A State-of-the-Art Survey on Reconfigurable Intelligent Surface Assisted Non-Orthogonal Multiple Access Networks

Zhiguo Ding, Lu Lv, Fang Fang, Octavia A. Dobre, George K. Karagiannidis, Naofal Al-Dhahir, Robert Schober, and H. Vincent Poor

***Abstract*—Reconfigurable intelligent surfaces (RISs) and non-orthogonal multiple access (NOMA) have been recognized as key enabling techniques for the envisioned sixth generation (6G) of mobile communication networks. The key feature of RISs is to intelligently reconfigure the wireless propagation environment which was once considered to be fixed and untunable. The key idea of NOMA is to utilize the users' dynamic channel conditions to improve spectral efficiency and user fairness. Naturally, the two communication techniques are complementary to each other and can be integrated to cope with the challenging requirements envisioned for 6G mobile networks. This survey provides a comprehensive overview of the recent progress on the synergistic integration of RISs and NOMA. In particular, the basics of the two techniques are introduced first, and then the fundamentals of RIS-NOMA are discussed for two communication scenarios with different transceiver capabilities. Resource allocation is of paramount importance for the success of RIS assisted NOMA networks, and various approaches, including artificial intelligence (AI) empowered designs, are introduced. Security provisioning in RIS-NOMA networks is also discussed as wireless networks are prone to security attacks due to the nature of the shared wireless medium. Finally, the survey is concluded with detailed discussions of the challenges for the practical implementation of RIS-NOMA, future research directions, and emerging applications.**

I. Introduction

With the rapid rollout of the fifth generation (5G) mobile networks, the focus of the research community is currently shifting towards the design of beyond 5G (B5G) and the sixth generation (6G) networks [1]–[4]. However, the envisioned performance metrics of B5G and 6G are much more demanding than those for the previous generations of wireless systems. For example, B5G and 6G systems are expected to support users with extremely diverse data rate requirements, where an augmented reality (AR) or virtual reality (VR) user might demand a peak data rate of several Tbps, but an Internet-of-Things (IoT) node could be served with a data rate in the kbps range. Another example is that the connection density supported in B5G and 6G systems is expected to be 100 times higher than that of 5G, e.g., more than 10^7 devices per square kilometre are to be connected [1]. To meet these demanding and diverse requirements for B5G and 6G, more intelligent techniques that can efficiently increase transmission reliability, improve system throughput, and support massive connectivity are needed.

Reconfigurable intelligent surfaces (RISs) and non-orthogonal multiple access (NOMA) have been recognized as two very promising communication techniques to meet the aforementioned challenges for the design of B5G and 6G systems. On the one hand, an RIS can intelligently reconfigure a mobile user's propagation environment, which ensures that the user's data rate and reception reliability can be significantly improved [5]–[13]. Note that these performance improvements are realized in a low-cost, energy-efficient, and spectrally efficient manner, since no additional spectrum needs to be acquired and the success of RISs is due to their ability to create favourable radio propagation conditions. On the other hand, NOMA can effectively increase spectral efficiency, support massive connectivity, improve user fairness, and reduce transmission latency, by encouraging dynamic spectrum sharing among mobile users and opportunistically exploiting their heterogeneous channel conditions and quality-of-service (QoS) requirements [14]–[22].

Naturally, these two important enabling technologies are complementary to each other, where the use of NOMA can improve the spectral efficiency and connectivity of RIS systems, and the use of RISs ensures that the users' propagation environments can be effectively and intelligently customized for the implementation of NOMA. In particular, many forms of NOMA have been developed to encourage dynamic spectrum sharing among mobile users by opportunistically exploiting the users' heterogeneous channel conditions. Conventionally, a user's channel conditions are viewed as a type of fixed and non-tunable phenomenon that is solely determined by the user's propogation environment. Therefore, if the users' channel conditions are not ideal for the application of NOMA, e.g., the users have similar channel conditions, the performance gain of NOMA over conventional orthogonal multiple access (OMA) can be quite limited. RISs enable a paradigm shift for the design of intelligent NOMA, since the use of an RIS ensures that the propagation environment can be effectively and intelligently customized for the needs of NOMA.

This paper provides a comprehensive overview of the opportunities and challenges that arise in connection with the integration of RISs and NOMA in the envisioned next-generation mobile networks. In particular, we will focus on the following aspects:

- The basics of the two considered communication techniques, namely RISs and NOMA, are reviewed first. In particular, the capability of NOMA to facilitate spectrum sharing among mobile users to improve spectrum efficiency is described, and the key idea behind RISs, i.e., reconfiguration of the users' propagation environment, is illustrated.
- Then, the fundamentals of RIS-NOMA are presented, where the benefits of the integration of RIS and NOMA are unveiled. Furthermore, the quasi-degradation criterion is used as a metric to illustrate how the use of RIS guarantees that NOMA can realize the same performance as dirty paper coding, but with lower computational complexity.
- Dynamic resource allocation is crucial for realizing the performance gains enabled by RIS-NOMA and is reviewed. In particular, RIS-NOMA offers additional

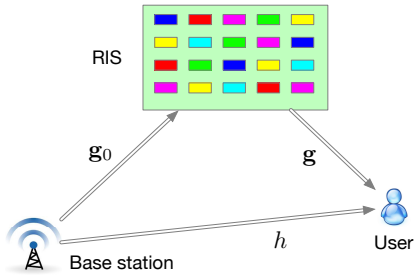


Fig. 1. An illustration example for RIS transmission.

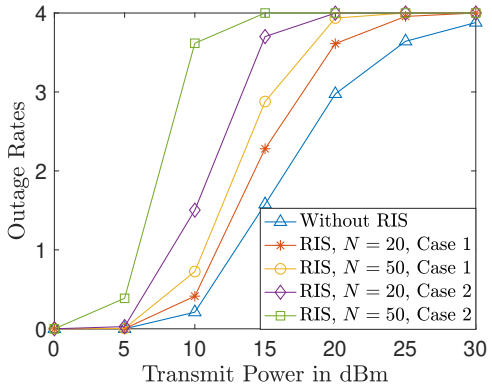


Fig. 2. Illustration of the performance of RIS transmission. The distance between the BS and the user is 15 m. For Case 1, the RIS, is located at a vertex of an equilateral triangle whose other two vertices are the user and the BS, and for Case 2, the RIS is located at the center of the aforementioned equilateral triangle. Both small scale Rayleigh fading and large scale path loss are considered. The path loss is modelled by $\frac{c_0}{d^\alpha}$, where α denotes the path loss exponent and c_0 denotes the reference power gain at the distance of 1 m. The path loss exponent for the channel between the BS and the user is 4.5, i.e., the direct link suffers severe blockage, the path loss exponents for the channels related to the RIS are 2.2, and $c_0 = -30$ dB [23]–[25]. The user's target data rate is 4 bit per channel use (BPCU). The noise power is -80 dBm.

degrees-of-freedom (DoFs) in the spatial, frequency, and time domains. We will show how these DoFs can be effectively exploited to improve the system performance by applying matching and game theoretic techniques. Artificial intelligence (AI)-empowered RIS-NOMA is presented. In particular, the benefits of using AI tools for realizing long-term performance gains in RIS-NOMA networks are illustrated, where conventional optimization tools serve as benchmarks.

- Security provisioning via RIS-NOMA is also covered in this survey. The application of RIS-NOMA to enhance physical layer security with respect to passive eavesdropping is studied first, where the impact of using an RIS on the secrecy performance is illustrated. Then, the application of RIS-NOMA to covert communications is investigated.
- While RIS-NOMA can offer various performance gains, its practical implementation faces many challenges, which are discussed here as well. In particular, the design of channel estimation schemes for RIS-NOMA networks is considered. This is a challenging problem, due to the passive nature of RIS arrays and the multi-user nature of NOMA. Low-complexity solutions for practical deployment of RISs in NOMA networks are introduced.
- RIS-NOMA has widespread applications in various communication network architectures, which are presented in this survey. In particular, the exploitation of unmanned aerial vehicles (UAVs) equipped with RISs to support aerial radio access networks is described. Other applications of RIS-NOMA, such as mobile edge computing (MEC) and simultaneously wireless information and power transfer (SWIPT), are presented in the survey as well.

The remainder of this survey is organized as follows. In Section II, the basics of RIS and NOMA are reviewed, and the fundamentals of RIS-NOMA are elaborated in Section III. Existing resource allocation approaches for RIS-NOMA are surveyed in Section IV, where AI empowered approaches are also discussed. Security provisioning and the practical implementation of RIS-NOMA are considered in Sections V and VI, respectively. The survey is concluded with a discussion of various future research directions and emerging applications for RIS-NOMA.

II. Basics of RISs and NOMA

In this section, the basics of the two considered communication techniques, RISs and NOMA, are briefly reviewed.

A. RISs

Unlike conventional information transmitters and receivers, RISs themselves do not have any information to send, but are deployed to assist information transceivers [5]–[7]. The basic idea of RIS systems can be illustrated by using the simple example shown in Fig. 1, where a single-antenna base station (BS) communicates with a single-antenna user via an RIS equipped with N reflecting elements. For this illustrative example, the signal received by the user can be expressed as follows:

$$y = (h + \mathbf{g}_0^H \mathbf{\Theta} \mathbf{g}) s + w, \tag{1}$$

where s denotes the information symbol, h denotes the channel gain between the BS and the user, $\mathbf{\Theta}$ is an $N \times N$ diagonal matrix, $\mathbf{g}_0$ denotes the channel vector between the RIS and the user, $\mathbf{g}$ denotes the channel vector between the BS and the RIS, and w denotes the noise. In the RIS literature, $\mathbf{\Theta}$ is termed the phase shifting matrix, since the RIS does not change the amplitudes of the reflected signals, but alters their phases only [8]–[10]. Denote the i-th main diagonal element of $\mathbf{\Theta}$ by $e^{j\theta_i}$, where θ_i is the phase shift caused by the i-th reflecting element on the RIS.

In Fig. 2, the performance achieved for the considered RIS-assisted single-user scenario is illustrated by using the outage rate as the performance metric. In particular, the outage rate is defined as follows: $R\left[1 - \mathrm{P}\left(\log\left(1 + \rho|\mathbf{g}_0^H \mathbf{\Theta} \mathbf{g}|^2\right) < R\right)\right]$, where $\mathrm{P}\left(\log\left(1 + \rho|\mathbf{g}_0^H \mathbf{\Theta} \mathbf{g}|^2\right) < R\right)$ denotes the probability of the outage event $\left(\log\left(1 + \rho|\mathbf{g}_0^H \mathbf{\Theta} \mathbf{g}|^2\right) < R\right)$, ρ denotes the transmit signal-to-noise ratio (SNR), and R denotes the user's target data rate. For the considered simple example, the phase shifts are obtained by first generating N sets of random phases and selecting the best set yielding the largest data rate.

Note that finding the optimal phase shifts are the key step for the design of RIS systems, and the design principles for optimizing the RIS phase shifts will be provided in detail in the following sections.

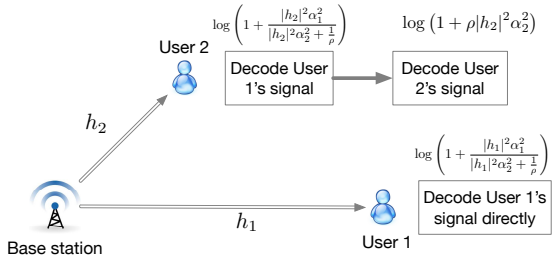
As can be observed from Fig. 2, the use of an RIS with $N = 20$ reflecting elements can already offer a significant performance gain over the scheme without RIS, e.g., with a transmit power of 20 dBm, the data rate achieved by the scheme without RIS is 1.5 bit per channel use (BPCU), but the RIS-assisted scheme can realize a data rate of 2.3 BPCU. By increasing the number of reflecting elements, the performance gain achieved with the RIS can be further improved. For example, Fig. 2 shows that the RIS with $N = 50$ reflecting elements yields two times the data rate compared to the scheme without RIS. It is also interesting to observe that the performance of the RIS assisted system is affected by the location of the RIS, where a more detailed discussions regarding how to optimize the deployment of RISs will be provided in Section VI.

B. NOMA

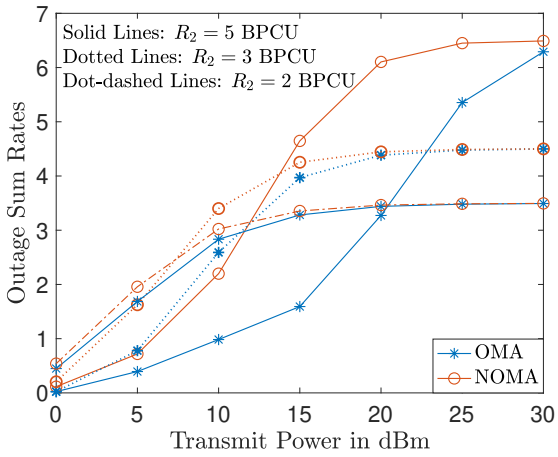
As a multiple access technique which allows multiple users to concurrently use scarce bandwidth resources, NOMA is fundamentally different from conventional OMA, which permits users to only individually occupy orthogonal bandwidth resource blocks. Instead, the key principle of NOMA transmission is to encourage spectrum sharing among users, e.g., multiple users are allowed to be served at the same time and frequency [14], [15]. The key idea behind NOMA can be illustrated based on power-domain NOMA[1] which encourages users with different channel conditions to share the spectrum [31], [32]. The success of power-domain NOMA is based on its capability to efficiently exploit the users' dynamic channel conditions, as explained in the following.

Consider a simple downlink scenario with one BS and two users, as illustrated in Fig. 3(a), where each node is equipped with a single antenna. Assume that user 1 has a weak channel gain, denoted by h_1, and user 2 has a strong channel gain, denoted by h_2, i.e., $|h_1|^2 \leq |h_2|^2$. In conventional OMA, each user occupies a different orthogonal bandwidth resource, such as a time slot, whereas in NOMA, both users are served simultaneously in the same resource. In particular, the BS sends a superimposed signal, i.e., $x = \alpha_1 s_1 + \alpha_2 s_2$, where s_i denotes user i's signal and α_i denotes the power allocation coefficient for user i's signal. Unlike conventional power allocation polices which allocate more power to users with strong channel conditions, NOMA allocates more power to the weak user, i.e., $\alpha_1 \geq \alpha_2$, in order to ensure that both users are connected. The two users adopt different detection strategies, depending on their channel conditions. As the weak user, user 1 treats user 2's signal as noise and tries to directly decode its own message with a rate of $\log\left(1 + \frac{|h_1|^2\alpha_1^2}{|h_1|^2\alpha_2^2 + \frac{1}{\rho}}\right)$. As the strong user, user 2 applies successive interference cancellation (SIC), i.e., it decodes user 1's signal first with a rate of $\log\left(1 + \frac{|h_2|^2\alpha_1^2}{|h_2|^2\alpha_2^2 + \frac{1}{\rho}}\right)$. If the first stage of SIC is carried out successfully, user 2 can remove user 1's signal and decode its own signal with a rate of $\log\left(1 + \rho|h_2|^2\alpha_2^2\right)$. On the other hand, the data rates achieved by OMA are given by $\frac{1}{2}\log\left(1 + |h_i|^2\right)$, for $i \in \{1, 2\}$, where the factor $\frac{1}{2}$ is due to the fact that each user can use the resource half of the time only.

[1]We note that there are various forms of NOMA, which exploit not only the users' different channel conditions, but also the users' heterogeneous QoS requirements [14], [15], [26]–[30].



(a) A two-user NOMA example



(b) Outage sum rates achieved by NOMA

Fig. 3. An illustrative example for NOMA. The distance between the BS and user 1 is 15 m, and the distance between the BS and user 2 is 10 m. Both small scale Rayleigh fading and large scale path loss are considered, with a path loss exponent of 4. User 1's target data rate is $R_1 = 1.5$ BPCU, and user 2's target date rate is denoted by R_2. The other simulation parameters are the same as the ones used for Fig. 2.

In Fig. 3(b), the outage sum rate is used as a metric for performance evaluation, where the outage sum rate is defined as $R_1\left(1 - \mathrm{P}_1\right) + R_2\left(1 - \mathrm{P}_2\right)$, with R_i as the user i's target data rate,

$$\mathrm{P}_1 = \mathrm{P}\left(\log\left(1 + \frac{|h_1|^2\alpha_1^2}{|h_1|^2\alpha_2^2 + \frac{1}{\rho}}\right) < R_1\right), \tag{2}$$

and

$$\mathrm{P}_2 = 1 - \mathrm{P}\left(\log\left(1 + \frac{|h_2|^2\alpha_1^2}{|h_2|^2\alpha_2^2 + \frac{1}{\rho}}\right) > R_1, \log\left(1 + \rho|h_2|^2\alpha_2^2\right) > R_2\right). \tag{3}$$

As can be observed from the figure, NOMA can yield a significant performance gain over OMA. For example, for $R_2 = 5$ BPCU and a transmit power of 15 dBm, the outage sum rate realized by NOMA is 4.6 BPCU, which is more than three times the data rate achieved by OMA. An important observation from the figure is that the performance gain of NOMA over OMA diminishes when the two users have similar target data rates. Thus, NOMA cannot only exploit the

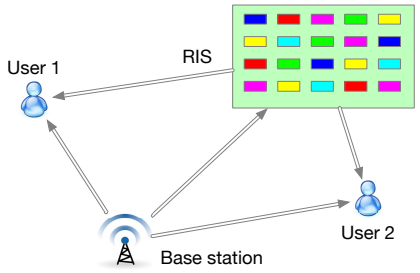


Fig. 4. Illustration for a two-user RIS-NOMA network.

users' heterogeneous channel conditions, but also other user heterogeneities, such as users' different mobility profiles and QoS requirements [14], [15], [26]–[30].

III. Fundamentals of RIS-NOMA

Although the concepts of RISs and NOMA have been developed separately in the literature, the two communication techniques are naturally complementary to each other. In this section, the fundamentals of RIS-NOMA are described and the benefits of RIS-NOMA are illustrated based on a simple two-user example. The two users are denoted by user 1 and user 2, respectively, as shown in Fig. 4. In particular, two scenarios, namely single-input single-output (SISO) and multiple-input single-output (MISO) transmission, are studied in the following subsections, respectively.

A. SISO-RIS-NOMA

Consider a SISO scenario, where each node in the network is equipped with a single antenna. For this simple scenario, the use of NOMA is clearly motivated since it can ensure that the two users are served simultaneously, whereas RIS-OMA can support only a single user at a time. The application of RISs to NOMA is motivated in the following.

Without RIS, the performance gain of NOMA over OMA is critically depending on the two users' channels, denoted by h_1 and h_2, respectively, as shown in the following. If power-domain NOMA is used and the users' channels are the same, the achievable sum rate is given by

$$\log\left(1+\frac{|h_1|^2\alpha_1^2}{|h_1|^2\alpha_2^2+\frac{1}{\rho}}\right)+\log\left(1+\rho|h_2|^2\alpha_2^2\right) \quad (4)$$
$$=\log\left(1+\rho|h_2|^2\right),$$

which is the same as the sum rate achieved by OMA.

Conventionally, the users' channel conditions are viewed as being fixed to be solely determined by the users' propagation environment. However, the use of RISs opens up opportunities for intelligently reconfiguring the users' propagation environment in order to facilitate the application of NOMA, which can lead to significant performance gains of NOMA over OMA. In particular, with RIS, the two users' effective channel gains are given by [5], [6]

$$\tilde{h}_i = h_i + \mathbf{g}^H\mathbf{\Theta}\mathbf{g}_i, \quad i\in\{1,2\}, \quad (5)$$

where $\mathbf{g}_i$ denotes the channel vector between the RISs and user i, and $\mathbf{g}$ and $\mathbf{\Theta}$ are defined in Section II. By intelligently tuning $\mathbf{\Theta}$, RIS-NOMA can introduce more DoFs for system design. This can be exploited to not only generate a significant difference in the users' channel gains but also to customize

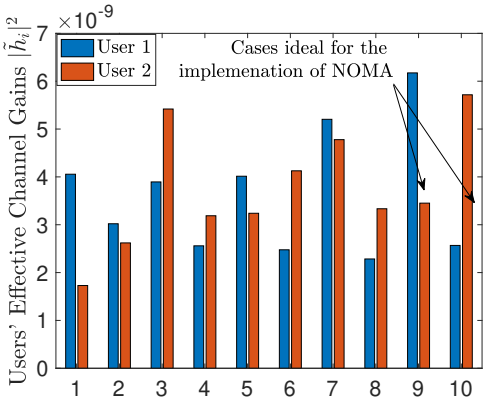
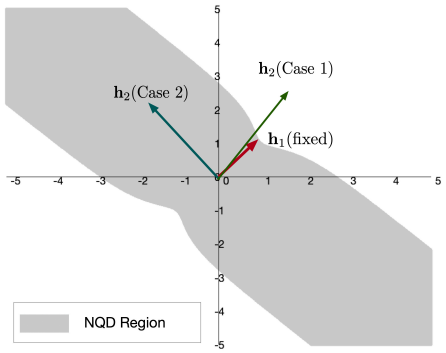


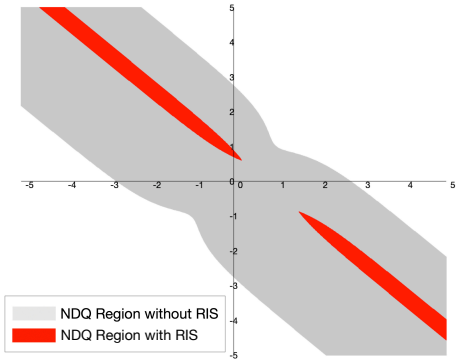
Fig. 5. Illustration of the benefit of using SISO-RIS-NOMA with $N = 100$. The BS and the two users are located at the vertices of an equilateral triangle with side length $d = 15$ m, and the RIS is located at the center of the triangle. For illustrative purposes, small scale fading is omitted for the users' direct channels to the BS, such that the users have the same channel gains, $h_1 = h_2$, because their distances to the BS are the same. The other simulation parameters are the same as the ones used for Fig. 2. The 10 cases are generated by using 10 random phase shifts.

the users' effective channel gains according to the users' QoS requirements, as shown in Fig. 5, where it is assumed that $h_1 = h_2$. As discussed before, this situation is not suitable for the application of power-domain NOMA. As can be observed from the figure, the use of RISs can effectively generate sufficient difference between the users' effective channel gains. For example, for case 9 shown in Fig. 5, user 1's channel gain is about 2 times larger than user 2's channel gain, which means that user 1 can act as the strong user and user 2 can act as the weak user in power domain NOMA. If user 2 has a more stringent target data rate and needs to be treated as the strong user, Fig. 5 shows that this can be realized by simply adjusting $\mathbf{\Theta}$.

The fundamentals of SISO-RIS-NOMA have been thoroughly analyzed in the literature from various perspectives. For example, in [33], the energy efficiency of downlink SISO-RIS-NOMA has been studied, where SISO-RIS-OMA has been used as a benchmark. As shown in [33], when the users have different data rate requirements, the use of RIS-NOMA can realize a reduction in the transmit power by almost a factor of two, even if the users have similar channel conditions, whereas the performance gain of RIS-NOMA diminishes if the users have similar target data rate requirements. Another performance criterion used in the literature is reception reliability, where the works in [34], [35] demonstrate that downlink SISO-RIS-NOMA can efficiently utilize the spatial DoFs to improve the reception reliability, compared to RIS-OMA. In particular, the use of SISO-RIS-NOMA can ensure that the diversity gain achieved by each user in the system is proportional to the number of reflecting elements of the RIS, even though the multiple users share the same bandwidth resource. The application of SISO-RIS-NOMA for uplink transmission has been studied in [36], where the sum rate is used as the metric. In particular, the authors of [36] have shown that the use of RIS-NOMA can yield significant performance improvement when the number of RIS elements is large. Another important finding in [36] is that the use of NOMA is particularly useful in crowded scenarios, e.g., the number of users a performance gain of 5 BPCU over OMA is achievable for the two-user case, and it can be increased to 8 BPCU when there are 6



(a) NQD without RISs



(b) An RIS with 10 reflecting elements

Fig. 6. Improvement of the quasi-degradation region by using RIS, with $\mathbf{h}_1 = [1 \quad 1]^T$ and the two users' target data rates being 1 BPCU. $\mathbf{G}$ and $\mathbf{g}_i$, $i \in \{1, 2\}$, are randomly generated. NQD represents non-quasi-degradation.

users.

B. MISO-RIS-NOMA

Consider an MISO scenario, where the BS is equipped with multiple antennas and each user has a single antenna. The benefit of the integration of RISs and NOMA can be clearly illustrated by using the quasi-degradation criterion, an insightful condition under which the use of MISO-NOMA results in the same performance as dirty paper coding (DPC) but with less complexity [37], [38].

The quasi-degradation criterion can be illustrated by considering a simple two-user downlink scenario, where the BS has two antennas, and each user has a single antenna. The two users' channel vectors are denoted by $\mathbf{h}_i$ and assumed to be areal-valued vectors. Furthermore, assume that user 1's channel vector is fixed, i.e., $\mathbf{h}_1 = \begin{bmatrix}1 & 1\end{bmatrix}^T$. The shaded region shown in Fig. 6(a) is obtained by highlighting the non-quasi-degradation (NQD) region, i.e., those realizations of $\mathbf{h}_2$ which do not meet the quasi-degradation condition [37], [38]. For example, Case 1 shown in the figure, i.e., the two users' channel vectors are almost aligned, is in the quasi-degradation region. This is an ideal case for using NOMA, since the users' channel vectors occupy a single spatial direction and one beam can be used to serve both users simultaneously. Case 2, i.e., the two users have orthogonal channels, is in the NQD region. This means that the use of NOMA for this case may result in a throughput loss compared to DPC, even though the use of NOMA can still offer other advantages, especially in overloaded scenarios in which the number of users exceeds the available resources. By using RISs and tuning the RIS phase shifting matrix, it is possible to reconfigure the users' channel vectors, such that they satisfy the quasi-degradation criterion. For example, Fig. 6(b) reveals that by using an RIS with 10 reflecting elements, the NQD region is diminishing, compared to the case without an RIS shown in Fig. 6(a). In other words, if there are a sufficient number of reflecting elements on the RIS, it can be ensured that the use of RIS-NOMA yields the same performance as DPC, regardless of what the users' original channel conditions are. Furthermore, since both NOMA users occupy only a single spatial dimension, the remaining dimensions can be used to accommodate additional users.

A more formal study of the impact of RISs on the quasi-degradation criterion was provided in [39], where a new expression of the quasi-degradation criterion for MISO-RIS-NOMA was developed. While the quasi-degradation criterion sheds light on whether NOMA can realize the same performance as DPC, it still has a few limitations. For example, the expressions for the quasi-degradation criterion reported in [39] have been developed by fixing the SIC decoding order, whereas recent studies have shown that opportunistically choosing SIC decoding orders can yield significant performance improvement in NOMA networks [40], [41]. Furthermore, the criterion has been investigated mainly for the simple two-user case, and its generalization to the more important overloaded cases, where the number of users exceeds the number of antennas at the BS, is still unknown.

Compared to SISO-RIS-NOMA and MISO-RIS-NOMA, the fundamentals of RIS-NOMA for more general setups, e.g., when there are multiple RISs or/and when each node is equipped with multiple antennas, have been less investigated in the literature, because of their challenging nature. One promising approach is to use RIS-NOMA as a type of add-ons, and combine it with conventional spatial division multiple access (SDMA). Figure 7 illustrates this approach, where conventional SDMA is used to form orthogonal beams and serve multiple primary users. Assume that there are additional secondary users to be connected. Conventionally, it is not possible to serve these additional users directly with the existing spatial beams, since the secondary users' channels may not be perfectly aligned with the primary users' channels due to the random radio propagation environment. The use of RISs opens the possibility to reconfigure the propagation environment and align the secondary users' channel vectors with the existing SDMA beams. As a result, the additional secondary users can be served without changing the SDMA legacy system, as shown in Fig. 7. In the next section, more sophisticated designs for RIS-NOMA will be considered by exploiting optimization based resource allocation.

IV. Dynamic Resource Allocation for RIS-NOMA

Driven by the benefits of RIS-NOMA systems, such as their high spectral/energy efficiency, resource allocation/optimization has attracted significant research attention

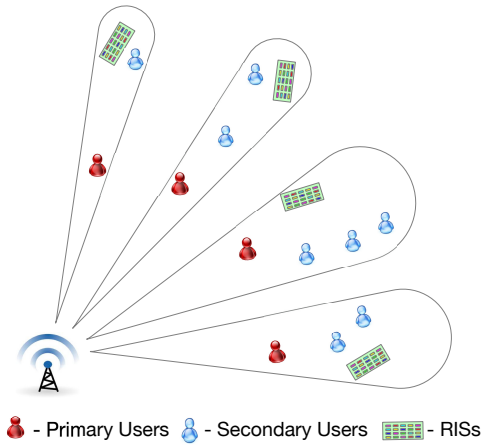


Fig. 7. Illustration of the benefit of RIS-NOMA, where primary users are served via conventional SDMA and the use of RIS-NOMA ensures that additional secondary users can be served by an existing SDMA beam simultaneously.

to further improve the communication performance. In particular, by jointly optimizing the communication resources, e.g., beamforming, power, and sub-channels, and the reflecting coefficients of each RIS, the strength of the received signal can be significantly improved. In this section, we review traditional resource allocation methods, the related optimization problems, and their solutions in the context of RIS-NOMA systems, including sum rate maximization, spectral/energy efficiency maximization, power minimization and physical secure communications issues, as summarized in Table I. Subsequently, we discuss AI/ML-based resource allocation in RIS-NOMA.

A. Throughput/Data Rate Maximization

Due to the non-convexity of the optimization problems formulated for throughput maximization, the alternating optimization algorithm proposed in [42] is widely adopted to design the transmit beamforming at the BSs and the RIS phase shifts in an alternating manner. Let us start with a simple system model, i.e., a SISO system. The average sum rate of a two-user downlink SISO-RIS-NOMA network was maximized by alternatingly applying a scheme with two phase shift adjustments (namely, dynamic phase adjustment and one-time phase adjustment) and a power allocation algorithm [43]. Considering a wireless powered RIS-NOMA system with multiple devices/users, the authors in [44] first proved that dynamic phase shifts are not needed for downlink NOMA and uplink NOMA systems, which simplifies the problem and reduces the signalling overhead. In addition to the optimization of the beamforming at the BS and the phase shifts of the RIS, the channel assignment and decoding order of the NOMA users were optimized to maximize the system throughput [45]. To compare three multiple access technologies, namely NOMA, frequency division multiple access (FDMA), and time division multiple access (TDMA), the weighted sum rate was studied for RIS systems in [46]. In this work, a joint optimization algorithm for the deployment location and the reflection coefficients of the RIS as well as the power allocation was proposed and shown to achieve near-optimal performance. Furthermore, the use of a BS equipped with multiple antennas was considered in RIS-NOMA systems [47], where a scheme for jointly optimizing active beamforming at the BS and the passive beamforming at the RIS was proposed in order to achieve the maximum sum rate. In addition to downlink case, the sum rate maximization problem was also studied for uplink RIS-NOMA systems [36].

RIS-NOMA is highly compatible with other key technologies for 6G communication systems, including mmWave [48], [49] and massive MIMO [50]. The authors in [48] considered the design of mmWave networks and proposed an optimization algorithm to maximize the system sum rate, where the power allocation and reflection coefficients are optimized alternately. The results shown in [48] illustrate the effectiveness of integrating RISs in mmWave-NOMA systems. Recall that the small wavelength of mmWave signals enables the deployment of a large number of antennas, which motivates the design of mmWave massive MIMO systems. [49] considered the application of RIS-NOMA in mmWave massive MIMO systems, where the effects of the power leakage and the per-antenna power constraint have been joint investigated. In this work, two multi-beam selection strategies were proposed to maximize the system weighted sum rate by jointly optimizing the active beamforming at the BS and the passive beamforming at the RIS. Regarding imperfect SIC in NOMA, the capabilities of RISs to manipulate wave polarization in dual-polarized MIMO-NOMA networks was investigated in [50], where the proposed novel strategy can alleviate the impact of imperfect SIC and exploit polarization diversity.

B. Power/Energy Efficient Resource Allocation

In future wireless communication networks, power efficiency and energy efficiency are important metrics for evaluating the network performance. For transmit power minimization, the performance of NOMA and OMA was compared in [33], which provide an important guideline for user pairing in RIS-aided systems with large numbers of users and resource blocks. Considering a single-cell RIS-NOMA network with multiple users, the authors in [53] proposed a novel difference-of-convex (DC) programming algorithm for the design of beamforming and phase shifts in order to minimize the total transmit power. To reduce the decoding complexity of NOMA transmission, users are generally grouped into small-size clusters in practice. Thus, a multi-cluster MISO RIS-NOMA system was studied in [54] aiming to minimize the total transmit power. In this work, a beamforming and phase shift optimization scheme based on second-order cone programming (SOCP) and alternating direction method of multipliers (ADMM) was proposed to minimize the total transmit power at the BS. Different from single-RIS assisted NOMA networks [39], a multi-RIS and multi-cluster NOMA network was investigated in [55], where the joint optimization of the beamforming at the BS, the power allocation to the NOMA users, and the phase shifts of the RISs was proposed to minimize transmit power at BS.

Energy efficiency (EE) is another important metric for future green communication networks. EE is defined as the ratio between the amount of information bits delivered and the amount of energy consumed. In particular, different from sum-

TABLE I
SUMMARY OF RESOURCE ALLOCATION SCHEMES

Category	System Model (Reference)	Optimization Variables	Main Results
Throughput or Data Rate Maximization	SISO-RIS-NOMA [43]	Power & phase shift	Proposes two phase shift adjustments to maximize the average sum rate of two users
	SISO-RIS-NOMA [45]	Power & phase shift & channel allocation & decoding order	Proposes matching-based channel allocation and a joint optimization algorithm to maximize the system throughput
	SISO-RIS-NOMA [46]	Deployment location & reflection coefficients of RIS & power allocation	Compares NOMA, FDMA, and TDMA with respect to their weighted sum rates
	MISO-RIS-NOMA [47]	Active beamforming at BS & phase shifts of RISs	Proposes an efficient alternating algorithm to maximize the system sum rate by considering ideal RIS and non-ideal scenarios
	SISO-RIS-NOMA [36]	Active beamforming at BS & phase shifts of RISs	Provides a near-optimal solution to maximize the uplink sum rate
	mmWave aided RIS-NOMA [48]	Active beamforming at BS & phase shifts of RISs & power allocation	Confirms the effectiveness of introducing RISs in mmWave-NOMA systems
	mmWave aided massive MIMO RIS-NOMA [49]	Beam selection & active beamforming at BS & phase shifts of RISs	Confirms the weighted sum rate performance gain
	Massive MIMO RIS-NOMA [50]	Phase shifts of RISs	Confirms that dual-polarized RISs can facilitate cross-polar transmissions
	UAV-aided RIS-NOMA [51]	Placement and transmit power of UAV (RIS) & phase shifts of RIS	Provides a simple UAV's horizontal position design and beamforming and the phase shift optimization to maximize the strong user's data rate
	Multi-UAVs aided multi-user RIS-NOMA [52]	Placement and transmit power of UAVs & phase shifts of RISs & decoding order of NOMA users	Proposes a block coordinate descent (BCD)-based iterative algorithm to maximize the sum rate for a multi-UAV enabled NOMA-RIS network
Power Minimization	SISO-RIS-NOMA [33]	Active beamforming at BS & phase shifts of RISs	Provides a guideline for user pairing in RIS-NOMA systems with a large number of users and resource blocks
	MISO-RIS-NOMA [53]	Active beamforming at BS & phase shifts of RISs	Proposes an alternating DC method to minimize transmit power
	MISO-RIS-NOMA [54]	Active beamforming at BS & phase shifts of RISs	Proposes SOCP-ADMM based algorithm to achieve significant performance gains over conventional semidefinite programming (SDP) based algorithm
	MISO-RIS-NOMA [55]	Active beamforming at BS & phase shifts of RISs & power allocation coefficients	Proposes a joint optimization of the beamforming, the power allocation and the phase shifts of RISs for a multi-RIS aided NOMA system
	MISO-RIS-NOMA [39]	Active beamforming at BS & phase shifts of RISs	Provides a closed-form beamforming design to achieve the same performance as dirty paper coding
EE Maximization	MISO-RIS-NOMA [56]	Active beamforming at BS & phase shifts of RISs	Provides a simple beamforming and phase shift optimization scheme to maximize the system energy efficiency

rate maximization and transmit power minimization, the aim of EE maximization is to achieve the optimal tradeoff between sum-rate maximization and power minimization, which leads to a fractional programming problem. Generally, the Dinkelbach algorithm can be used to iteratively achieve the system maximum EE by transforming the fractional form objective function into a subtractive-form one, which has been well investigated for NOMA systems [57]. Moreover, DC programming can also be utilized to maximize the EE of NOMA systems with matching theory-based subchannel allocation [58]. Driven by the benefits of RIS-NOMA, the EE maximization problem was first studied for two-user RIS-NOMA networks [56], where successive convex approximation (SCA) based beamforming optimization and SDR-based phase shift optimization schemes were proposed for maximization of the EE. For complex RIS-NOMA networks, such as multi-cell systems, game theory and matching theory are efficient tools to address the user association, user clustering, and subchannel allocation problems with a lower complexity compared to an exhaustive search. This can be combined with convex optimization and ML techniques for resource optimization in RIS-NOMA systems.

C. AI/ML-based Beamforming Optimization for RIS-NOMA

As one of the most promising technologies in 6G, AI includes advanced ML techniques, such as supervised learning, unsupervised learning, and reinforcement learning (RL), to smartly address the dynamic and uncertain environments in RIS-NOMA systems. ML techniques, especially RL, can be exploited for beamforming optimization, phase shift optimization, and power allocation in RIS-NOMA systems. To reduce the high computation complexity of the traditional resource allocation methods such as SDR and SCA, deep learning (DL) and RL can be utilized for optimization in RIS networks. In particular, DL can be used to optimize the RIS reflection matrix based on the sampled channel knowledge [59]. A deep neural network (DNN) was proposed for RIS configuration in an indoor RIS communication system [60]. In this work, the trained DNN is fed with the measured position information of the target user to determine the optimal phase configurations of the RIS.

Intergrating NOMA into RIS can further improve the spectral efficiency and massive connectivity. However, as decoding complexity increases with the number of users, user grouping/clustering are widely adopted in NOMA systems, especially dynamic user grouping/clustering policies. Recently, intelligent user grouping/clustering methods, such as deep Q-network (DQN) aided user clustering [61], have been proposed for NOMA systems to adapt to dynamic network conditions. Regarding the deployment design of MISO-RIS-NOMA networks, a decaying double deep Q-network ($\mathrm{D^3QN}$) based algorithm was proposed to optimize the position and phase shifts of an RIS in [62]. In this work, the EE was maximized by the proposed $\mathrm{D^3QN}$ method, which is capable of overcoming the issue of large overestimation of action values in conventional Q-learning.

To support ultra-fast communication in RIS-NOMA networks, the resource allocation, including the user grouping/clustering, subchannel assignment, beamforming design, power allocation, and phase shift optimization, must embrace intelligence to adapt to the fast time-varying wireless environment. Therefore, ML algorithms, such as DQN, have been utilized to provide smart control for RISs in NOMA networks [63]. In particular, in a MISO-RIS-NOMA system, a K-means based method and DQN based algorithm were proposed [63] for user clustering and the joint design of the phase shift matrix and the power allocation, respectively. However, DQN can only handle discrete actions, such as quantized phase shifts. As a sophisticated deep RL method, deep deterministic policy gradient (DDPG) combines DQN and the policy gradient algorithm, and is applicable to the DQN problem with continuous actions, such as continuous power allocation. DDPG is an effective actor-critic, model-free RL algorithm and is applicable to communication problems modelled based on Markov decision process. There are three main elements in RL, i.e., the state space (wireless environment), action space (power allocation), and reward (sum rate). The agent (generally the BS) interacts with the dynamic wireless environment to generate good actions to maximize the long-term reward of the system, such as the sum rate. Thus, the phase shifts of RISs can be designed by DDPG to maximize the system sum rate in RIS-NOMA [64], [65].

V. Security Provisioning for RIS-NOMA

Thanks to their potential of intelligently reconfiguring the wireless propagation environment, RISs have also been applied in NOMA networks to improve wireless communication security and privacy. In the following, physical layer security and covert communication in connection with RIS-NOMA are studied, respectively.

A. Physical Layer Security With RIS-NOMA

In conventional NOMA networks, guaranteeing perfect security (i.e., having a positive secrecy rate) for all users is a challenging task [66]. Consider a multiuser NOMA scenario, where the quality of the channels from a transmitter to some of the legitimate users is worse than that of the wiretap channel from the transmitter to the eavesdropper. This situation can happen frequently in practical NOMA networks, since users generally have different geographical locations and the eavesdropper may be closer to the transmitter. In this scenario, positive secrecy rates cannot be guaranteed if the transmitter and these users have one antenna only and cannot employ secure beamforming. This thus motivates the application of RISs to NOMA for physical layer security enhancement, the benefits of which are summarized next.

1) Improving Secrecy Rate: Theoretically, by deploying the RIS in the vicinity of the legitimate NOMA users or the eavesdropper and by properly designing the passive RIS beamforming, the signals received from the direct and reflected paths can be added constructively at the users but destructively at the eavesdropper [67]–[70]. Therefore, the legitimate signal reception quality is enhanced while the eavesdropping capability is degraded, and an increased secrecy rate for RIS-NOMA can be achieved. Note that the design of active beamforming at the transmitter and passive beamforming at the RIS depends on the availability of channel state information (CSI). Compared to a scenario without secrecy considerations, enabling an efficient beamforming design to ensure RIS-NOMA communication security requires additional CSI knowledge, i.e., the eavesdropper's CSI (refered to as ECSI) is required. However, the instantaneous ECSI is difficult to obtain in practice since i) the eavesdropper may intentionally remain silent and only its location information may be known corresponding to statistical ECSI, and ii) although the local oscillator power leakage from the eavesdropper's radio frequency front-end can be used for ECSI estimation, the acquired ECSI is expected to be outdated and imperfect. In [71], a robust secure beamforming strategy with artificial noise (AN) injection for RIS-NOMA was proposed by exploiting imperfect ECSI. The obtained results reveal that even with RIS, the use of AN is still helpful in degrading the eavesdropper's capability and improving the secrecy rate, particularly when the quality-of-service (QoS) requirements of the users are stringent. If only statistical CSI is available, it is not possible to maximize the secrecy rate directly. As an alternative, one can utilize the secrecy outage probability as the performance metric and impose an upper bound on it to limit the eavesdropping. In [72], a joint transmit and RIS beamforming scheme for a distributed RIS deployment was proposed for RIS-NOMA, where the minimum non-adaptive secrecy rate (which is fixed and predetermined) among all users was maximized subject to a secrecy outage probability constraint. A significant secrecy rate improvement was obtained compared to the scenario without RIS [72].

2) Creating Signal or Constellation Overlapping: In fact, the intentionally superimposed transmission of multiple signals via the RIS can result in desirable signal overlapping at the eavesdropper, and therefore, the eavesdropper has to adopt interference decoding and its signal reception quality is degraded. This can be demonstrated for a secure RIS-enabled two-way communication scenario, which can be regarded as a special case of NOMA [73]. As depicted in Fig. 8(a), the two users simultaneously transmit their signals following the uplink NOMA principle and the RIS reflects the incident signals. On the one hand, each user can remove the interfering signal (a copy of which has been transmitted by the user itself)

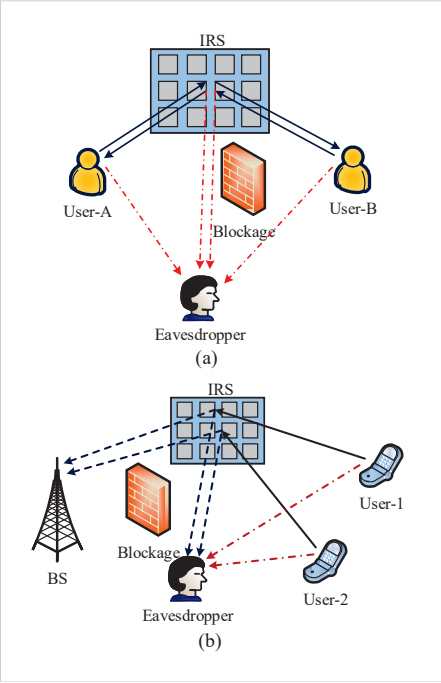


Fig. 8. Physical layer security with RIS-NOMA. (a) Signal overlapping. (b) Constellation overlapping.

via self-interference cancellation without degrading its own signal reception quality. On the other hand, the eavesdropper receives a superimposed signal version and it has to decode one signal by treating the other signal as interference. Such an interfering signal can be used as a useful jamming signal to confuse the eavesdropper. Thus, by increasing the transmit power and/or the number of RIS's reflecting elements, the average secrecy rate can be improved [73]. Furthermore, the fundamental operating principle of RISs opens up a new design opportunity, i.e., causing the inter-user interference in NOMA to be constructively and destructively received at the legitimate user and the eavesdropper, respectively. To clarify this, we consider the uplink NOMA scenario shown in Fig. 8(b). Two users transmit signals simultaneously to the BS via an RIS, where the RIS adjusts its phase to guarantee that the users' signal constellations fully overlap at the eavesdropper but do not non-overlap at the BS [74]. Thus, even if the eavesdropper may have a high receive SNR to decode the superimposed signal successfully, the individual signals have to be recovered by random guess, which results in a poor bit error rate performance and degrades the eavesdropping capability significantly. To this end, a joint design of the power allocation and RIS beamforming is needed, which is an interesting direction for future research.

3) Generating Artificial Jamming Signals: Different from the conventional solutions that utilize RISs to reflect the confidential signal and the AN (which is generated by the transmitter) for physical layer security [75], [76], if appropriately modulated, the RIS can mimic a backscatter transmitter to generate artificial jamming when the signal is reflected to degrade the eavesdropper's reception, which is referred to as RIS-aided backscatter jamming [77]. By carefully optimizing the RIS's reflection coefficients, the information leakage to the eavesdropper can be minimized and reliable legitimate transmission can be guaranteed. This strategy is particularly appealing due to the fact that the transmission security is achieved without the consumption of additional communication resources to construct the AN, but by reusing the existing confidential signal and modulating it into a helpful jamming signal via the RIS to confuse the eavesdropper. Considering this merit, RIS-aided backscatter jamming is expected to be employing in NOMA networks for secrecy enhancement. However, in NOMA networks, inter-user interference and backscatter jamming co-exist, and both undoubtedly deteriorate the legitimate users' reception. Thus, extra efforts are required for interference management to balance communication reliability and security. One possible approach is to operate the RIS in a hybrid signal processing mode, i.e., partitioning the elements of the RIS into two groups: the elements in one group work in the reflection mode to improve the legitimate reception quality while the elements in the other group work in the backscatter mode to generate the jamming signal for degradation of the eavesdropper. Based on this, future works may jointly optimize the element-wise mode selection and the corresponding modulation and reflection coefficients to maximize the overall system performance.

B. Covert Communication With RIS-NOMA

Under certain circumstances, safeguarding the content of communication by physical layer security techniques is not sufficient, and the communication entity may want to hide itself from being detected. This calls for covert wireless communication, which aims at hiding the amount of confidential data that is exchanged between a transmitter and a receiver, subject to a negligible detection probability of a warden. Several existing research works have shown that both NOMA and RISs can benefit communication covertness, since 1) with NOMA, the transmission of a non-covert user can serve as a shield for a covert user [78], and 2) the RIS enables signal strength boosting at the receiver and signal cancellation at the warden without consuming additional communication resources, thus guaranteeing green covertness [79].

However, solely using NOMA or RISs for covert communication has some disadvantages. In particular, other uncertainty sources such as a random transmit power [78] or noise uncertainty at the warden [79] are required to achieve covertness. Assume a transmitter communicates with a receiver by applying RIS assisted OMA or NOMA without RIS. Covertness cannot be achieved in either scenario since in both cases, the warden can exactly measure its non-zero received power to detect the covert transmission. This thus motivates the integration of NOMA and RISs towards covert communication, where the NOMA signaling of a non-covert user and the phase-shift uncertainty of the RIS are jointly exploited as the new cover medium to hide the covert user's transmission [80]. As validated in [80], by appropriately generating a random phase shift pattern at the RIS, uncertainties exist in the warden's received signal power. In fact, the detection error probability of the warden increases with the transmit power and the number of RIS elements.

VI. Practical Implementation of RIS-NOMA

In this section, two practical implementation issues for RIS-NOMA, namely channel estimation and RIS deployment, are discussed in detail.

A. Channel Estimation

To fully reap the various performance gains brought by RIS-NOMA, efficient channel estimation schemes are required since both active transmit beamforming at the BS and passive reflection beamforming at the RIS rely on CSI. However, this is a non-trivial task, due to the fact that i) the RIS does not have any RF chains to enable pilot signal transmission for channel estimation, and ii) there is a large number of RIS-related channel coefficients, due to the large number of RIS elements and the multi-user nature of NOMA, that need to be estimated. In the literature, there are two kinds of RIS channel estimation schemes depending on whether the RIS is integrated with sensing devices or not [81], as elaborated in the following.

1) Semi-Passive RIS Channel Estimation: In this scheme, additional sensing devices (e.g., low-cost sensors) are integrated into the reflecting elements of the RIS. Hence, by letting the BS and each user transmit pilot signals, the channels from the BS/users to the RIS can be estimated based on their received pilots, and the corresponding reverse channels can be easily obtained by exploiting channel reciprocity in a time-division duplexing (TDD) system. To reduce the cost and energy consumption of semi-passive RIS, the number of integrated sensors is usually less than that of the reflecting elements. This indicates that the estimated channels are only a subset of all channels, and the available estimated channels are to be used to recover the high-dimensional full channels. In fact, there may exist sparsity or spatial correlations between the estimated and the actual channels. To this end, advanced signal processing techniques, e.g., compressed sensing and machine learning, can be applied to solve this problem. In [59], [82], compressed sensing and deep learning based semi-passive RIS channel estimation schemes were proposed, where the channel sparsity and the mapping information of the sampled channels were exploited to construct the full channels. The results show that the proposed schemes achieve a comparable rate performance as the case with perfect channel knowledge, while employing a negligible training overhead and a small number of active sensors. By assuming sparse channels in the beamspace domain, the authors in [83] developed an efficient AO method to explicitly estimate the channels of the RIS elements attached to a single RF chain, where the impact of the channel sensing time on the mean-squared error (MSE) performance was analytically evaluated. Note that the above solution approaches are only applicable in TDD systems, while the design of semi-passive RIS channel estimation schemes in frequency-division duplexing systems is still an open issue.

2) Fully-Passive RIS Channel Estimation: If the RIS is fully passive and does not possess any active sensors, it is infeasible to estimate the CSI between the BS/users and the RIS directly. As an alternative approach, one can estimate the cascaded BS-RIS-user channels instead. Basically, the cascaded channel knowledge can be estimated following an on/off RIS reflection pattern based scheme [84], where in each training signal transmission, only one RIS reflection element is turned on while the others are turned off, and this process is carried out until all the RIS reflection elements have been turned on. As a result, the cascaded BS-RIS-user CSI for each reflecting element is obtained in a sequential manner. We note that this scheme inevitably incurs a very large training overhead for pilot signal transmission, i.e., N pilot symbols are needed if the RIS is equipped with N reflecting elements. To overcome this issue, an RIS reflecting element grouping aided channel estimation scheme was proposed in [85], [86], where adjacent RIS reflecting elements were grouped into a sub-surface, as channels associated with the same sub-surface are usually spatially correlated. As such, only the cascaded channels for each sub-group have to be estimated, which largely reduces the training overhead. The above fully-passive channel estimation schemes consider only a single user, and need to be suitably modified to fit multi-user NOMA networks. Intuitively, one can estimate the cascaded channels for K users by sequentially using the single-user schemes K times, which, however, significantly increases the training overhead especially when K is large. A potential approach is to treat the K users as one equivalent K-antenna user and exploit certain RIS channel properties (e.g., low-rank, spatial correlation, and sparsity) to facilitate the cascaded channel matrix decomposition for channel estimation [87]–[89]. Another promising approach is to exploit the fact that all the users share the same BS-RIS channel, and hence, one user can be selected as the reference user and her cascaded channel is estimated first. Subsequently, the remaining $K-1$ users' cascaded channels can be obtained by individually scaling the reference user's cascaded channel [81]. The optimal user selection criterion and a mechanism to learn the relationship between the users' cascaded channels should be further investigated.

B. Deployment Strategy

From an implementation perspective, the RIS deployment plays an important role for boosting the performance of NOMA networks. Given a fixed number of the reflecting elements, the RIS can be either deployed in a *centralized* manner with all the reflecting elements forming a large RIS located close to the BS-/user-side, or in a *distributed* manner where the reflecting elements form multiple RISs and each is located near one of the users [81], as shown in Fig. 9. Thus far, there is no consensus whether the centralized RIS deployment or the distributed RIS deployment is more preferable for NOMA networks, and in fact, both deployment strategies have their own benefits and drawbacks.

- *Centralized RIS deployment*: This strategy is appealing for the cluster-based NOMA scenario where the direct communication links from the BS to the users are highly correlated (e.g., in mmWave and THz communications) [90]. In this scenario, the BS may not have enough DoFs to enable user clustering via the transmit beamforming. This motivates the use of a centralized RIS to provide additional DoFs. Theoretically, a centralized RIS has a large number of reflecting elements to facilitate

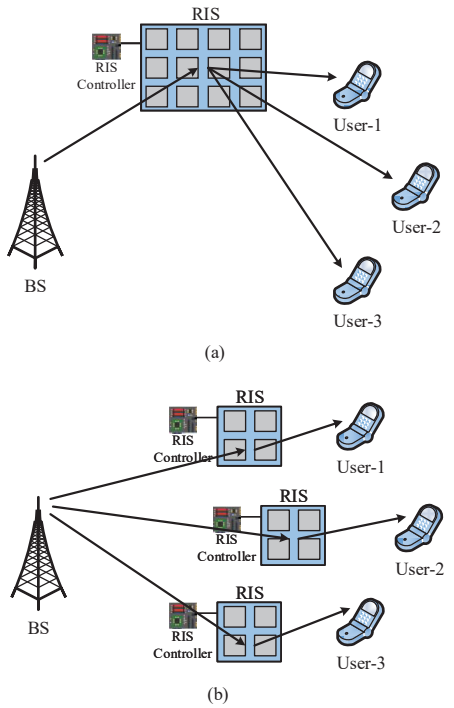


Fig. 9. RIS deployment strategies. (a) Centralized. (b) Distributed.

beamforming, and a high passive array gain is obtained for user clustering [90]. As indicated in [91], the gain of NOMA over OMA is more pronounced when the users have distinct channel conditions. The centralized RIS deployment can help achieve this. By deploying the RIS in an asymmetric manner to enlarge the disparities among the users' channels, a larger sum rate of NOMA is achieved compared to OMA [46]. Furthermore, for a special two-user NOMA scenario under the twin channel condition (i.e., the BS-RIS-user channels are similar to each other), it was shown in [92] that the centralized RIS deployment always offers a larger multi-user achievable rate region than the distributed RIS deployment. However, considering the size and hardware constraints, it is not always feasible to deploy a huge centralized RIS, while deploying multiple moderate or small-size RISs at their respective user clusters is more manageable [81].

- *Distributed RIS deployment*: This strategy is ideal for the scenario where the locations of the users are clustered. The advantage of applying distributed RISs in this scenario lies in the fact that different NOMA user clusters are associated with different RISs and are separated by large distances, and thus, the signals reflected by one RIS have little impact on the reception performance of other unintended user clusters [90]. Moreover, with proper placement, the distributed RIS deployment will very likely create more line-of-sight (LoS) paths between the BS and the users compared to the centralized RIS deployment, when users are randomly located [35]. Despite the above merits, the distributed RIS deployment suffers from a high signaling overhead since the RIS phase-shift design requires information exchange between the BS and multiple RIS controllers.

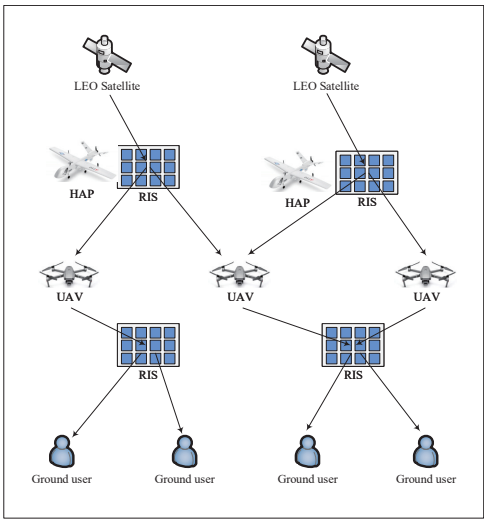


Fig. 10. Application of RIS-NOMA to aerial radio access.

To be specific, relying on centralized or distributed RIS deployment alone may not satisfy all the requirements of NOMA transmission, and a more general hybrid RIS deployment may be desirable [93]. In the hybrid design, a large centralized RIS is deployed close to the BS to provide a high passive beamforming gain, while multiple small RISs are deployed close to the users to support multiple LoS channels. Hence, the hybrid RIS deployment combines the complementary advantages of the centralized and distributed strategies and can achieve a tradeoff between them. A joint optimization framework for transmit beamforming, all RISs' reflection beamforming, and RIS element allocation (i.e., for a fixed number of RIS elements, the number of elements assigned to the centralized RIS and each distributed RIS) is required to improve the overall system performance. This requires further investigation in the future.

VII. Future Directions and Emerging Applications of RIS-NOMA

In this section, we explore the potential of employing RIS-NOMA in various emerging wireless network architecture.

A. RIS-NOMA Assisted Aerial Radio Access Networks

Aerial radio access networks are a promising strategy to complement the conventional terrestrial cellular communications [94]. Comprising airborne components such as unmanned aerial vehicles (UAVs), high-altitude platforms (HAPs), and low earth orbit (LEO) satellites, aerial radio access networks can quickly establish a flexible access infrastructure and support the development of seamless 6G mobile communication systems, where the use of RISs and NOMA can be particularly important as shown in Fig. 10.

Being equipped with communication and signal processing capabilities, UAVs can act as moving BSs or relays to facilitate reliable and efficient communication with multiple users [95], [96]. When the users are roaming continuously, it is challenging for the UAV to periodically reposition itself according to the mobility of the users (as this could consume significant amounts of energy) since the UAV is usually powered by batteries without constant energy supply. RISs can be exploited

to tackle this challenge. Theoretically, one can adjust the phase shifts of the RIS instead of controlling the movement of the UAV to help establish reliable communication links between the UAV and the users. Moreover, with additional array gains provided by the RIS, only a small number of onboard antennas (which reduces the energy consumption of the UAV) are needed for the UAV to support the required array gains at the users [97], [98]. On the other hand, by mounting RISs on UAVs to enable over-the-air intelligent reflection, 360° panoramic/full-angle reflection can be realized, i.e., aerial RISs can help reflect signals between any pair of nodes on the ground, which is in sharp contrast to conventional terrestrial RISs that can only serve users in a half space [99]. To support massive connectivity in UAV networks, NOMA can be employed to encourage spectrum sharing among users over the same time/frequence resource block [100]. Nevertheless, RIS-NOMA assisted UAV communications introduces new challenges. First, for the general scenario with multiple UAVs and/or RISs, the achievable rate of each user depends not only on the desired signal power level but also on the interference level. Thus, it is important to jointly design the three dimensional UAV trajectory/placement, RISs reflection beamforming, and user association to maximize the achievable rate of the users. In addition, to fully reap the NOMA gains, additional channel condition/QoS-based decoding order design needs to be considered. This makes the joint UAV trajectory/placement, RISs reflection beamforming, and NOMA decoding order selection problem highly-coupled and difficult to resolve. Compared to the fixed RIS deployment, e.g., on a building, deploying an RIS on a UAV can lead to a more significant improvement of the channel conditions by adjusting the UAV's mobility and/or the RIS reflection coefficients. In [51], a simple design of a UAV aided RIS-NOMA system was proposed to maximize the strong user's data rate based on the UAV's optimized horizontal position. Different from [51], RISs can be also integrated into UAV enabled communication systems to enhance the channel quality. The authors in [52] considered a more practical scenario, i.e., a multi-UAV and multi-user RIS-NOMA system [52]. In this work, the placement and transmit power of UAVs, the reflection matrix of the RIS, and the NOMA decoding order of the users were jointly optimized to maximize the entire system sum rate.

AI/ML can also be used to provide real-time control for UAV enabled RIS-NOMA networks. By smartly controlling the UAV movement and adjusting the phase shifts of the RISs via ML techniques, the energy consumption of UAV enabled RIS-NOMA networks can be minimized [101]. Recently, a novel RIS framework, namely transmitting and reflecting RISs (STAR-RISs), was proposed, where signals are reflected and transmitted simultaneously [102], which is different from the traditional reflection-only RIS framework. Considering the decoding order of the NOMA users, the authors in [102] investigated the sum rate maximization by optimizing the power allocation and the active and passive beamforming in STAR-RIS systems. However, the current research works on STAR-RISs rely on convex optimization methods, which lead to high computation complexity for the time-varying channels. However, ML techniques can be used for the intelligent control of RIS-NOMA systems.

Furthermore, there is increasing enthusiasm regarding HAPs and LEO satellites for providing access services in remote areas, while do not have full terrestrial BS coverage, both from industry to academia. For example, SpaceX Corp implements the Starlink project, which establishes a constellation of thousands of small LEO satellites to support global network services [103]. Several companies are devoted to the research and development of HAPs, including the Zephyr Platform of Airbus [104], Stratobus Platform of Thales [105], and Solar HAPS of HAPSMobile [106]. It is also envisioned that the use of RIS-NOMA can benefit both HAPs and LEO satellite communications, e.g., in creating strong LoS links and enabling massive connectivity between LEO satellites and HAPs, as well as between HAPs and terrestrial devices. Therefore, communication performance can be further enhanced, e.g., in terms of spectrum efficiency, energy consumption, and network coverage. However, studies of RIS-NOMA assisted HAP and LEO satellite communications are still in a nascent stage and advanced design and resource allocation algorithms have to be developed.

B. RIS-NOMA Assisted MEC

MEC allows computation offloading to local servers for reducing the energy consumption and processing delay [107], [108]. The computation offloading performance gains can be further improved if NOMA and RISs are properly applied, as explained in the following.

- From the delay perspective, the use of NOMA allows multiple users to complete their offloading at the same time or one user to offload multiple tasks to multiple MEC servers simultaneously, thus effectively reducing the offloading latency [109], [110]. In addition, from the energy perspective, by asking the strong user to first offload a portion of its task when the weak user is offloading with NOMA, and then allowing the strong user to offload its remaining task using a dedicated time slot, the energy consumption for offloading can be significantly reduced [109], [111]. Moreover, the energy consumption of MEC networks can be substantially decreased by exploiting NOMA uplink transmission [112] and NOMA downlink transmission [113].
- When the communication links for data and computation offloading are strongly attenuated, the use of RISs reflection beamforming can improve the channel quality to offload the computation tasks without incurring additional energy consumption, which facilitates low latency and high energy efficiency [114]–[116]. Furthermore, by utilizing the RISs to dynamically reconfigure the communication links for computation offloading, the offloading decisions and polices can be flexibly determined, thereby improving computation resource usage [81].

Hence, applying RIS-NOMA to MEC can facilitate more intelligent offloading management, by combining the benefits of RISs and NOMA in a constructive manner to further enhance the system performance. In [117], a new flexible time-sharing RIS-NOMA scheme for computation offloading was proposed, which allows users to flexibly divide their

offloading data into two parts transmitted via RIS-NOMA and RIS-OMA, respectively. The results show that RIS-OMA outperforms RIS-NOMA in terms of sum delay minimization when the users' cloud-computing time is sufficiently long and/or the rate improvement of RIS-OMA over RIS-NOMA is sufficiently large. In [118], the joint optimization of the RIS reflection beamforming, power control, transmission rate and time, and NOMA decoding order for RIS-NOMA assisted MEC was presented to minimize the total energy consumption. It was shown that compared to conventional RIS-OMA assisted MEC, the proposed solution can dramatically decrease the total energy consumption of the system. Note that for a large-size MEC network with a massive number of users, edge servers, and RISs, how to efficiently associate the users, the edge servers, the RISs and how to optimally design the power allocation and RISs reflection beamforming are still not known, and merit further investigations.

C. RIS-NOMA Assisted Wireless Power Transfer (WPT)

WPT has been applied in the IoT networks to prolong the battery lives of the devices and resolve the energy limitation problem [119], [120]. On one hand, in recent years, the application of RISs to improve the efficiency of WPT to energy receivers has been studied. As indicated in [121], [122], by properly designing the RIS reflection beamforming and deploying RISs to create LoS paths between the power beacon and the energy receivers, the received power of self-sustainable devices in the downlink can be enhanced, and this in turn increases the throughput performance of the uplink transmission due to the large available transmit power. Moreover, with joint transmit beamforming and RIS reflection beamforming, an improved rate-energy tradeoff can be achieved in simultaneous wireless information and power transfer (SWIPT) networks [123]–[125].

On the other hand, NOMA is also appealing in WPT and SWIPT networks since it enables the simultaneous access of massive numbers of users to improve the throughput, fairness, and energy efficiency [126], [127]. More recently, RISs have been integrated with NOMA and WPT/SWIPT for further performance improvement. In [44], RIS-NOMA assisted WPT is investigated, where multiple devices first harvest energy from a hybrid access point and then transmit information using NOMA to the hybrid access point. An optimization framework for the RIS phase shifts and the resource allocation is developed to maximize the sum throughput of all the devices. Interestingly, it is found that the downlink WPT and uplink information transmission share the same RIS phase shifts, and the proposed design can realize spectrum and energy efficient wireless powered communication networks. In [128], a hybrid NOMA and OMA scheme for RIS assisted WPT networks was proposed to balance complexity and performance. In particular, the users are grouped into multiple clusters, where the users in the same cluster transmit information using NOMA while the users in different clusters transmit information via OMA. Efficient user grouping and resource allocation algorithms were presented, which guarantee a sufficiently large network throughput. In [129], a BS power minimization problem for RIS-NOMA assisted SWIPT networks was studied, and an efficient two-stage algorithm was proposed to jointly optimize the NOMA decoding order, BS transmit beamforming, power splitting ratio, and RIS reflection beamforming. It is shown that compared with non-RIS assisted networks, the proposed design can reduce the BS transmit power by 51.13%. Despite the existing works discussed above, there are still several challenges, such as discrete RIS phase shifts, imperfect CSI and/or SIC, non-linear energy harvesting, which have not been thoroughly investigated for RIS-NOMA networks yet and require further research.

VIII. Conclusions

In this survey, we have provided a comprehensive overview of recent progress on the efficient integration of RISs and NOMA. In particular, the basics of the two techniques were reviewed first, and then the fundamentals of integrating RISs and NOMA were discussed for both SISO and MISO communication scenarios. Because resource allocation is crucial to the success of RIS-NOMA networks, various approaches for resource allocation design, including AI-empowered designs, were described. Security provisioning and practical RIS-NOMA implementation issues were also studied. Finally, the survey was concluded by a detailed discussion of promising future research directions and potential applications of RIS-NOMA.

References

[1] X. You, C. Wang, J. Huang *et al.*, "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 110301, pp. 1–74, Feb. 2021.

[2] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Tech. Mag.*, vol. 14, no. 3, pp. 28–41, Jul. 2019.

[3] Y. Xiao, G. Shi, Y. Li, W. Saad, and H. V. Poor, "Toward self-learning edge intelligence in 6G," *IEEE Commun. Mag.*, vol. 58, no. 12, pp. 34–40, Dec. 2020.

[4] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, Oct. 2020.

[5] M. D. Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. de Rosny, A. Bounceu, G. Lerosey, and M. Fink, "Smart radio environments empowered by AI reconfigurable meta-surfaces: An idea whose time has come," *EURASIP J. on Wirel. Commun. Netw.*, vol. 129, pp. 1–20, May 2019.

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

[7] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, 2020.

[8] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

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

[10] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Nov. 2020.

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

[12] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable intelligent surface-assisted uplink sparse code multiple access," *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2058–2062, Jun. 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.*, pp. 1–1, 2021.

[14] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Springer International Publishing, 2019.

[15] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[16] 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.

[17] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[18] W. Liang, S. X. Ng, J. Shi, L. Li, and D. Wang, "Energy efficient transmission in underlay CR-NOMA networks enabled by reinforcement learning," *China Commun.*, vol. 17, no. 12, pp. 66–79, Dec. 2020.

[19] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2017.

[20] Y. Wang, Z. Tian, and X. Cheng, "Enabling technologies for spectrum and energy efficient NOMA-MmWave-MaMIMO systems," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 53–59, Oct. 2020.

[21] A. S. d. Sena, D. Carrillo, F. Fang, P. H. J. Nardelli, D. B. d. Costa, U. S. Dias, Z. Ding, C. B. Papadias, and W. Saad, "What role do intelligent reflecting surfaces play in multi-antenna non-orthogonal multiple access?" *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 24–31, Oct. 2020.

[22] M. Mohammadkarimi, M. A. Raza, and O. A. Dobre, "Signature-based nonorthogonal massive multiple access for future wireless networks: Uplink massive connectivity for machine-type communications," *IEEE Veh. Tech. Mag.*, vol. 13, no. 4, pp. 40–50, Dec. 2018.

[23] C. You, B. Zheng, and R. Zhang, "Wireless communication via double IRS: Channel estimation and passive beamforming designs," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 431–435, 2021.

[24] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, 2020.

[25] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, "Physics-based modeling and scalable optimization of large intelligent reflecting surfaces," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2673–2691, 2021.

[26] R. Hadani and A. Monk, "OTFS: a new generation of modulation addressing the challenges of 5G," arXIv:1802.02623.

[27] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix OTFS," *IEEE Trans. Veh. Tech.*, vol. 68, no. 1, pp. 957–961, Jan. 2019.

[28] Z. Ding, "Harvesting devices' heterogeneous energy profiles and QoS requirements in IoT: WPT-NOMA vs BAC-NOMA," *IEEE Trans. Commun.*, to appear in 2021.

[29] Q. N. Le, A. Yadav, N. P. Nguyen, O. A. Dobre, and R. Zhao, "Full-duplex non-orthogonal multiple access cooperative overlay spectrum-sharing networks with SWIPT," *IEEE Trans. Green Commun. and Netw.g*, to appear in 2021.

[30] J. Tang, J. Luo, J. Ou, X. Zhang, N. Zhao, D. K. C. So, and K. K. Wong, "Decoupling or learning: Joint power splitting and allocation in MC-NOMA with SWIPT," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5834–5848, Jun. 2020.

[31] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, London, UK, Sept. 2013.

[32] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[33] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753–757, Jan. 2020.

[34] Z. Ding, R. Schober, and H. V. Poor, "On the impact of phase shifting designs on IRS-NOMA," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1596–1600, Apr. 2020.

[35] Z. Ding and H. V. Poor, "A simple design of IRS-NOMA transmission," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1119–1123, Feb. 2020.

[36] 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, Sept. 2021.

[37] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1263–1266, Jun. 2016.

[38] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6174–6189, Dec. 2016.

[39] J. Zhu, Y. Huang, J. Wang, K. Navaie, and Z. Ding, "Power efficient IRS-assisted NOMA," *IEEE Trans. Commun.*, to appear in 2021.

[40] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems: Part I - state of the art and recent findings," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2373–2377, Nov. 2020.

[41] ——, "Unveiling the importance of SIC in NOMA systems: Part II - new results and future directions," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2378–2382, Nov. 2020.

[42] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Commun. Conf.*, 2018, pp. 1–6.

[43] Y. Guo, Z. Qin, Y. Liu, and N. Al-Dhahir, "Intelligent reflecting surface aided multiple access over fading channels," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 2015–2027, 2021.

[44] Q. Wu, X. Zhou, and R. Schober, "IRS-assisted wireless powered NOMA: Do we really need different phase shifts in DL and UL?" *IEEE Wireless Commun. Lett.*, pp. 1–1, 2021.

[45] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7170–7183, 2020.

[46] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Joint deployment and multiple access design for intelligent reflecting surface assisted networks," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[47] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884–6898, 2020.

[48] J. Zuo, Y. Liu, E. Basar, and O. A. Dobre, "Intelligent reflecting surface enhanced millimeter-wave NOMA systems," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2632–2636, 2020.

[49] P. Liu, Y. Li, W. Cheng, X. Gao, and X. Huang, "Intelligent reflecting surface aided NOMA for millimeter-wave massive MIMO with lens antenna array," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4419–4434, 2021.

[50] A. S. De Sena, P. H. J. Nardelli, D. B. Da Costa, F. R. M. Lima, L. Yang, P. Popovski, Z. Ding, and C. B. Papadias, "IRS-assisted massive MIMO-NOMA networks: Exploiting wave polarization," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

[51] S. Jiao, F. Fang, X. Zhou, and H. Zhang, "Joint beamforming and phase shift design in downlink UAV networks with IRS-assisted NOMA," *J. Commun. Inf. Networks*, vol. 5, no. 2, pp. 138–149, Jun. 2020.

[52] X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent reflecting surface enhanced multi-UAV NOMA networks," arXIv: 2101.09145.

[53] M. Fu, Y. Zhou, and Y. Shi, "Intelligent reflecting surface for downlink non-orthogonal multiple access networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshops*, 2019, pp. 1–6.

[54] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA reconfigurable intelligent surface-aided downlink communication networks," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 664–674, 2021.

[55] X. Xie, F. Fang, and Z. Ding, "Joint optimization of beamforming, phase-shifting and power allocation in a multi-cluster IRS-NOMA network," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2021.

[56] F. Fang, Y. Xu, Q.-V. Pham, and Z. Ding, "Energy-efficient design of IRS-NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 14 088–14 092, 2020.

[57] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2874–2885, 2017.

[58] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, 2016.

[59] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44 304–44 321, 2021.

[60] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," arXIv:1905.07726.

[61] H. Li, F. Fang, and Z. Ding, "DRL-assisted resource allocation for NOMA-MEC offloading with hybrid SIC," *Entropy*, vol. 23, no. 5, 2021.

[62] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1057–1071, 2021.

[63] X. Gao, Y. Liu, X. Liu, and L. Song, "Machine learning empowered resource allocation in IRS aided MISO-NOMA networks," arXIv:1905.07726.

[64] Z. Yang, Y. Liu, Y. Chen, and N. Al-Dhahir, "Machine learning for user partitioning and phase shifters design in RIS-aided NOMA networks," arXIv:2101.01212.

[65] M. Shehab, B. S. Ciftler, T. Khattab, M. Abdallah, and D. Trinchero, "Deep reinforcement learning powered IRS-assisted downlink NOMA," arXIv:2104.01414.

[66] L. Lv, H. Jiang, Z. Ding, L. Yang, and J. Chen, "Secrecy-enhancing design for cooperative downlink and uplink NOMA with an untrusted relay," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1698–1715, Mar. 2020.

[67] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.

[68] Z. Chu, W. Hao, P. Xiao, and J. Shi, "Intelligent reflecting surface aided multi-antenna secure transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 108–112, Jan. 2020.

[69] L. Dong and H.-M. Wang, "Enhancing secure MIMO transmission via intelligent reflecting surface," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7543–7556, Jan. 2020.

[70] S. Yan, X. Zhou, D. W. K. Ng, J. Yuan, and N. Al-Dhahir, "Intelligent reflecting surface for wireless communication security and privacy," arXIv: 2103.16696.

[71] Z. Zhang, L. Lv, Q. Wu, H. Deng, and J. Chen, "Robust and secure communications in intelligent reflecting surface assisted NOMA networks," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 739–743, Mar. 2021.

[72] Z. Zhang, J. Chen, Q. Wu, Y. Liu, L. Lv, and S. Sun, "Securing NOMA networks by exploiting intelligent reflecting surface," arXIv: 2104.03460.

[73] L. Lv, Q. Wu, N. Al-Dhahir, and J. Chen, "Secure two-way communications via intelligent reflecting surfaces," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 744–748, Mar. 2021.

[74] L. Lv, H. Jiang, Z. Ding, Q. Ye, N. Al-Dhahir, and J. Chen, "Secure non-orthogonal multiple access: An interference engineering perspective," *IEEE Network*, to be published, DOI: 10.1109/MNET.011.2000539.

[75] X. Guan, Q. Wu, and R. Zhang, "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 778–782, Jun. 2020.

[76] H.-M. Wang, J. Bai, and L. Dong, "Intelligent reflecting surfaces assisted secure transmission without eavesdropper's CSI," *IEEE Signal Process. Lett.*, vol. 27, pp. 1300–1304, 2020.

[77] C. Xu, J. Liu, and Y. Cao, "Intelligent reflecting surface empowered physical layer security: Signal cancellation or jamming?" *IEEE Internet Things J.*, to be published, DOI: 10.1109/JIOT.2021.3079325.

[78] L. Tao, W. Yang, S. Yan, D. Wu, X. Guan, and D. Chen, "Covert communication in downlink NOMA systems with random transmit power," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 2000–2004, Nov. 2020.

[79] X. Lu, E. Hossain, T. Shafique, S. Feng, H. Jiang, and D. Niyato, "Intelligent reflecting surface (IRS)-enabled covert communications in wireless networks," *IEEE Network*, vol. 34, no. 5, pp. 148–155, Sep./Oct. 2020.

[80] L. Lv, Q. Wu, Z. Li, Z. Ding, N. Al-Dhahir, and J. Chen, "Covert communication in intelligent reflecting surface-assisted NOMA systems: Design, analysis, and optimization," arXIv: 2012.03244.

[81] 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.

[82] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Deep learning for large intelligent surfaces in millimeter wave and massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[83] G. C. Alexandropoulos and E. Vlachos, "A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 9175–9179.

[84] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4659–4663.

[85] 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.

[86] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2020.

[87] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[88] H. Liu, X. Yuan, and Y.-J.-A. Zhang, "Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, Nov. 2020.

[89] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, to be published, DOI: 10.1109/TCOMM.2021.3072729.

[90] Y. Liu, X. Mu, X. Liu, M. D. Renzo, Z. Ding, and R. Schober, "Reconfigurable intelligent surface (RIS) aided multi-user networks: Interplay between NOMA and RIS," arXIv: 2011.13336.

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

[92] S. Zhang and R. Zhang, "Intelligent reflecting surface aided multi-user communication: Capacity region and deployment strategy," *IEEE Trans. Commun.*, to be published, DOI: 10.1109/TCOMM.2021.3079128.

[93] C. You, B. Zheng, and R. Zhang, "How to deploy intelligent reflecting surface in wireless network: BS-side, user-side, or both sides?" arXIv: 2012.03403.

[94] N.-N. Dao, Q.-V. Pham, N. H. Tu, T. T. Thanh, V. N. Q. Bao, D. S. Lakew, and S. Cho, "Survey on aerial radio access networks: Toward a comprehensive 6G access infrastructure," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1193–1225, Second Quarter 2021.

[95] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2109–2121, Mar. 2018.

[96] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. IEEE*, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.

[97] S. Li, B. Duo, X. Yuan, Y. Liang, and M. D. Renzo, "Reconfigurable intelligent surface assisted UAV communication: Joint trajectory design and passive beamforming," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 716–720, May 2020.

[98] 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.*, to be published, DOI: 10.1109/TWC.2021.3070014.

[99] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Aerial intelligent reflecting surface: Joint placement and passive beamforming design with 3D beam flattening," *IEEE Trans. Wireless Commun.*, to be published, DOI: 10.1109/TWC.2021.3056154.

[100] Y. Liu, Z. Qin, Y. Cai, Y. Gao, G. Y. Li, and A. Nallanathan, "UAV communications based on non-orthogonal multiple access," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 52–57, Feb. 2019.

[101] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042–2055, 2021.

[102] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint design for simultaneously transmitting and reflecting (STAR) RIS assisted NOMA systems," arXIv:2106.03001.

[103] https://www.starlink.com.

[104] https://www.airbus.com/defence/uav/zephyr.html.

[105] https://www.thalesgroup.com/en/worldwide/space/news/whats/stratobus.

[106] https://www.hapsmobile.com/en/.

[107] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, Aug. 2017.

[108] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.

[109] Z. Ding, P. Fan, and H. V. Poor, "Impact of non-orthognal multiple access on offloading of mobile edge computing," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 375–390, Jan. 2019.

[110] Z. Yang, C. Pan, J. Hou, and M. Shikh-Bahaei, "Efficient resource allocation for mobile-edge computing networks with NOMA: Completion time and energy minimization," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7771–7784, Nov. 2019.

[111] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, "Joint power and time allocation for NOMA-MEC offloading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6207–6211, Jan. 2019.

[112] K. Wang, F. Fang, D. B. d. Costa, and Z. Ding, "Sub-channel scheduling, task assignment, and power allocation for OMA-based and NOMA-based MEC systems," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2692–2708, 2021.

[113] F. Fang, K. Wang, Z. Ding, and V. C. M. Leung, "Energy-efficient resource allocation for NOMA-MEC networks with imperfect CSI," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3436–3449, 2021.

[114] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, Nov. 2020.

[115] Y. Cao and T. Lv, "Intelligent reflecting surface enhanced resilient design for MEC offloading over millimeter wave links," arXIv: 1912.06361.

[116] T. Jiang and Y. Shi, "Over-the-air computation via intelligent reflecting surfaces," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[117] F. Zhou, C. You, and R. Zhang, "Delay-optimal scheduling for IRS-aided mobile edge computing," arXIv: 2011.03690.

[118] Z. Li, M. Chen, Z. Yang, J. Zhao, Y. Wang, J. Shi, and C. Huang, "Energy efficient reconfigurable intelligent surface enabled mobile edge computing networks with NOMA," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 427–440, 2021.

[119] Y. Zeng, B. Clerckx, and R. Zhang, "Communications and signals design for wireless power transmission," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2264–2290, May 2017.

[120] Z. Ding, C. Zhong, D. W. K. Ng, M. Peng, H. A. Suraweera, R. Schober, and H. V. Poor, "Application of smart antenna technologies in simultaneous wireless information and power transfer," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 86–93, Apr. 2015.

[121] B. Lyu, P. Ramezani, D. T. Hoang, S. Gong, Z. Yang, and A. Jamalipour, "Optimized energy and information relaying in selfsustainable IRS-empowered WPCN," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 619–633, Jan. 2021.

[122] Y. Zheng, S. Bi, Y.-J. A. Zhang, X. Lin, and H. Wang, "Joint beamforming and power control for throughput maximization in IRS-assisted MISO WPCNs," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 8399–8410, May 2021.

[123] 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, Aug. 2020.

[124] ——, "Weighted sum power maximization for intelligent reflecting surface aided SWIPT," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 586–590, May 2020.

[125] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface enhanced MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[126] P. D. Diamantoulakis, K. N. Pappi, Z. Ding, and G. K. Karagiannidis, "Wireless-powered communications with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8422–8436, Dec. 2016.

[127] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[128] D. Zhang, Q. Wu, M. Cui, G. Zhang, and D. Niyato, "Throughput maximization for IRS-assisted wireless powered hybrid NOMA and TDMA," *IEEE Wireless Commun. Lett.*, to be published, DOI: 10.1109/LWC.2021.3087495.

[129] Z. Li, W. Chen, Q. Wu, K. Wang, and J. Li, "Joint beamforming design and power splitting optimization in IRS-assisted SWIPT NOMA networks," arXIv: 2011.14778.