Long-Term surface-Partition Optimization:

Next, we optimize the long-term surface-partition for the average sum-rate maximization. Applying the phase-shift design in (22), the expected effective channel power gains \( \mathbb{E} \{ (c_k)^2 \} \) are characterized as follows.
Lemma 3: For the PTE protocol, the expected effective channel power gain of user $k$ is given by
\[
E\{(c_k)^2\} = \delta_k + \delta_{BS} \delta_{Sk} M_k + \frac{\pi^2 \delta_{BS} \delta_{Sk} M_k(M_k - 1)}{16} ~ f(\kappa_1, \kappa_2) + \frac{\pi^2 \delta_{BS} \delta_{Sk} f(\kappa_1, \kappa_2)}{16} M_k,
\]
where $f(\kappa_1, \kappa_2) = \frac{L_\frac{1}{2}(\kappa_1 L_\frac{1}{2}(\kappa_2))^2}{(\kappa_1 + 1)(\kappa_1 + 1)}$, and $L_\frac{1}{2}()$ denotes the Laguerre polynomial.

Proof: See Appendix C. □

From Lemma 3, we can derive that the PTE protocol is less sensitive to the Rician factor since $f(\kappa_1, \kappa_2) = \frac{L_\frac{1}{2}(\kappa_1 L_\frac{1}{2}(\kappa_2))^2}{(\kappa_1 + 1)(\kappa_1 + 1)}$ is little affected by $\kappa_1$ and $\kappa_2$. Based on Lemma 3, the expected achievable rate of user $k$ can be approximated as
\[
E(R_k^b) \approx \left(1 - \frac{M + K}{T_c}\right) \log_2 \left(1 + \frac{p_k E\{(c_k)^2\}}{\sum_{j < k} p_j E\{(c_j)^2\} + \sigma^2}\right) = \bar{R}_k.
\]

Therefore, the long-term surface-partition optimization problem can be approximated as
\[
(P4) : \max_{\{M_k\}} \sum_{k=1}^K \bar{R}_k
\]
\[
\text{s.t. } \bar{R}_k \geq \gamma_k, \quad E\{(c_k)^2\} > E\{(c_j)^2\}, \forall k < j. \tag{27c}
\]

Problem (P4) is a non-convex mixed-integer non-linear programming (MINLP), which is generally hard to solve. Although the optimal solution can be obtained by enumerating all possible surface-partition strategies, this exhaustive-search based algorithm entails a prohibitively high computational complexity, which is unaffordable when $M$ is large. Moreover, the conventional integer relaxation and rounding approach can not be applied, since the relaxed problem is still not convex. To tackle these issues, we propose an efficient two-step heuristic algorithm to suboptimally solve problem (P4), whose details are introduced as follows.

1) Initial STAR-RIS Surface-Partition Under Rate and Decoding-Order Constraints: It is expected that NOMA transmission benefits more from the multiplexing gain when the channels of the users are more distinct. Hence, in order to improve the NOMA gain, it is desirable to allocate more elements to the users with higher decoding orders to create asymmetric channel conditions. However, with this strategy, the weak users’ rate requirements may not be met if their effective channel power gains are too small. To tackle this, in the first step, we aim to determine the minimum number of elements allocated to each user for satisfying their rate requirements and the decoding order constraints. Specifically, we first select a small number of elements (e.g., the minimum number of elements in a group) to be assigned to user $K$ (the weakest user), denoted by $M_K^{(0)}$. Based on $M_K^{(0)}$, we obtain the minimum number of elements assigned to other users, i.e., $\{M_k^{(0)}\}$ that satisfy $E\{(c_k)^2\} > E\{(c_{k+1})^2\}, \forall k < K$; while the remaining elements are allocated to user 1 which is of the highest decoding order. Next, we decide whether the rate requirements are satisfied under the current surface-partition strategy by checking the feasibility constraint in (7). If it is infeasible, we gradually increase $M_k^{(0)}$ by $M_k^{(0)} = M_k^{(0)} + 1$ and repeat the above steps until all users’ target rates are reached.

2) Progressive Surface-Partition Refinement for Sum-Rate Enhancement: Although the rate requirements are satisfied in the first step, the obtained surface-partition strategy may not be the optimal choice for sum-rate maximization. To improve the solution, we give the following lemma.

Lemma 4: Given the initial surface-partition strategy in Step 1, and BS power allocation in (7), if an element assigned to user $k$ is reassigned to another user $k'$, then $R_k^b$ increases and $R_k^b$ decreases, while the achievable rates of the other users remain unchanged.

Proof: First, it is observed from the expression of $R_k^b$ that the expected channel power gain and the transmit power of other users are not affected by the reassignment. Therefore, their average achievable rates do not change. Second, after this reassignment, the expected channel power gain of user $k$ and $k'$, i.e., $E\{(c_k)^2\}$ and $E\{(c_j)^2\}$, will increase and decrease, respectively. It is observed from (26) that the rate of user $k$ monotonically increases with $E\{(c_k)^2\}$. Hence, the average rates of users $k$ and $k'$ follow the same trend as their expected channel power gains. This completes the proof. □

Lemma 4 implies that the user rate can be adjusted by reallocating the elements from one to another user. This inspires us to design a progressive surface-partition refinement algorithm to gradually improve the system sum-rate. Specifically, we denote the initial surface-partition strategy obtained in Step 1 as $\{M_k^{(0)}\}$, where $i$ is the initial iteration number; and obtain the power allocation and the corresponding average sum-rate according to Proposition 1, denoted by $\{p_i\}$ and $R_i$, respectively. Next, we consecutively reassign an element associated with user 1 to another user $k$ as $k = 2, 3, \ldots, K$, and calculate the average sum-rate after reassignment, denoted...
by $R^{(i)}$. If $R^{(i')} > R^{(i)}$, we update the STAR-RIS surface-partition strategy and BS power allocation, and then enter the next iteration. This process is repeated until the system sum-rate can not be improved anymore or the power allocation problem in (P3) is infeasible.

The main procedures of the proposed solution for solving (P4) are summarized in Algorithm 2. Generally, the worst-case complexity of the proposed algorithm is $O(KM)$, which is much smaller than that of the exhaustive-search based algorithm, i.e. $O(M^K)$.

Algorithm 2 Proposed Low-Complexity Algorithm for Solving Problem (P4)

1: Step 1:
2: Initialize $t = 0$, a small number of elements assigned to user $K$, $M^{(i)}_K$.
3: Calculate the minimal $M^{(i)}_k$ that satisfies the decoding order constraint.
4: repeat
5: $M^{(i+1)}_k = M^{(i)}_k + 1, \forall k$.
6: $t = t + 1$.
7: until (7) is feasible.
8: Step 2:
9: Initialize $l = 0$, $M^{(i)}_k$ as that obtained from the last step.
10: repeat
11: $M^{(i')}_1 = M^{(i)}_1 - 1, M^{(i')}_k = M^{(i')}_k + 1, k = 2, \ldots, K$.
12: if $R^{(i)} > R^{(i)}$ then
13: $l = l + 1$.
14: $M^{(i')}_k = M^{(i')}_k, p^{(i')}_k = p^{(i')}_k, \forall k$.
15: endif
16: until there is no rate increase.

D. Comparison Between the Proposed Protocols

In Table I, we summarize and compare the proposed two TTS transmission protocols in terms of the required CSI, design parameters, optimization methods, and channel estimation overhead. Compared with the BTE protocol, the PTE protocol requires a larger channel estimation overhead, while is less sensitive to the variations of the channel statistics. Specifically, the BTE protocol is only effective when the channels are LoS-dominant, while results in a considerable rate loss in rich scattering environments, e.g., when $\kappa_1, \kappa_2 \to 0$, as explained in Remark 1. In contrast, the proposed PTE protocol is applicable to any fading conditions. Particularly, it provides a squared-order beamforming gain, i.e., $O(M^2)$, even for Rayleigh fading channels. This is because the expected effective channel power gain, i.e., $E\{(c_k)^2\}$ always contains the term (i.e., the third term in (25)) that is proportional to $(M_k)^2$, as shown in Lemma 3.

V. SIMULATION RESULTS

In this section, numerical results are provided to validate the effectiveness of our proposed TTS transmission protocols and the STAR-RIS MS design. We consider a three-dimensional system setup, where the BS is located at the (0, 0, 1) meters (m) and the reference center of the STAR-RIS is at (50, 0, 1) m. We assume that two T users and two R users are randomly located within a circle centered at the STAR-RIS with a radius of 5 m, unless otherwise stated. The distance-dependent path loss is modeled as $\delta(d) = \rho_0(d/d_0)^{-\alpha}$, where $d$ is the individual link distance, $\rho_0$ denotes the path loss at the reference distance $d_0 = 1$ m, and $\alpha$ denotes the path loss exponent. We set $\rho_0 = -30$ dB, $\alpha_k = 3.5$, $\alpha_{BS} = 2$, $\alpha_{Sk} = 2.2$. The noise power is set as $\sigma^2 = -80$ dBm and the Rician factors are set as $\kappa_1 = \kappa_2 = \kappa$. Besides, the achieved sum-rates of the proposed protocols are averaged over 5000 random channel realizations.

A. Convergence Behavior of the Proposed Algorithms

In Fig. 4, we verify the convergence behavior of the proposed algorithms. Specifically, Fig. 4(a) shows the violation of the equality constraints and the average sum-rate of Algorithm 1 versus the number of outer loop iterations with $\eta = 10^{-1}$ and $\omega = 10$. It is observed that the constraint violation decreases quickly as the number of outer loop iterations increases, and ultimately reaches the predefined accuracy (i.e., $\epsilon = 10^{-7}$). Besides, the sum-rate increases quickly and converges after 4 iterations. Fig. 4(b) plots the system sum-rate versus the times of element reassignments for the second step of Algorithm 2. With the proposed progressive refinement
algorithm, the system sum-rate is observed to monotonically increase once the rate requirements are met in the first step.

B. Accuracy of the Expected Rate/Channel Power Gain Approximation

In Fig. 5, we evaluate the accuracy of the approximation for the expected sum-rate of the proposed protocols, i.e., \( \sum_{k=1}^{K} \bar{R}_k \) and \( \sum_{k=1}^{K} \tilde{R}_k \). Specifically, the approximations of the expected rates of each user for the proposed BTE protocol and PTE protocol are given as (9) and (26), respectively. Two sets of experiments are carried out under \( P_{\text{max}} = 30 \) dBm, \( \gamma_k = 2 \) bps/Hz and \( P_{\text{max}} = 25 \) dBm, \( \gamma_k = 1 \) bps/Hz, respectively. It is observed that, for all cases and both protocols, the approximated sum-rate is only slightly larger than the simulation result, since the approximation using the Jensen inequality serves as an upper-bound for the exact average achievable rate.

In Fig. 6, we further evaluate the accuracy of the approximation for the expected channel power gains of the PTE protocol given in (25). The surface-partition strategy is optimized under \( M = 50 \), \( \gamma_k = 2 \), \( \kappa = 1 \). Under this setup, according to Algorithm 2, the number of elements assigned to each user is 36, 9, 3 and 2, respectively. The exact expected channel power gains, i.e., \( |c_k|^2 \), are obtained by averaging over 10000 random channel realizations for each user. As observed, our analysis matches well with the exact results for users 1, 3, 4, regardless of the Rician factor \( \kappa \). This is because these users’ desired signals are almost not affected by other transmission/reflection elements that are not assigned to them. For user 2, its exact channel power gain is slightly larger than the made approximation. This is because, in this case, the subsurface associated with user 1 serves as a random scatterer, whose scattered signals are combined constructively with user 2’s desired signals.

C. Effectiveness of Decoding Orders

To show the effectiveness of our decoding order design, we consider the exhaustive search-based algorithm as a benchmark for both protocols. We show the achievable sum-rates attained by different algorithms versus the number of total elements in Fig. 7. It is observed that the proposed design for both protocols achieve very close performance with their respective upper-bounds, which thus validates that our design is efficient for the considered setups.

D. Performance Comparison

To demonstrate the benefits brought by STAR-RIS and NOMA, we consider the following benchmark schemes:

- **STAR-RIS assisted FDMA/TDMA system**: Two types of OMA schemes are considered, namely, frequency division multiple access (FDMA) and time division multiple access (TDMA). For FDMA, the operation modes and the phase-shifts are configured in the same way as for our proposed PTE protocol, while the users are served over orthogonal frequency RBs. For TDMA, each coherence block is further divided into several equal sub-blocks, each of which is responsible for serving one user, with all the STAR elements operating in the transmission/reflection mode. Note that the channel estimation overhead for FDMA and TDMA is \( M+K \) and \( K(M+1) \), respectively.
Fig. 7. Average achievable sum-rate versus the number of total elements $M$ with $T_c = 500$, $\kappa = 3$, $\gamma_k = 1.8 \text{ bps/Hz}$.

Fig. 8. Average achievable sum-rate versus the number of total elements $M$ with $T_c = 400$, $\kappa = 3$, $\gamma_k = 1.8 \text{ bps/Hz}$.

- Conventional RIS-assisted NOMA system (CR-NOMA): In this case, one reflecting-only RIS and one transmitting-only RIS (each with $M/2$ elements) are co-located for achieving full-space coverage. This scheme can also be regarded as a special case of mode switching, where half of the elements are fixed to operate in the transmission/reflection mode.

- Instantaneous full CSI without considering channel estimation overhead (Ideal CSI): In this case, we assume that the instantaneous full CSI is perfectly known at the BS for system optimization. To evaluate the performance of the proposed CSI acquisition methods, we neglect the channel estimation overhead of this benchmark scheme (which requires $KM + K$ channel training symbols in practice), thus its achievable sum-rate serves as the performance upper-bound.

We first consider a simple scenario with two users to show the superiority of NOMA over OMA. In Fig. 8, we plot the sum-rate of the users versus the channel coherence time $T_c$ with $M = 50$, $\kappa = 3$, $\gamma_k = 3 \text{ bps/Hz}$. Several important observations are made as follows. Firstly, it is observed that the performance of the proposed BTE protocol is not affected by $T_c$, since by exploiting S-CSI, almost all the time is utilized for data transmission. In contrast, the achieved sum-rate of the proposed PTE protocol increases with $T_c$ and is larger than that of the BTE protocol when $T_c$ is sufficiently large, i.e., $T_c \geq 500$. Moreover, as $T_c$ grows, the impact of the channel estimation overhead on the system performance weakens. In this case, utilizing the I-CSI for the STAR-RIS beamforming design becomes more effective. Secondly, STAR-RIS assisted systems based on NOMA largely outperform those based on FDMA and TDMA, since NOMA provides a larger multiplexing gain than FDMA/TDMA. This also indicates that NOMA is more suitable in practical systems for supporting more users. Finally, although the TDMA scheme can dynamically adjust STAR-RIS phase-shifts over time, it incurs a larger channel estimation overhead for each user, which is not practically effective.

Next, we consider the more general case with four users to show the benefits of deploying STAR-RISs. In Fig. 9, we depict the sum-rate brought by different transmission protocols versus the number of total elements $M$. It is observed that the ideal CSI scheme offers the best performance. The performance gap between the ideal CSI case and the proposed BTE protocol is because the S-CSI can not well characterize the variations of the instantaneous channels, whereas for the proposed PTE protocol, the gap is mainly attributed to the training time of channel estimation. One can observe that there exists a tradeoff between the total number of elements and the system sum-rate of the PTE protocol. Specifically, the average rate under the PTE protocol first increases and then decreases with $M$. This can be explained as follows. When $M$ is small, a larger surface achieves a higher passive beamforming gain. However, as $M$ becomes larger, the penalty due to channel estimation overhead starts to dominate, which becomes the main performance bottleneck when $M$ exceeds a threshold (i.e., $M > 50$). On the other hand, the sum-rate of the BTE protocol monotonically increases with $M$, because with S-CSI, the overhead for estimating the instantaneous effective channels in each coherence block is very small. Furthermore, as observed, the considered STAR-RIS aided systems significantly outperform the conventional RIS-aided systems for different $M$, and the performance gain increases with $M$. In a nutshell, the two proposed transmission protocols for STAR-RIS are generally preferred for the scenarios with a large and a small number of elements, respectively.
Fig. 10 shows the rate performance of the proposed protocols versus the Rician factor \( \kappa \). It is observed that both protocols for STAR-RIS-NOMA systems outperform their counterparts for CR-NOMA systems, since STAR-RIS provides more flexibility in controlling the number of transmission/reflection elements, thus achieving a larger NOMA gain. Moreover, the performance of the BTE protocol increases drastically at first and then tends to be saturated. This is because the S-CSI can well characterize the actual channel condition in the high-Rician-factor regime, but may be ineffective when the NLoS components are dominant, as discussed in Remark 1. On the other hand, the sum-rate achieved by the PTE protocol slightly increases with \( \kappa \), which indicates that the proposed PTE protocol is not sensitive to the fading condition. To summarize, the proposed BTE protocol outperforms the PTE protocol when the LoS components of the channel coefficients are dominant, and vice versa.

VI. Conclusion

In this paper, we proposed two efficient TTS transmission protocols for different channel setups in a STAR-RIS aided NOMA communication system under MS. Specifically, for the LoS-dominant channel setup, a BTE protocol was first proposed, where the long-term STAR-RIS transmission- and reflection-coefficient vectors are optimized based on the S-CSI, while the short-term BS power allocation in each coherence block is designed based on the estimated effective CSI. Moreover, a penalty-based method was proposed to optimize the STAR-RIS beamforming coefficients. Next, in scattering channels, we further proposed a customized PTE protocol, where the long-term operation modes of the STAR-RIS elements are optimized based on the large-scale path-loss, while the short-term BS power allocation and STAR-RIS phase-shifts are designed to cater to the estimated partial I-CSI. Moreover, a low-complexity surface-partition algorithm was devised to obtain the elements allocation assigned to serve each user. Numerical results validated the effectiveness of our proposed designs. Particularly, both transmission protocols can greatly reduce the channel estimation overhead while achieving high communication rates.

APPENDIX A: PROOF OF PROPOSITION 1

Given the effective channel power gains of all users, the short-term optimization problem is equivalent to a power allocation problem in a single-cluster NOMA system. Following the method in [45], to maximize the system sum-rate under users’ target rates, each user should be first allocated the minimum power for meeting its rate requirement, while the remaining power should be allocated to the user with the highest decoding order. Hence, the optimal power allocation is obtained by solving the following equations:

\[
\begin{align*}
    p_2 &= (2^{\gamma_2} - 1) \left( p_1 + \frac{\sigma^2}{|c_2(v_s)|^2} \right), \\
    & \vdots \\
    p_K &= (2^{\gamma_K} - 1) \left( \sum_{k=1}^{K-1} p_k + \frac{\sigma^2}{|c_K(v_s)|^2} \right), \\
    \sum_{k=1}^{K} p_k &= P_{\text{max}}.
\end{align*}
\]

By solving the above equations, we obtain the optimal BS power allocation given in (7). Regarding the optimal \( p_k^* \), there are three cases: 1) \( p_k^* \geq (2^{\gamma_k-1})\sigma^2 \), and \( p_k > 0, \forall k > 1 \). In this case, all the users’ rate requirements are met. 2) \( 0 \leq p_k^* < (2^{\gamma_k-1})\sigma^2 \). In this case, the rate requirements of all users except for the first user can be guaranteed. 3) Any of the \( p_k^* \) has negative values, which indicates that \( P(1.1) \) is infeasible.

The original problem is feasible if all the power allocation factors \( p_k^* \) are positive values and the strongest users’ rate requirements are satisfied.

APPENDIX B: PROOF OF LEMMA 1

The expected effective channel power gains can be decomposed as follows

\[
\mathbb{E}\left\{ |h_k + r_k^H \Theta_s g|^2 \right\} \\
\equiv \mathbb{E}\{ |h_k|^2 \} + |x_1|^2 + \mathbb{E}\{ |x_2|^2 \} + \mathbb{E}\{ |x_3|^2 \} + \mathbb{E}\{ |x_4|^2 \},
\]

where

\[
\begin{align*}
    x_1 &= \sqrt{\kappa_1 \kappa_2 \beta_{BS} \delta_{Sk} \frac{\delta_{Sk}}{(\kappa_1 + 1)(\kappa_2 + 1)}} f_k^H \Theta_s g, \\
    x_2 &= \sqrt{\kappa_1 \beta_{BS} \delta_{Sk} \frac{\delta_{Sk}}{(\kappa_1 + 1)(\kappa_2 + 1)}} f_k^H \Theta_s g, \\
    x_3 &= \sqrt{\kappa_2 \beta_{BS} \delta_{Sk} \frac{\delta_{Sk}}{(\kappa_1 + 1)(\kappa_2 + 1)}} f_k^H \Theta_s g, \\
    x_4 &= \frac{\delta_{BS} \delta_{Sk} \frac{\delta_{Sk}}{(\kappa_1 + 1)(\kappa_2 + 1)}} f_k^H \Theta_s g,
\end{align*}
\]

and (a) is due to the fact that the small-scale fading channel coefficients \( h_k, g, r_k \) are independent. Next, we have

\[
|x_1|^2 = \frac{\kappa_1 \kappa_2 \beta_{BS} \delta_{Sk} \frac{\delta_{Sk}}{(\kappa_1 + 1)(\kappa_2 + 1)}} |w_k^H V_s|^2,
\]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
\[ E\{|x_2|^2\} = \frac{\kappa_1\delta_{BS}}{(\kappa_1+1)(\kappa_2+1)} \tilde{r}_k \Theta_s E\{(\hat{g}^H\hat{g})\Theta_s \tilde{r}_k \} = \frac{\kappa_1\delta_{BS}}{(\kappa_1+1)(\kappa_2+1)} \sum_{m=1}^{M} \beta_m^2, \quad (31) \]

where (b) holds since \( E\{(\hat{g}^H\hat{g})\} = I_M, \Theta_s \Theta^H_s = \text{diag}(\beta_s) \). The remaining terms can be obtained similarly as \( E\{|x_3|^2\} = \frac{\kappa_2^2\delta_{sk1}}{(\kappa_1+1)(\kappa_2+1)} \sum_{m=1}^{M} \beta_m^2 \), \( E\{|x_4|^2\} = \frac{1}{(\kappa_1+1)(\kappa_2+1)} \sum_{m=1}^{M} \beta_m^2 \).

Additionally, we have \( E\{|h_k|^2\} = \delta_k \). Therefore, by substituting (29)-(32) into (29), we arrive at (8).

**APPENDIX C: PROOF OF LEMMA 3**

Applying the binomial expansion theorem, we have

\[ E\{(c_k)^2\} = E\left\{ \left(|h_k| + \sum_{m=1}^{M_k} |q_k|_m \right)^2 \right\} = E\{|h_k|^2\} + E\left\{ \sum_{m=1}^{M_k} |q_k|_m^2 \right\} + 2E\left\{ \sum_{m=1}^{M_k} |h_k| |q_k|_m \right\}. \quad (32) \]

where the second term can be further expanded as

\[ x_2 = E\left\{ \sum_{m=1}^{M_k} |q_k|_m^2 \right\} + E\left\{ \sum_{m=1}^{M_k} \sum_{n=1, n \neq m}^M |q_k|_m |q_k|_n \right\}. \quad (33) \]

First, it is easy to have \( E\{|h_k|^2\} = \delta_k, E\{|q_k|_m^2\} = \delta_{BS}\delta_{sk} \). Then, we note that \( h_k, q_k \) are independent variables with the following results

\[ E\{|h_k|^2\} = \frac{\sqrt{\pi \delta_k}}{2}, \quad (34a) \]

\[ E\{|q_k|_m^2\} = \frac{\pi \delta_{BS}\delta_{sk}}{4(\kappa_1+1)(\kappa_2+1)} L_{\frac{1}{2}}(-\kappa_1) L_{\frac{1}{2}}(-\kappa_2), \forall m \in M_k, \quad (34b) \]

where (34b) is due to the fact that \( |q_k|_m \) is the product of two independent Rician variables, which follow the same distribution with the element in \( g \) and \( \tilde{r}_k \), respectively.

Substituting the above results into (32) yields the desired result in (25).

**REFERENCES**

[1] C. Wu, S. Shi, C. You, Y. Liu, and S. Zhang, “STAR-RIS aided NOMA communication system with statistical CSI,” in Proc. IEEE Globecom Workshops (GC Wkshps), Rio de Janeiro, Brazil, Dec. 2022, pp. 100–105.

[2] M. Di Renzo et al., “Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead,” IEEE J. Sel. Areas Commun., vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[3] Y. Liu et al., “Reconfigurable intelligent surfaces: Principles and opportunities,” IEEE Commun. Surveys Tuts., vol. 23, no. 3, pp. 1546–1577, 3rd Quart., 2021.

[4] 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, May 2021.

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

[6] C. Pan et al., “Multicell MIMO communications relying on intelligent reflecting surfaces,” IEEE Trans. Wireless Commun., vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[7] Y. Liu et al., “STAR: Simultaneous transmission and reflection for 360º coverage by intelligent surfaces,” IEEE Wireless Commun., vol. 28, no. 6, pp. 102–109, Dec. 2021.

[8] W. Zhang et al., “Intelligent omnidirectional communications for full-dimensional wireless communications: Primitives, technology, and implementation,” IEEE Commun. Mag., vol. 60, no. 2, pp. 39–45, Feb. 2022.

[9] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, “Nonorthogonal multiple access for 5G and beyond,” Proc. IEEE, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[10] Z. Ding, P. Fan, and H. V. Poor, “Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions,” IEEE Trans. Veh. Technol., vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[11] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, “Intelligent reflecting surface meets OFDM: Protocol design and rate maximization,” IEEE Trans. Commun., vol. 68, no. 7, pp. 4522–4535, Jul. 2020.

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

[13] M. Hua, L. Yang, Q. Wu, C. Pan, C. Li, and A. L. Swindlehurst, “UAV-assisted intelligent reflecting surface symbiotic radio system,” IEEE Trans. Wireless Commun., vol. 20, no. 9, pp. 5769–5785, Sep. 2021.

[14] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, “Sum rate maximization for IRS-assisted uplink NOMA,” IEEE Commun. Lett., vol. 25, no. 1, pp. 234–238, Jan. 2021.

[15] G. Yang, X. Xu, Y.-C. Liang, and M. Di Renzo, “Reconfigurable intelligent surface-assisted non-orthogonal multiple access,” IEEE Trans. Wireless Commun., vol. 20, no. 5, pp. 3137–3151, May 2021.

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

[17] T. L. Jensen and E. De Carvalho, “An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2020, pp. 5000–5004.

[18] C. You, B. Zheng, and R. Zhang, “Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement,” IEEE J. Sel. Areas Commun., vol. 38, no. 11, pp. 2604–2620, Nov. 2020.

[19] B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design,” IEEE Trans. Wireless Commun., vol. 19, no. 12, pp. 8315–8329, Dec. 2020.

[20] Z. Wang, L. Liu, and S. Cui, “Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis,” IEEE Trans. Wireless Commun., vol. 19, no. 10, pp. 6607–6620, Oct. 2020.

[21] X. Wei, D. Shen, and L. Dai, “Channel estimation for IRS-assisted wireless communications—Part II: An improved solution based on double-structured sparsity,” IEEE Commun. Lett., vol. 25, no. 5, pp. 1403–1407, May 2021.

[22] C. You, B. Zheng, and R. Zhang, “Fast beam training for IRS-assisted multiuser communications,” IEEE Wireless Commun. Lett., vol. 9, no. 11, pp. 1845–1849, Nov. 2020.

[23] 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, Aug. 2019.
[24] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, “Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization,” IEEE Trans. Wireless Commun., vol. 20, no. 1, pp. 2–17, Jan. 2021.

[25] 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, May 2022.

[26] Q. Tao, S. Zhang, C. Zhong, W. Xu, H. Lin, and Z. Zhang, “Weighted sum-rate of intelligent reflecting surface aided multiuser downlink transmission with statistical CSI,” IEEE Trans. Wireless Commun., vol. 21, no. 7, pp. 4925–4937, Jul. 2022.

[27] K. Zhi et al., “Two-timescale design for reconfigurable intelligent surface-aided massive MIMO systems with imperfect CSI,” IEEE Trans. Inf. Theory, vol. 69, no. 5, pp. 3001–3033, May 2023.

[28] A. Abrardo, D. Dardari, and M. Di Renzo, “Intelligent reflecting surfaces: Sum-rate optimization based on statistical position information,” IEEE Trans. Commun., vol. 69, no. 10, pp. 7121–7136, Oct. 2021.

[29] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, “STAR-RISs: Simultaneously transmitting and reflecting reconfigurable intelligent surfaces,” IEEE Commun. Lett., vol. 25, no. 9, pp. 3134–3138, Sep. 2021.

[30] 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, May 2022.

[31] 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. 5, pp. 3036–3040, Sep. 2021.

[32] 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, Jan. 2023, doi: 10.1109/TWC.2022.3197079.

[33] C. Zhang, W. Yi, Y. Liu, Z. Ding, and L. Song, “STAR-IOS aided NOMA networks: Channel model approximation and performance analysis,” IEEE Trans. Wireless Commun., vol. 21, no. 9, pp. 6861–6876, Sep. 2022.

[34] H. Niu, Z. Chu, F. Zhou, and Z. Zhu, “Simultaneous transmission and reflection reconfigurable intelligent surface assisted secrecy MISO networks,” IEEE Commun. Lett., vol. 25, no. 11, pp. 3498–3502, Nov. 2021.

[35] 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, Oct. 2022.

[36] 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. Surveys Tuts., vol. 24, no. 2, pp. 1035–1071, 2nd Quart., 2022.

[37] C. Wu, C. You, Y. Liu, X. Gu, and Y. Cai, “Channel estimation for STAR-RIS-aided wireless communications,” IEEE Commun. Lett., vol. 26, no. 3, pp. 652–656, Mar. 2022.

[38] C. Huang, A. Zappone, M. Debbah, and C. Yuen, “Achievable rate maximization by passive intelligent mirrors,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Calgary, AB, Canada, Apr. 2018, pp. 3714–3718.

[39] J. Xu, Y. Liu, X. Mu, R. Schober, and H. V. Poor, “STAR-RISs: A correlated T&R phase-shift model and practical phase-shift configuration strategies,” IEEE J. Sel. Topics Signal Process., vol. 16, no. 5, pp. 1097–1111, Aug. 2022.

[40] G. Zhou, C. Pan, H. Ren, P. Popovski, and A. L. Swindlehurst, “Channel estimation for RIS-aided multilayer millimeter-wave systems,” IEEE Trans. Signal Process., vol. 70, pp. 1478–1492, Apr. 2022.

[41] Y. Cai, Q. Shi, B. Champagne, and G. Y. Li, “Joint transceiver design for secure downlink communications over an amplify-and-forward MIMO relay,” IEEE Trans. Commun., vol. 65, no. 9, pp. 3691–3704, Sep. 2017.

[42] M. Grant and S. Boyd. (2014). CVX: MATLAB Software for Disciplined Convex Programming. [Online]. Available: http://cvxr.com/cvx

[43] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[44] H. Cho and J. Choi, “Alternating beamforming with intelligent reflecting surface element allocation,” IEEE Wireless Commun. Lett., vol. 10, no. 6, pp. 1232–1236, Jun. 2021.

[45] Z. Chen, Z. Ding, X. Dai, and R. Zhang, “An optimization perspective on the superiority of NOMA compared to conventional OMA,” IEEE Trans. Signal Process., vol. 65, no. 19, pp. 5191–5202, Oct. 2017.

Chenyu Wu received the B.E. and Ph.D. degrees from the School of Electronics and Information Engineering, Harbin Institute of Technology (HIT), Harbin, China, in 2018 and 2023, respectively. He is currently an Assistant Professor with HIT. His current research interests include non-orthogonal multiple access, reconﬁgurable intelligent surface-aided networks, UAV communications, and convex optimization.

Changsheng You (Member, IEEE) received the B.Eng. degree from the University of Science and Technology of China (USTC) in 2014 and the Ph.D. degree from The University of Hong Kong (HKU) in 2018. He was a Research Fellow with the National University of Singapore (NUS). He is currently an Assistant Professor with the Southern University of Science and Technology. His research interests include intelligent reflecting surfaces, UAV communications, edge learning, and mobile-edge computing. He received the IEEE Communications Society Asia-Pacific Region Outstanding Paper Award in 2019, the IEEE ComSoc Best Student Paper Award in 2020, and the IEEE ComSoc Best Tutorial Paper Award in 2023. He is a Guest Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and an Editor of IEEE COMMUNICATIONS LETTERS, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. He is listed as a Highly Cited Chinese Researcher and an exemplary reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Yuanwei Liu (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2016. He was with the Department of Informatics, King’s College London, from 2016 to 2017, where he was a Post-Doctoral Research Fellow. He has been a Senior Lecturer (Associate Professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London, since Aug. 2021, where he was a Lecturer (Assistant Professor) from 2017 to 2021. His research interests include non-orthogonal multiple access, reconﬁgurable intelligent surface, integrated sensing and communications, and machine learning. He has been a Web of Science Highly Cited Researcher since 2021, an IEEE Communication Society Distinguished Lecturer, an IEEE Vehicular Technology Society Distinguished Lecturer, the academic Chair for the Next Generation Multiple Access Emerging Technology Initiative, the rapporteur of ETSI Industry Specification Group on Reconﬁgurable Intelligent Surfaces, and the UK representative for the URSI Commission C. He was listed as one of the 35 Innovators Under 35 China in 2022 by MIT Technology Review. He received IEEE ComSoc Outstanding Young Researcher Award for EMEA in 2020. He received the 2020 IEEE Signal Processing and Computing for Communications (SPCC) Technical Committee Early Achievement Award, IEEE Communication Theory Technical Committee (CTTC) 2021 Early Achievement Award. He received IEEE ComSoc Outstanding Nominee for Best Young Professionals Award in 2021. He is the co-recipient of the Best Student Paper Award in IEEE VTC2022-Fall, the Best Paper Award in ISWCS 2022, and the 2022 IEEE SPCC-TC Best Paper Award. He serves as the Co-editor-in-Chief of IEEE ComSoc TC Newsletter, an Area Editor of IEEE Communications Letters, an Editor of IEEE Communications Surveys & Tutorials, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING. He serves as the (leading) Guest Editor for Proceedings of the IEEE on Next Generation Multiple Access, IEEE JSAC on Next Generation Multiple Access, IEEE JSTSP on Intelligent Signal Processing and Learning for Next Generation Multiple Access, and IEEE Network on Next Generation Multiple Access for 6G. He serves as the Publicity Co-Chair for IEEE VTC 2019-Fall, Symposium Co-Chair for Cognitive Radio & AI-Enabled Networks for IEEE GLOBECOM 2022 and Communication Theory for IEEE GLOBECOM 2023. He serves as the chair of Special Interest Group (SIG) in SPCC Technical Committee on Signal Processing Techniques for Next Generation Multiple Access, the vice-chair of SIG in SPCC Technical Committee on Near Field Communications for Next Generation Mobile Networks, and the vice-chair of SIG WTC on Reconﬁgurable Intelligent Surfaces for Smart Radio Environments.
Shuai Han (Senior Member, IEEE) received the M.E. and Ph.D. degrees in information and communication engineering from the Harbin Institute of Technology in 2007 and 2011, respectively. He is currently a Full Professor with the Department of Electronics and Communication Engineering, Harbin Institute of Technology. As a PI, he has more than 20 grants on wireless networks and positioning. Over his academic career, his students and he have contributed in various fields in wireless networks. He began his university studies in communication engineering at the Harbin Institute of Technology in 2000. And he completed his post-doctoral work in electrical and computer engineering at the Memorial University of Newfoundland, Canada, in 2012. He has authored or coauthored more than 100 technical papers in major journals and conferences. His research interests include wireless communications and quantum communication security, satellite IoT and integrated satellite-terrestrial communication networks. He is a member of 2020–2021 R10 Awards and Recognition Committee. He has served as a Co-Chair for technical symposia of international conference, IEEE GC 2023, ICC 2023, IEEE GC 2021, IEEE GC 2019, IEEE ICC 2018, IEEE VTC FALL 2016. He has also served as the TPC Chair for some international conferences, including the AICON2019 and MLICOM2018. He is a Senior Member of the IEEE Communication Society, the Chair of the IEEE BTS Chapter, the Vice Chair of the IEEE Harbin ComSoc Chapter, the Vice Chair of the IEEE Harbin VTS Chapter, and a Secretary of the IEEE IoT-AHSN TC. He is an Associate Editor of IEEE CHINA COMMUNICATIONS, IEEE ACCESS, Journal of Communications and Information Networks, Journal of Telemetry, Tracking and Command, and Journal of Signal Processing. He has served as a guest editor for many IEEE magazines and journals.

Marco Di Renzo (Fellow, IEEE) received the Laurea (cum laude) and Ph.D. degrees in electrical engineering from the University of L’Aquila, Italy, in 2003 and 2007, respectively, and the Habilitation Diriger des Recherches (Doctor of Science) degree from University Paris-Sud (currently Université Paris-Saclay), France, in 2013. Currently, he is a CNRS Research Director (Professor) and the Head of the Intelligent Physical Communications Group, Laboratory of Signals and Systems (L2S), CNRS, CentraleSupelec, Université Paris-Saclay, Paris, France. At Université Paris-Saclay, he serves as the Coordinator of the Communications and Networks Research Area of the Laboratory of Excellence DigiCosme, a member of the Admission and Evaluation Committee of the Ph.D. School on Information and Communication Technologies, and a member of the Evaluation Committee of the Graduate School in Computer Science. He is a Founding Member and the Academic Vice Chair of the Industry Specification Group (ISG) on Reconfigurable Intelligent Surfaces (RIS) within the European Telecommunications Standards Institute (ETSI), where he serves as the Rapporteur for the work item on communication models, channel models, and evaluation methodologies. He is a fellow of the IET and AAIA; an Ordinary Member of the European Academy of Sciences and Arts, an Ordinary Member of the Academia Europaea; and a Highly Cited Researcher. Also, he holds the 2023 France-Nokia Chair of Excellence in ICT. He was a Fulbright Fellow with The City University of New York, USA, a Nokia Foundation Visiting Professor, and a Royal Academy of Engineering Distinguished Visiting Fellow. His recent research awards include the 2021 EURASIP Best Paper Award, the 2022 IEEE COMSOC Outstanding Paper Award, the 2022 Michel Monpetit Prize conferred by the French Academy of Sciences, the 2023 EURASIP Best Paper Award, the 2023 IEEE ICC Best Paper Award (wireless), the 2023 IEEE COMSOC Fred W. Ellersick Prize, the 2023 IEEE COMSOC Heinrich Hertz Award, and the 2023 IEEE VTS James Evans Avant Grade Award. He served as the Editor-in-Chief for IEEE COMMUNICATIONS LETTERS (2019–2023). Currently, he is serving on the Advisory Board.