Non-Terrestrial Communications Assisted by Reconfigurable Intelligent Surfaces

By JIA YE\textsuperscript{1}, Student Member IEEE, JINGPING QIAO\textsuperscript{2}, Member IEEE, ABLA KAMMOUN\textsuperscript{1}, Member IEEE, and MOHAMED-SLIM ALOUINI\textsuperscript{1}, Fellow IEEE

ABSTRACT | Non-terrestrial communications have emerged as a key enabler for seamless connectivity in the upcoming generation networks. This kind of network can support high data rate communications among aerial platforms (i.e., unmanned aerial vehicles (UAVs), high-altitude platforms (HAPs), and satellites) and cellular networks, achieving anywhere and anytime connections. However, there are many practical implementation limitations, especially overload power consumption, high probability of blockage, and dynamic propagation environment. Fortunately, the recent technology reconfigurable intelligent surface (RIS) is expected to be one of the most cost-efficient solutions to address such issues. RIS with low-cost elements can bypass blockages and create multiple line-of-sight (LoS) links and provide controllable communication channels. In this article, we present a comprehensive literature review on the RIS-assisted non-terrestrial networks (RANTNs). First, the framework of the RANTNs is introduced with detailed discussion about distinct properties of RIS in NTNs and the two deployment types of RIS, that is, terrestrial RISs (TRISs), and aerial RISs (ARISs), and the classification of RANTNs, including RIS-assisted air-to-ground (A2G)/ground-to-air (G2A), ARIS-assisted ground-to-ground (G2G), and RIS-assisted air-to-air (A2A) communications. In combination with next-generation communication technologies, the advanced technologies in RANTNs are discussed. Then, we overview the literature related to RANTNs from the perspectives of performance analysis and optimization, followed by the widely used methodologies. Finally, open challenges and future research direction in the context of the RANTNs are highlighted.

KEYWORDS | High-altitude platform stations (HAPSs); non-terrestrial network (NTN); reconfigurable intelligent surface (RIS); unmanned aerial vehicles (UAVs).

NOMENCLATURE

3-D Three-dimensional.
3GPP Third Generation Partnership Project.
5G/6G Fifth generation/sixth generation.
A2A/A2G/ Air-to-air/air-to-ground/ground-to-air/
G2A/G2G ground-to-ground.
ABS Aerial base station.
AoI Age-of-information.
BER/SER Bit error rate/symbol error rate.
CR Cognitive radio.
CSI Channel state information.
DF Decode-and-forward.
DNN/DRL/ Deep neural network/deep reinforce-
DQN ment learning/deep Q-network.

Digital Object Identifier 10.1109/JPROC.2022.3169690

0018-9219 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
I. INTRODUCTION

As the 5G wireless network deployed worldwide successfully, the state of the art of the 5G beyond and the 6G network architecture is being enthusiastically investigated by researchers from both academia and industry. Future wireless communication networks are expected to provide a much more satisfying service to people by building uninterrupted and ubiquitous connectivity to everyone, everything, and everywhere with ultrahigh data rate, extremely high reliability, and low latency. However, due to the high deployment cost of typical GBSs and wired backhauling infrastructure, it is expensive and unprofitable to build worldwide connectivity via terrestrial networks, especially in remote areas with sparse users. Considering the limitations of the conventional terrestrial cellular networks, the NTNs supported by the developed aerial platforms become a fundamental technology in future communication blueprint [1], [2]. This promising vision has been confirmed by the 3GPP and reported in Technical Report (TR) 38.811 [3]. NTNs are able to tackle several tough problems in recent networks, such as the surge in throughput demands, coverage holes, and terrestrial network failures [4]. Specifically, practical experiments and projects have been initiated to provide ubiquitous Internet services, such as Google Loon [5] and Thales Stratobus.

Compared to the traditional ground wireless communication networks, NTNs tend to provide better channel conditions than terrestrial fading channels because aerial platforms usually have a strong LoS connection with ground nodes [6]. It will be easier to predict the CSI in 3-D positions based on the location information of terrestrial devices and communication performance [7]. Intuitively, the aerial platforms play a dominant role in the NTNs with 3-D mobility, flexibility, and adaptable altitude, who are often regarded as ABSs [8]–[10], relay platforms [11]–[13], or user equipment [14], [15] in NTNs. Different types of platforms possess distinct operating features, such as frequency bands or wavelength, the operating altitude, conditions than terrestrial fading channels because aerial platforms usually have a strong LoS connection with ground nodes [6]. It will be easier to predict the CSI in 3-D positions based on the location information of terrestrial devices and communication performance [7]. Intuitively, the aerial platforms play a dominant role in the NTNs with 3-D mobility, flexibility, and adaptable altitude, who are often regarded as ABSs [8]–[10], relay platforms [11]–[13], or user equipment [14], [15] in NTNs. Different types of platforms possess distinct operating features, such as frequency bands or wavelength, the operating altitude, flight duration, and the size of platforms. They can be divided into three types: UAVs, high-altitude platforms (HAPs), and satellites without loss of generality.

A. Unmanned Aerial Vehicles

They operate at low altitudes of a few hundred meters with a coverage radius of around 2 km [16]. The fully controllable maneuverability 3-D of UAVs enables the height and horizontal position adjustment to fit in different circumstances, which provides flexibility in easy deployment [17]. Because UAVs are battery powered and given the limited onboard energy, their lifetime in communication networks is generally time-limited, ranging from a few minutes to a few hours [18], [19]. UAVs are of great help in establishing a temporary or specific communication network because they can fly directly over the users, thereby introducing higher LoS communication probability and communication stability compared to the conventional terrestrial networks. It is expected that UAVs will bring considerable benefits to many fields, including, but not limited to, emergency communications, smart city construction, rapid network recovery, and dense communication user networks.

B. High-Altitude Platforms

They are network nodes that operate in the stratosphere at an altitude of around 20 km with a coverage radius...
of about 50 km, which enables HAPs to stay at a quasi-stationary position relative to the Earth [20]. Solar power coupled with energy storage has been regarded as the primary means of providing energy for HAPs since they have large surfaces suitable to accommodate solar panel films. Since the required power to stabilize HAPs is greatly reduced due to the low speed of weak wind in the stratosphere, HAPs can support flight duration varying from several hours to a few years. Typically, HAPs used in communications enable fast Internet access and computation offloading, as well as data analysis to millions of devices in urban, suburban, and remote areas. Moreover, they also perform as the large-scale intelligent relays to build fast, reliable, and efficient connections between the satellites [21]. As HAPs are distributed between UAVs and satellites, they can also act as the distributed data centers for recording, monitoring UAVs’ and satellites’ actions.

C. Satellite

The communication satellites are divided into three primary types according to their orbital altitudes. Specifically, the LEO satellites are typically distributed in a circular orbit of about 160–2000 km, and the MEO satellites are in orbit somewhere between 2000 and 35 786 km above the Earth’s surface. Both of them move in relation to the surface of the Earth with a coverage radius exceeding 500 km. The GEO satellites are located in a fixed position in the sky and cover about one-third of the Earth’s surface, and they have the same orbital period as the rotation rate of the Earth in 35 785 km from the Earth’s surface. The power source of satellites is the energy collected from the solar panels, which are exposed to direct solar radiation or to indirect radiation from albedo. Batteries of satellites are installed alongside the solar panels to store energy, which can then be used when the satellite regularly passes through the shadow of the Earth or provide sufficient power during periods of peak demand. Consequently, satellites have long operating lifetimes in the range of 10–20 years on average and provide a complementary role in wireless communication coverage. Satellites can backhaul the traffic load from the edge of the network, broadcast the popular content to the edge, and ensure direct connectivity in remote areas where terrestrial infrastructure is difficult or impossible to implement. Moreover, they can also support 5G service onboard moving platforms, such as aircraft, vessels, and trains.

The composed NTN systems, through its various aerial platforms, can enhance the network reliability by ensuring service continuity, guarantee the service ubiquity in unserved or underserved areas, as well as enable the 5G service scalability with efficient multICASTing or broadcasting. Although aerial platforms bring numerous advantages, they are not yet at their cutting edge of technology, with several vital problems waiting for handles. The specific challenges of NTNs can be summarized as follows.

1) **Weak connectivity:** Down-tilted antennas impair the connectivity at GBSs or GUs. Since the lobes of GBS mainly point to the ground optimizing coverage, the aerial platforms flying over high altitude cannot build a reliable connection with the GBS. Moreover, the complex and uncontrollable wireless environment, especially a crowded area, increases the probability of the LoS links between GUs and aerial platforms being blocked, which can proliferate problems with coverage and connectivity. For example, the human bodies, buildings, and trees severely attenuate the signal strength, which also proposes a new design challenge for the development of aerial platforms in the current environment. In most cases, it is hard to find a position where the ABSs are able to provide LoS connections to all ground devices.

2) **Interference:** However, once the GBS antenna’s main beam is directed toward the intended aerial platforms to concentrate on the transmitted power and provide high signal gain, strong LoS propagation conditions in the aerial area cause only small-signal losses, which impose unintentional but serious interference to aerial users in the same direction. This is a fatal problem, especially in cases requiring a critical data transmission quality. The cooperation work of multiple aerial devices and various aerial platform types also brings cochannel interference and thereby causing A2G transmission performance degradation. Therefore, severe interference in the aerial area is a major hindrance to the coverage extension of the cellular system to the sky.

3) **Severe path loss:** In addition, the long-distance-related large path loss in the 3-D plane is a problem that cannot be ignored in NTNs. The long transmission distance and latency between satellites and terrestrial devices may impose some degradation on the quality of the received signals or the delivered data rate. Also, future communication networks are exploring the use of higher frequencies with wider spectrum bandwidths to tackle the increasing throughput demands and the spectrum scarcity problem. A major challenge hindering their widespread use is their short communication distance and vulnerability to blockages in the propagation environment.

4) **Information leakage:** More eavesdropping threats are brought by the dominating LoS channels built by aerial platforms. Due to the broadcast nature of wireless transmissions and the dominating LoS channels in NTNs, the legitimate receivers are more vulnerable to malicious eavesdroppers, which greatly increases the potential risk of security in NTNs.

5) **Power constraint:** Moreover, the current size and weight of all kinds of aerial platforms are influenced by the carried power supply component. For instance, UAVs are generally time-limited due to the limited onboard energy. The energy-hungry issue seriously affects the promotion and popularization as the flying
process consumes much more power than the hovering process. Due to the size and computation capability of UAVs, it is also impractical to perform power optimization with high complexity. Besides, under 6G communication in a high-frequency band, the system requires sophisticated RF transceiver units and complex signal processing to achieve high-performance communication, which requires the system to have sufficient energy.

6) Hardware limitations: It is well known that mounting multiple antennas at wireless transceivers can boost the communication system performance significantly due to the potential to exploit multiplexing gains offered by spatial DoFs. However, the size, weight, and power constraints of aerial platforms hinder the deployment of advanced MIMO techniques for mitigating the detrimental fading effects.

Therefore, it is necessary to carry out innovative research to facilitate the realization of NTNs. In this context, RIS is envisioned as a cost-effective and energy-efficient reliable alternative to be integrated with aerial platforms. The RIS, also called intelligent reflecting surface, is a planar array comprising a large number of low-cost and nearly passive reflecting elements. RISs can modify the wireless communication environment by introducing some changes in the phase, amplitude, frequency, or polarization of the incident electromagnetic wave. The destructive effect of multipath fading can be counteracted through RIS in a controllable fashion in which they reflect, refract, and scatter radio signals. These features can be leveraged to transform the propagation environment into a smart space that can be programmable for the benefit of the communication application.

The real-time reconfiguration RISs are developed based on the invention of advanced microelectromechanical systems and metamaterials [22], which enables the fixed phase shifters on surfaces to adapt the phase modification in time-varying wireless propagation environments. Since the important research effort is being devoted to developing hardware architectures based on this technology, RIS now can be implemented by various materials, including reflect arrays [23], [24], liquid crystal meta-surfaces [25], ferroelectric films, or even metasurfaces [26]. Compared to existing related technologies, such as active intelligent surface-based massive MIMO [27], multiantenna relay [28], and backscatter communication [29], RIS can reflect ambient RF signals in a passive way, without incurring additional energy cost for signal processing operations or transmitters, and without noise amplification. On the other hand, the lightweight and conformal geometry of RIS can enable their installment onto the facades of almost any object, which provides superior compatibility and high flexibility for practical deployment [30]. Also, integrating RIS into the existing networks is transparent without any change in the hardware and software of the existing devices. Therefore, RIS stands out among existing advanced technologies with advantages, such as overcoming unfavorable propagation conditions, enriching the channel with more multipaths, increasing the coverage area, improving the received signal power, avoiding interference, enhancing security/privacy, and consuming very low energy.

Motivated by the numerous aforementioned RIS’s attendant benefits and the advanced features of aerial platforms, researchers have envisioned that RIS usage in NTNs will offer great agility, flexibility, rapid deployment, and support for the wireless network in a cost-effective manner. RIS enables digitally tuned reflections of incident signals to improve service for isolated users in NTNs. In this way, integrating RIS into NTNs allows several benefits [31]. First, RISs can be rapidly deployed to increase operators’ revenue per user by filling coverage holes and meeting users’ high-speed broadband needs. Second, compared to conventional relay-assisted NTNs, the data transmission between devices will experience less intermediate delays and additional interference to relay the information due to its full-duplex relaying mode. Third, RIS greatly reduces the power consumption for communications since directing signals can be realized in a nearly passive way without an RF source or power amplifier, and only a minimum amount of power is required for the RIS control unit. In addition, due to the thin and lightweight materials of RISs, the aerial platforms’ payload is light even if a large number of reflecting elements are deployed. The greatly reduced energy consumption and payload at aerial platforms yield extended flight duration and reduced aerial platform deployment costs. Finally, RISs consist of low-cost electronics, which does not put a heavy burden on hardware cost.

D. Related Works

There are a few review papers that addressed the integration of RIS with aerial platforms, but their focus is different from this work. In particular, the review papers in [32] and [33] cover the application of RISs in wireless networks at large. Among the applications that were discussed are the use of RISs to enhance the performance between terrestrial networks and UAVs and the wireless relay networks aided by UAV-carried RIS. In this context, the work in [32] describes the role of RIS in reducing UAVs’ energy consumption and discusses the related design challenges, while the work in [33] provides a survey of state of the art related to TRISs and ARISs to assist UAV communications. In the same vein, the tutorial in [34] discusses the advantages and challenges of UAV-carried RIS, whereas the survey in [35] advocates the benefits of using flying RIS to improve the physical layer security in wireless networks. All these works focused on the use of RIS in wireless communications in general and cited as an example the case of UAV communications. The potential application of RIS in networks involving HAPs and satellites appeared only in recent surveys in [36] and [37].
Fig. 1. Survey framework and outline of the main topics.

However, all these works, focusing on integrated RIS-aerial networks for 6G, did not cover important aspects in relation to the potential applications, the challenges, and the future research directions of RIS-assisted nonterrestrial networks (RANTNs). In addition, mathematical tools for designing and analyzing RANTNs have not been extended in any of the published surveys. This lies behind the main motivation of the present work. A systematic organization of this article is shown in Fig. 1, and the main contributions can be summarized as follows.

1) We start by reviewing the two different deployments of RISs in NTNs. TRIS with RISs typically coated on terrestrial objects and ARIS flying over the air. Both of these deployments induce the three kinds of networks existing in the current literature, which are RIS-assisted A2G/G2A, ARIS-assisted G2G, and RIS-assisted A2A communications. This review helps the reader get a general idea of the concept of RANTNs with a clear classification of all existing situations involved in this framework.

2) We provide an extensive review of works studying the combination of existing promising technologies with RANTNs. This aspect has been generally overlooked in existing review papers. In particular, based on the literature, we show that the combination of advanced technologies with RANTNs not only allows for significant gains in performances but also may help solve many limitations of these technologies. We believe that the provided insights and discussion are useful pieces researchers can use to identify practically relevant ideas.

3) Considering the performance analysis and optimization of RANTNs, we classified the related works according to the deployment of RIS, and the performance metric under study presented briefly their associated system models and summarized their main results.

4) Then, we identify the mathematical tools used to carry out performance analysis and optimize RANTNs. It is worth mentioning that optimizing and studying RANTNs is more complex as it involves the joint design of RIS and aerial platforms. Our summary of these tools can be used by researchers to explore existing methods or compare between them.

5) We discuss limitations of the current literature often based on simplistic, yet impractical assumptions. In particular, we highlight the importance of accounting for the mobility of aerial platforms by using an adequate time-varying channel model and the complexity of jointly optimizing the RIS and the aerial platforms’ designs. Both of these aspects are generally overlooked by existing studies, which may call into question their applicability to practical deployments.

Overall, this survey starts from a brief description of the concept of RANTNs to go on to presenting existing works investigating the combination of RANTNs with advanced technologies and posing performance analysis and optimization problems. It also provides a short review of the used mathematical tools before opening up unexplored research directions.

The rest of this article is organized as follows. Section II introduces and compares the two deployment types of RIS studied in current NTN systems. After this, this article divides the RANTNs into three categories depending on the positions of transmitters and receivers. Section III discusses the integration of other promising technologies into RANTNs, which is followed by the performance analysis and optimization of RANTNs in Sections IV and V, respectively. The methodology and tools used to enhance the system performance are concluded in Section VI. Section VII summarizes some practical challenges and future research directions, and Section VIII concludes this article. A list of abbreviations used in this survey is given in the Nomenclature.

II. FRAMEWORKS OF RANTNs

Under the assistance of RISs, the innovative NTN systems composed of various aerial platforms are shown in Fig. 2. The RIS can contribute alongside aerial platforms to introducing rich scattering of LoS links for GUs by optimizing the phase shifts of the reflecting elements. It is
a promising technology regarding performance enhancements, especially when the target receiver is far from the serving transmitter or the direct link between the source and the destinations is blocked. In the absence of channel knowledge, RIS can be used for spatial diversity to combat channel impairments (fading and path loss). For example, by appropriately deploying RISs, it is possible to improve the SNR of users located at the edge of the GBS/ABS coverage area. In this way, the cell coverage of the GBS/ABS can be effectively extended by optimizing the positions and phase shift vectors of RISs.

Before introducing the various NTN systems, we first review the distinct functionality and characteristics of RIS in NTNs.

A. Reconfiguration Intelligent Surface

As a key component in RANTNs, RISs can constructively combine the reflected waves at the destination or compensate for blocked communication links to produce an enhanced transmission and counteract the channel attenuation. The joint beamforming at transmitters and RISs provides the sharp directivity from the transmitter to receivers, which enhances the signal propagation at the desired direction while suppressing the undesired diffusion to other directions. On the other hand, through an adequate control of the reflected direction of the radio waves, RISs can be used to reduce the interference stemming from other users. In addition, RIS is leveraged to enhance the communication quality of the legitimate links and weaken that of the eavesdropping links. Due to a vast amount of low-cost reflection elements, RIS allows for more benefits than conventional relays based on sending and receiving data between devices through forwarding.

1) Less power consumption: RIS consumes much less power than other advanced communication technologies for communications since directing signals can be realized in a nearly passive way without an RF source or power amplifier, and only a minimum amount of power is required for the RIS control unit.

2) Shorter transmission delay: The data transmission through RIS will experience less intermediate delays to relay the information compared to conventional relays, which commonly adopts half-duplex mode. This is because, in a decode and forward half-duplex relaying mode, the transmission is executed over two time slots, while RIS requires only one time slot since the RIS operates in a full-duplex relaying mode.

3) Simpler hardware: The passive nature of the RIS makes its relaying strategy overcome any antenna noise amplification and self-interference, which translates into less computation and lower power consumption than for active full-duplex relays. In addition, RIS provides a promising but inexpensive solution to mimic the massive MIMO gain. As a
result, the NTNs with single-antenna aerial platforms can enjoy beamforming gain under the assistance of TRIS/ARIS without the need to deploy multiple antennas on aerial platforms. In other words, RANTNs with single-antenna aerial platforms are good substitutes for the NTNs with multiple-antenna aerial platforms, as they present lower hardware complexity at the aerial platforms and allow for achieving the same or even better system performance.

4) Longer communication duration: The simplification in communication circuits and minimization of consumed energy indicates that energy-limited devices, being they ground devices or aerial platforms, can reserve more energy for information processing and transmission, which yields extended active duration for commutation at the energy-limited devices.

Despite many advantages mentioned above, the realization of RANTNs faces several new challenges resulting from the controllability of reflecting elements at RIS and the aerial platforms. As shown in Fig. 3(a) and (b), the additional control unit is required to deploy aerial platforms and adjusting the passive beamforming at RIS. Generally, the control can be managed at two levels: ground control stations and aerial platforms.

1) Ground Control Station: The ground controller consists mainly of a processing unit that analyzes the sensed data from the aerial platform, the users’ localization conditions, and the information exchange with GBS and/or gateways. Given a set of global policies (e.g., flying regulations and power) and objectives (e.g., coverage and SNR), the processing unit ensures the joint management of the aerial platform’s flying and communication functions. Moreover, the ground control station also coordinates RISs to monitor the channel estimation and location prediction [38], including the locations of aerial platforms and users, as well as the CSI of the propagation channels. With the information of the required transmission parameter, the ground control station continuously controls the altitude of the aerial platforms and the phase shift of the reflecting elements to serve the target devices and maintain their required QoS [39].

2) Onboard Aerial Platform Control: The onboard aerial platform’s controller mainly consists of the flight control unit. It receives motion commands from the ground control station to ensure the platform’s stabilization and dynamical deployment adjustment. For example, an embedded microcontroller is assumed in [40], allowing the aerial platform to communicate with the source node through a separate reliable wireless control link, thus enabling the instantaneous control of the aerial platform. Another possibility is to embed an active antenna onto the UAV to receive a control signal from the ground control station [41]. In most cases, a specific channel is assigned between the ground control station and aerial platforms for control signaling [42]. If there is an RIS integrated on the aerial platform, an additional RIS control unit will be required at the onboard aerial platform [43]. The RIS control unit receives the optimized RIS configuration, translates it into a switch control map, and applies it to the metasurfaces layer for directing incident signals to targeted directions.

Typically, it is assumed that the required phase shifts are transferred to the memory of the RIS controller. However, such an approach demands a fully synchronized and reliable control link between the computing node and the RIS. This can be achieved in stationary use cases or in other cases where the control link can be easily provided [44]. More interestingly, a sensor can also be integrated into the RIS system, which, for example, collects environmental information such as temperature, humidity, and illuminating light and sends it to a smart controller at the RIS via a wired link. Then, the controller transmits the collected information to the GBS by adjusting the ON/OFF state of the RIS [45]. The reflecting element with the “ON state” can be denoted as “1,” which indicates that the reflecting element is active, and its phase shift can be adjusted to reflect the incident signal to the desired direction. The reflecting element with “OFF state” is denoted as “0,” which indicates that the reflecting element is not active and cannot reflect the incident signals. Under this assumption, the RIS sends its own data to the receiver by controlling its ON/OFF state, and the receiver side can decode the RIS information by using the difference in channel response caused by the ON/OFF state. Due to the
mutualistic relationship between the transmitter and the RIS, the concept of symbiotic communications is borrowed from biology, which not only supports the RIS information transmission but also enhances the primary transmission.

Besides the phase shift control, where to deploy RIS is an important question in the design of NTNs. Since the lightweight and conformal geometry of RIS enables the installment onto the facades of almost all ground and aerial objects, the deployment of RIS can be divided into TRIS and ARIS.

a) Terrestrial RIS: As one of the key technologies to support A2G communications in RANTNs, TRIS paves the way for communication between air platforms and GU s with satisfactory service quality. The realization of TRIS into NTNs is completed by coating RIS to the facades of buildings for outdoor communications or by installing it on the wall for indoor communications. As shown in Fig. 3(a), TRIS is composed of multiple low-cost passive reflecting elements, copper backplane, and control circuit, and the phase shift of each reflection elements can be adjusted by the controller [46]. The ground control station estimates the CSI and angle of arrival between users and the aerial platform to determine the best TRIS configuration setup.

By integrating TRIS in NTNs, concatenated virtual LoS links between aerial platforms and mobile users can be formed via passively reflecting the incident signals, leading to extended coverage and reduced UAV’s movement. With the aid of TRIS, one can adjust the phase shift of the TRIS instead of controlling the movement of aerial platforms for forming concatenated virtual LoS propagation between the aerial platforms and the GU s. Therefore, the aerial platforms can maintain hovering status rather than deliberately altering their route and flying close to their users to establish strong communication links, which is usually time and energy-consuming. The movement of aerial platforms happens only when concatenated virtual LoS links cannot be formed even with the aid of the TRIS, which in turn leads to the reduction of aerial platforms’ energy dissipation and maximizes the endurance of the aerial platforms.

b) Aerial RIS: With the deepening research on RIS, academia has proposed a new aerial architecture, the ARIS, to assist the information transmission, where RIS can be carefully mounted on an aerial platform to create an intermediate reflection layer between end devices. Besides the advantages mentioned above, ARISs enable lighter aerial platforms payload without signal processing circuits, even when a large number of reflectors is deployed. This comes from the fact that RIS is made of thin and lightweight materials. The decrease in the weight of satellites and HAPs enabled by RIS saves a huge amount of cost per operation due to the fact that the aerial platform launch cost is related to its weight. For instance, the cost of launching satellites into an orbit is related to the weight of the satellite system with $6000 per kg. On the other hand, given the reduced communication components and payload, the required stabilization energy is minimized.

Integrating RIS on aerial platforms can be done in several ways depending on the platform’s shape. For instance, the RIS can coat the outer surface of the aerial platform or it can be installed as a separate horizontal surface at the bottom of an aerial device, as shown in Fig. 3(b). This enables intelligent reflection and MIMO transmissions from the sky, making aerial platforms act as aerial relay stations. In addition, a special case of RIS, known as intelligent omnisurface (IOS), has antenna elements on both sides of the metasurface and can reflect incident signals coming from the opposite directions. An unobstructed IOS reflects the incident signals on both sides of the sheet and can cover dead zones, providing massive 260° coverage and higher spectral efficiency. The aerial platforms provide this capability by carrying the IOS underneath them and being able to fly at suitable heights to provide reflective RF surfaces where needed. The greatest strength point of the IOS is its ability to control the direction of the departure signal from the RIS to the potential receivers without any blind spots by adjusting the phase shift vectors on either side.

c) Comparisons: It is obvious that whether the RIS is deployed on terrestrial objects or on aerial platforms, the integration of RIS can enhance the performance of NTN systems. Here, from the perspective of TRIS and ARIS features, we briefly introduce the comparison between them.

1) Deployment: One of the most difficult tasks in the practical implementation of TRIS is finding the appropriate place. The TRIS installation process involves a lot of issues, such as site rent, the impact of the urban landscape, and the willingness of owners to install large RIS on their properties. However, the deployment problem does not bother ARISs flying in the air, which enjoys higher deployment flexibility.

2) Coverage: TRIS deployed on the walls or facades of buildings can at most serve terminals located in half of the space, that is, both the source and destination nodes must lie on the same side of the RIS. Compared to the conventional TRIS, ARIS is able to achieve 360° panoramic full-angle reflecting, i.e., one ARIS can, in principle, manipulate signals between any pair of nodes located on the ground.

3) LoS connection: In a complex environment like urban areas, the radio signal coming from a source node has to be reflected many times before reaching the desired destination node, even with the presence of a sufficient number of TRISs. This leads to significant signal attenuation since each reflection, even by TRIS, would cause signal scattering to undesired directions. Compared to TRIS, the ARIS can be deployed more flexibly with an elevated position to cater to the real-time communication environment. The LoS links between ARIS and the ground ends can be more easily established compared to the ones between TRIS and ground ends, especially in the crowded urban
scenario. Zhang et al. [41] found that while TRIS yields an LoS probability of around 50%, an ARIS results in a probability of over 60%. The trajectory’s planning of aerial platforms can be more flexibly optimized to further improve the communication performance, thereby offering a new DoF for performance enhancement via 3-D network design. ARIS is usually able to achieve desired signal manipulation by one reflection only, even in the complex urban environment, due to its high likelihood of having LoS links with the ground nodes. This greatly reduces the signal power loss due to multiple reflections in the case of TRIS.

4) **Channel variation**: ARIS outperforms TRIS at the cost of experiencing higher fluctuating channels. Specifically, the position variation of the aerial platform leads to the channel variation in both hops in the ARIS-assisted system, while only the link between the aerial platform and TRIS will be impacted in the TRIS-assisted system. Due to the movement of aerial platforms, the signal misalignment happens in both hops of ARIS-assisted system, while it occurs in only one hop of TRIS-assisted system. Therefore, the ARIS passive beamforming is more ineffective than the TRIS passive beamforming. Moreover, the channel estimation/tracking to acquire the ARIS channels along their 3-D trajectories is more challenging than in TRIS deployed at fixed locations.

5) **Lifetime**: The limited endurance of aerial platforms makes ARIS quite challenging to build a reliable network that is available for a long time. The power consumption minimization problem will be more important to ARIS to extend the service time. This problem is overcome by TRIS, which can be continuously powered by backhaul links.

### B. Frameworks of RANTNs

In RANTNs, RISs are deployed to enhance the information transmission between aerial platforms and GUs, ground devices and ground devices, as well as aerial platforms and aerial platforms. In general, the RIS-assisted communications included in RANTNs can be divided into three types according to the types of end devices: RIS-assisted A2G/G2A communications, ARIS-assisted G2G communications, and RIS-assisted A2A communications. In the following, we will introduce the frameworks of these three communication types assisted by RIS.

1) **RIS-Assisted A2G/G2A Communications**: The RIS-assisted A2G/G2A communications have been widely investigated in the existing literature by treating the aerial platforms as one of the end terminals. Specifically, the aerial platforms can act as ABSs to serve GUs with the assistance of TRISs or ARISs or they can also act as aerial users served by GBS. Therefore, based on the functionality of the aerial platforms, the use cases of RIS-assisted A2G/G2A communications can be further divided into two cases.

---

![Fig. 4. A2G/G2A communications with ABSs.](image-url)

#### a) Aerial base station

In this scene, ABSs replace GBSs. In this way, the ground space is saved and the entire system is miniaturized. As shown in Fig. 4, one common scenario involving ABSs is the one where UAVs, HAPs, and satellites act as ABS to build uplink and downlink transmission links with a set of GUs, typically representing IoTDs. These ABSs can solely serve one user or multiple GUs simultaneously by adopting suitable multiple access schemes, such as TDMA [47], [48] and FDMA [49]. To further improve the performance, an ARIS/TRIS can be used to provide an additional link between ABS and the GU when the LoS link suffers from severe propagation deterioration. The use of ARIS/TRIS leads to several major benefits. First, it enables the ABS with limited capabilities and service duration to provide high data rate transmission with significantly improved coverage, capacity, and connectivity. Second, the size and weight of ABS can also be greatly reduced compared to the one adopting conventional MIMO technology, thus extending the battery life of the ground device because of reduced processing energy [50].

In addition to information transmission, under the assistance of TRIS, these ABSs can even provide power supply to ground IoTDs. In this context, ABSs serve in the scene of vast IoTDs by transferring all the necessary information and energy to all of the IoTDs with high efficiency and high security. The IoTDs distributed in every aspect of our life ranging from smart cities, home appliances, and transportation can use the harvested energy to gather and disseminate information from the surrounding environment and upload it to ABSs.

Moreover, few works proposed some interesting frameworks involving multiple TRISs [51], [52], multiple ABS–GU pairs [53], and multiple A2G groups [54], which are worth more attention. However, the consideration of multiple TRISs remains little explored in the context of A2G/G2A communications, while we are not aware of any works considering multiple ARISs. Overall, the A2G/G2A communications composed of ABSs, RISs, and ground terminals are expected to find a place in future networks. In practice, it will be useful for instance when the terrestrial infrastructures are destroyed due to natural disasters.
or had not been installed. It is also applicable in a crowded area for grand events, where the surrounding GBSs are overloaded with heavy communication traffic.

b) Aerial devices: The A2G/G2A communications also include the use cases where aerial platforms act as aerial devices served by GBSs. This framework is similar to the one shown in Fig. 4 but with aerial platforms acting as aerial users served by GBS. In the same way, RIS can be used to combat the path loss and improve the channel gains by creating additional transmission links or mitigating the interference at other unserved aerial users that share the same spectrum in the open air communication environment [55].

A typical scenario of this framework is represented by the case when these aerial devices act as IoTDs such as cameras and communication devices to collect environment information from the air and then send it to GBSs with the help of RIS. Nevertheless, the aerial platforms can also act as relaying stations employed for transmitting the signal to the destination by using forwarding protocols to further enhance the communication quality of the RIS-assisted system [56], [57]. Specifically, the transmitted signals from the transmitter are first forwarded at the aerial platform/reflected at RIS and then reflected at RIS/forwarded at the aerial platform, prior to being received by the receiver. Typically, the TRIS can be employed to reflect the signals transmitted from a ground source to a UAV, and the UAV acts, in turn, as a relaying station employed for transmitting the signal to the destination by using a DF protocol [6], [56]–[58]. In this way, such a network combines the reconfigurable channel abilities of RISs and the flexible deployment of aerial platforms in a novel way.

2) RIS-Assisted G2G Communication: In the G2G NTNs, the aerial platforms assist data transmission between ground devices rather than acting as transmitters or receivers as in A2G/G2A NTNs. As shown in Fig. 5, the aerial platforms are equipped with RIS, which are called ARIS to refer to their being in the air. They serve in this case to assist either the transmission between GBS and served GUs [41], [59]–[62] or the transmission between communicating pairs [63]. The main role of ARIS in this kind of systems is to extend the coverage of the GBSs, enhance communication reliability, and improve spectral efficiency by the additional DoF and the flexible deployment of ARIS. Such a functionality is typically beneficial to communication scenarios in which a sea of IoTDs are required to interact and exchange data with GUs but are often equipped with limited capabilities and cannot communicate over long distances in a reliable manner [39].

Two less common yet interesting scenarios have been considered within the framework of RIS-assisted G2G communications. The first one is wireless traffic backhauling between a remote GBS and a core network through ARIS. Compared to conventional backhauling, such a solution avoids the high cost of fiber-optic infrastructure [64]. The second one is that of cell-free massive MIMO systems [38], which constitutes the core architecture of next-generation mobile communication networks. In these systems, a number of users in a certain area are served by multiple APs simultaneously. To overcome the limitation of intercell interference encountered in traditional cellular mobile networks, these APs apply advanced beamforming technologies. In this context, the use of ARIS can help enhance communication links between APs and users with blocked direct links.

3) RIS-Assisted A2A Communication: Connections are also needed to be built between aerial platforms, as shown in Fig. 6. In this case, ARIS is considered to be the most effective way to assist the transmission since both end terminals are distributed in the air space. In this context, the use of ARIS in LEO satellite communication systems to assist the intersatellite link connection is a promising tool to address the high power consumption and low diversity order problem [65]. The RIS carried by HAPs and LEO satellites provides seamless and ubiquitous connectivity for deep space networks. However, the studies on A2A communication assisted by RIS are still in their infancy. More valuable frameworks regarding the UAV-UAV, UAV-HAP, UAV-satellite, HAP-HAP, and HAP-satellite transmission are important topics that are worth deeper investigation.
Summary: In this section, we classified RANTNs into three categories according to the types of end devices, namely, RIS-assisted A2G/G2A communications, RIS-assisted G2G communications, and RIS-assisted A2A communications. We reviewed the novel frameworks proposed by existing works with great potential to be applied in the 5G beyond/6G network. However, additional research efforts are still in need to build more powerful RANTNs to fully explore the potential of RISs.

III. RIS-ASSISTED NTNs WITH ADVANCED TECHNOLOGIES
Combined with the advanced technologies in the next-generation wireless communication, including but are not limited to optical communication technologies, mmWave/THz communications, wireless power transfer, multiple access schemes, the communication performances of RANTN can be further enhanced. Thus, we will review the current research work on RANTN with advanced technologies in this section.

A. OWC System
Conventional RF-based wireless communication is now seriously challenged by the overcrowding RF spectrum, leading to insufficient capacity to support the ever-increasing wireless data traffic. The idea of OWC has been presented as a promising solution for obtaining larger bandwidth and offsetting the frequency spectrum crowding problem [66], [67].

1) Free-Space Optics: FSO, which utilizes the coherent lightwave of the laser diode to transfer information in free space, is an attractive technology due to its numerous advantages such as high rate capability, license-free, wide spectrum available, long-distance, high EE, reduced interference, directivity, inherent security, and robustness [68]. These advantages make the FSO system a potential supplement candidate for new generation wireless communication. However, using a laser as a propagating medium requires FSO to have LoS between the light source and the photodetector. This issue can be solved by using some relaying methods to satisfy the coverage requirement of FSO communication [69], [70].

On the lookout for constructing a flexible and energy-saving communication system, the combination of the described FSO communication into RANTNs remains an attractive option. In this context, Jia et al. [61] considered an FSO communication assisted by ARIS and showed that the system performance is influenced by the atmospheric conditions, the configurations of laser devices, and the steady ability of UAV. In particular, the EC decreases as the index of refraction structure parameter increases, indicating stronger atmospheric turbulence. On the other hand, EC decreases when the radius of the light source or the UAV vibration increases, as they cause pointing error losses. Moreover, the EC will be reduced when the photoelectrical responsibility decreases because the energy transformation efficiency drop leads to lower electrical SNR. The authors also show that by changing the location of ARIS, the link distance variation will cause the area of beam footprint on photodetector to change, resulting in markedly different ECs.

2) Visible Light Communication: As far as optical communication is concerned, VLC represents a potential candidate that can be deployed in RANTNs. Based on utilizing the incoherent lightwave of light-emitting diode to transfer information in the wireless channel, this technology has been presented as a popular solution for short-distance broadband wireless communication [71], [72]. It presents the advantages of incurring low deployment costs while providing ultrahigh data rates and operating in the unlicensed spectrum without health hazards [73], [74]. Besides, compared with traditional RF, VLC can provide communication and illumination simultaneously. However, there are several challenges toward VLC implementation, including the limited coverage range, the signal loss caused by little movement, the misalignment between the light-emitting diodes transmitter and photodetector receiver, and the need for LoS [72]. The RANTNs can help overcome these drawbacks and improve the performance of VLC. On the technological level, it is feasible to apply a metalens or crystal liquid-based RIS into the VLC communication system to shape the environment of the incident light signals through dynamic artificial muscles and the refractive index [75].

A study about the TRIS-assisted VLC communication system was conducted in [76], where multiple ABSs are assumed to provide communication and illumination for GUs simultaneously. Each ABS is only under the assistance of one RIS at most but can serve multiple users simultaneously. The impact of the ABSs’ deployment height was fully investigated in this work. Specifically, when the height of the ABSs increases, the total transmitting power decreases first and then increases later as the height of the ABSs increases. This is in large part due to the fact that the cosine of emission angles and the communication distance, both increasing when the ABSs fly from low to high, possess an opposite effect on the performance. While the increase in the cosine of emission angles enhances the system performance, the increase in the communication distance leads to a higher path loss and thus degrades the performance. Moreover, the number of TRISs also plays a nonnegligible positive role in the total transmit power of the considered system. The authors derived the conclusion that even when TRISs are associated randomly at each time and used zero-phase shifts, the total transmitting power can be reduced by 21.73% on average.

B. Millimeter Wave/THz
One of the unique features of future 6G wireless networks is the use of frequency bands above 100 GHz, which brings mmWave and THz communication to our life [77]. The large available bandwidth enables them
to be the essential components to support the ultrahigh data transmission for applications including, but not limited to virtual reality, high-definition video broadcasting. However, high-frequency bands experience tough environmental conditions due to molecular absorption, resulting in severe attenuation and high path losses \([78], [79]\). Furthermore, the small wavelength of mmWave and THz spectrum yields the high susceptibility to blockage caused by common objects, such as buildings and foliage, which seriously attenuates the reliability and availability of wireless communication services \([80]\). In order to bypass obstacles and prolong the communication range, the proposed RANTNs have been considered as an energy-efficient solution for mmWave and THz communications.

1) TRIS: Sun et al. \([81]\) and Guo et al. \([82]\) adopted TRIS to assist the secure communication between an ABS and a legitimate GU operating in mmWave frequency, while Wang et al. \([52]\), Pan et al. \([83]\), and Jiang and Jafarkhani \([84]\) adopted TRIS to assist the downlink transmission between an ABS and multiple GUs operating over mmWave or THz frequencies. As indicated in \([83]\), the THz frequencies experience severe path losses at some distance-dependent locations. As such, path loss peaks are going to appear when the communication distance varies according to the movement of the aerial platform. Consequently, terrestrial THz transmission approaches cannot be directly applied to A2G communications. Considering the scenario of a UAV serving multiple users through a TRIS, Pan et al. \([83]\) proposed a solution based on dividing the THz band into subbands, each allocated to a specific UAV-user link. The subband allocation is updated for each UAV's position to avoid path-loss peaks.

2) ARIS: Due to the blockage-prone nature of mmWave/THz signals, mobile ARISs are a better option to further maintain LoS links and enhance communication than stationary RISs. ARISs can improve the reliability of mmWave/THz transmissions by optimizing their location intelligently.

a) G2G communications: In this context, the works in \([41], [42]\), and \([85]–[87]\) investigated a novel downlink framework using ARIS to assist high-frequency communications between GBS and GUs. By leveraging the mobility of ARIS and adjusting its position, the direct NLoS links between GBS and users can be replaced by two connected LOS links. Compared to a system assisted by the TRIS, the works in \([41]\) and \([42]\) noted that a significant improvement in LoS connection probability can be achieved. Specifically, Zhang et al. \([41]\) observed that the LoS downlink probability naturally increases with the altitude of the ARIS and can be further improved by adopting a suitable trajectory design. However, as shown by the work in \([42]\), improving the probability of LoS links by increasing the altitude of the ARIS does not result in better performances due to the higher experienced path loss in the high-frequency band. If the altitude of the ARIS increases beyond a certain threshold, the gain brought by ARIS is lost, the performance becoming equivalent to that of TRIS. However, with ARIS, it is possible to improve the downlink rate faster than with TRIS by increasing the transmit power. Furthermore, to ensure a better LoS link, the ARIS is found to change its position more frequently in the case of a moving GU than in the case of a static GU.

b) G2A communications: Besides G2G communications, ARIS operating over high-frequency bands can be used for G2A communications. In this context, Abuzainab et al. \([88]\) considered the setting of a multi-antenna GBS serving a mobile aerial user in the THz band. Due to the movement of the aerial user, the LoS link between the GBS and the aerial user may not always be maintained. In such a situation, the GBS can serve the aerial user through ARIS. A design based on a proactive hand-off and beam selection that predicts the best communication link (direct or RIS assisted), as well as the best beamforming vector, was proposed in \([88]\).

c) A2A communications: The use of THz band provides high data rates for intersatellite links, as the path loss due to molecular absorption is a nonissue for space-based applications of THz waves. However, a significant decrease in the received power can occur when the transmitter and receiver are misaligned. The misalignment fading, originating from the sharp beams of THz communication and the LEO satellites moving with high velocity, has been considered in \([65]\) for an RIS-empowered THz band in the inter-LEO satellite communications. The derived results in \([65]\) showed that the impact of the misalignment loss could be reduced by using antennas with large bandwidths. To compensate for the high path loss experienced in high carrier frequencies, increasing the number of ARIS elements on each satellite is a cost-effective solution to reach the desired QoS.

C. Wireless Power Transfer

Wireless power transfer has emerged as an effective solution to recharge low-power devices with a limited lifetime \([89]–[91]\). Collecting energy from electromagnetic radiation carried by radio signals allows for charging devices wirelessly, thereby ensuring full operation while avoiding the cost of wired charging devices. This technology turns out to be useful for devices with limited power capacity, such as IoT devices in the IoT networks, allowing them to be charged without wired charging circuits, thus paving the way for miniaturization. One major challenge in wireless power transfer is to ensure that capacity-limited devices harvest enough energy to support necessary operation without increasing the transmit power. The new paradigm, RIS, thus can deal with the challenge for the reason that RISs can create additional communication links to establish enhanced wireless charging/harvesting zones for wirelessly powered devices. This has motivated researchers to apply the RIS technology into wireless power transfer networks \([92], [93]\) and further combine it with SWIPT \([94], [95]\) and wirelessly powered-communication.
networks [96]. In addition to RIS, aerial platforms may support the communication scenarios with energy-limited devices since aerial platforms can supply such devices with power through downlink power transfer. The joint use of RIS and aerial platforms offers further opportunities to provide more energy charging and information transfer capacity and introduce new challenging tradeoff problems to ensure fairness between the information transmission performances and EH performance.

1) Wirelessly Powered GUs: The work in [97] considered a TRIS-assisted cell-free network, where the ground IoTDs are powered by the downlink signals from tethered ABSs directly and then send uplink signals to ABSs through the TRISs. In the considered scenario, the deployment and the number of ABSs play more important roles than the ones of TRIS. It has been noted in [48] that using the UAV to charge IoTDs directly is not a good option for two major reasons. On the one hand, outfitting the UAV with an omnidirectional antenna to spread energy waves results in low received power, which, in turn, causes a long charging time. On the other hand, using a directional antenna at the UAV is not a feasible option since, in this case, the UAV has to know the exact positions of all IoTDs. Such a condition cannot be satisfied in vast IoTDs settings. As a solution, the work in [48] considered the use of PDs relaying data and energy transmission between a UAV and vast ground IoTDs under the assistance of TRISs. Based on a set of simulations, the work in [48] illustrates the significant performance improvement in energy and time consumption brought by the deployment of TRISs and the optimization of the UAV trajectory. To better enhance the power transfer and information transmission from the GBS to the IoT, both ARISs and a TRIS are deployed in [98]. The IoT applies the power splitting scheme to perform communication and energy supply simultaneously, where a portion of the received signals is utilized for signal decoding, while the remaining part is used for EH. The required minimum harvested power at the IoT is also considered and plays a negative role in the achievable rate. It is because the energy supply service gradually becomes the primary work for the proposed system, and the communication service is the secondary work as the required harvesting power increases. A power splitting scheme has also been considered in the recent work [99] in which a battery-limited ABS is designed to communicate with ground information receivers and power EH users through an RIS.

2) Wirelessly Powered RIS: In addition to wirelessly powering the GUs, UAVs can power TRIS. In this context, the work in [100] considered a design in which an UAV with fixed transmit power charges both GUs and TRIS. The design uses the harvest-then-transmit protocol according to which the GUs and TRIS harvest energy and use it later for uplink transmissions. This work shows that the optimized trajectory of the UAV is such that it flies close and hovers around to the TRIS and GUs to enhance the energy and information transfer. Wirelessly powering the ARIS has also been investigated in recent works. Although ARIS consumes a small amount of energy, it is advisable to avoid drawing its needs from aerial platforms because of their limited onboard energy. Instead, the RIS can self-power by harvesting energy from the unreflected fraction of the incident signals and converting it into electrical energy via the rectifier [42]. Structurally, the self-powered ARIS is composed of an adjustable antenna and a rectifier. By varying the impedance of the antennas, the mismatch between the antenna structures with the carrier wave can reflect back a portion of the incident signal, while the rectifier can harvest the remaining part to power the reflector. Under this assumption, the amplitude of the reflection coefficient of each element on RIS is not unit anymore but is less than 1. The simulation results given in [42] showed that the energy harvested during transmission is sufficient for ARIS to self-power without drawing any energy from the UAV when the GBS transmit power is high enough.

3) Wirelessly Powered Aerial Platforms: Aerial platforms with limited batteries can also be charged by transmitted signals during their serving duration. In this context, the work in [101] considered the scenario in which an ABS with a high-capacity battery and a UAV carrying RIS with a low-capacity battery leave the charging station to provide communication services for GUs. The UAV-carried RIS needs to be wirelessly charged by the ABS. A complex design involving cooperation between ARIS and ABS is proposed. Its aim is to help the UAV-carried RIS decide whether to be charged or terminate when its battery is low, and the ABS has just enough energy for charging. This work shows that for the optimized design, both the ARIS and ABS almost run out of energy simultaneously and try to terminate near the charging station. Interestingly, results also show that the system performance is not always likely to improve with the increase of the battery capacity of the ABS, which can be explained by the involved complex power charging decision problem. In NTN networks, smart and efficient management of the energy is thus crucial. HAPs or satellites can also be powered by solar energy in addition to energy harvested from RF signals [43]. NTN networks are expected not only to improve terrestrial communication but also to enable deep space communication. They can act as a relay to exchange information between low-power sensor networks used in deep space networks, and they can power them by making use of RIS to transmit power from solar power satellites to the planetary surface [102].

D. Nonorthogonal Multiple Access

Today's wireless networks allocate radio resources to users based on the OMA schemes, such as TDMA [47], [48] and FDMA [103], [104]. However, as the number of users increases, OMA-based approaches may not meet the stringent emerging requirements of very high spectral efficiency, very low latency, and massive device
connectivity [105]. Sparked by the concept of superimposing the signals of multiple associated users at different power levels, NOMA has been invoked for improving the spectrum efficiency, guaranteeing user fairness, and supporting massive connectivity of wireless networks [106]. The key idea is to use the power domain for multiple access to exploit the spectrum more efficiently by opportunistically exploring the users’ different channel conditions. In this way, it allows all users to use the total communication resources, including frequency, space, and spreading code simultaneously. Central to NOMA, it is the use of successive interference cancellation (SIC) at the strongest code simultaneously. 

In this way, it allows all users to use the total communication resources, including frequency, space, and spreading code simultaneously. Central to NOMA, it is the use of successive interference cancellation (SIC) at the strongest code simultaneously. 

The key idea is to use the power domain for multiple access to exploit the spectrum more efficiently by opportunistically exploring the users’ different channel conditions. In this way, all users to use the total communication resources, including frequency, space, and spreading code simultaneously. Central to NOMA, it is the use of successive interference cancellation (SIC) at the strongest code simultaneously. 

In this way, it allows all users to use the total communication resources, including frequency, space, and spreading code simultaneously. Central to NOMA, it is the use of successive interference cancellation (SIC) at the strongest code simultaneously. 

1) TRIS: In [108], an aerial user and a GU share the same NOMA protocol with the help of TRIS. To handle the interference caused to the GU by spectrum sharing, the GBS performs the GBS-GU interference cancellation: it first detects the GBS-GU signal by treating the GU's signal as noise and then detects the GU's signal in an interference-free manner by subtracting the demodulated UAV's signal from the received composite signal. Liu et al. [109] and Diamanti et al. [110], [111] integrated NOMA techniques into the dynamic A2G/G2A communication scenario with multiple roaming GUs, where a TRIS is employed on the facade of a high-rise building to form concatenated virtual LoS propagation between the UAV and the users. Contrary to conventional NOMA, the decoding order of the considered TRIS-assisted NOMA system could not be obtained from the order of the GU's channel gains. Instead, other ordering techniques are employed, including the sorted received SINR [109], the cascaded channel gains [111], and the received signal strengths [110]. Furthermore, the multigroup A2G network employing NOMA to serve multiple users in each group under the assistance of a single TRIS was studied in [54]. The SIC decoding order among users in each group is determined by the distance between users and the serving UAV when the experienced path loss dominates the effective channel gains of users. This assumption facilitates the design as it avoids the need for instantaneous CSI. Compared to OMA-based networks and the interference-free network with orthogonal operating frequency bands and time slots, the results given in Fig. 7 showed that the RIS gain is more pronounced in NOMA-based networks than in other transmission schemes. The reason why OMA achieves the worst performance is that it provides limited resource blocks for each user compared with NOMA and experiences intergroup interference compared with the interference-free scheme. The works in [54] and [109] concluded that NOMA is more efficient in terms of energy consumption and spectrum efficiency than the OMA scheme. As demonstrated in [54], the reasons are twofold. First, NOMA supports serving many users simultaneously, leading to better spectral efficiency. Second, UAV and TRIS introduce new DoFs brought by optimal UAV placement and RIS configuration that allows for efficient implementation of NOMA.

2) Aerial RIS: The work in [112] considered the uplink of a multicell system in which multiple GUs communicate with GBSs through an ARIS. A successive cancellation technique is employed at the GBSs with a decoding order based on the cascaded channel gains. Focusing on the downlink, Jiao et al. [113] proposed an NOMA-based design to assist communication between a GBS and two different users through an ARIS. The authors showed that the ARIS-assisted NOMA system outperforms the traditional NOMA system even with random phase shifts. In particular, a comparison with ARIS-assisted OFDMA has shown that ARIS OFDMA can only improve the performance of the strongest user, while the ARIS-assisted NOMA is capable of improving the performance of both users. Targeting the minimization of the total consumed power, Khalili et al. [87] considered a heterogeneous network assisted by multiple ARISs to support macro BS downlink transmissions operating over microwave frequencies based on OMA overlaid by downlink transmissions of small GBSs operating over millimeter-wave frequencies based on NOMA. An important finding of the works in [87] and [109] is that the UAV requires fewer movement actions under the NOMA scheme because the transmit rate of GUs in NOMA networks is higher than that in OMA networks, which indicates that the data demand constraint is more likely to be satisfied in NOMA networks.

E. Cognitive Radio

In addition to the NOMA scheme, another efficient method to improve spectral efficiency is the CR. In CR systems, the spectrum is shared by two different types of users, primary users, and secondary users. Specifically, the secondary users can opportunistically access the spectrum.
bands owned by the primary licensed users on the condition that they control the interference leakage to primary licensed users [114]. This technology allows for efficient use of the spectrum to cope with the ever-increasing number of wireless devices. This has motivated several researchers to investigate its combination with RANTNs. In this line, the work in [115] considered a primary network wherein a satellite communicates with a satellite GU coexisting with a secondary terrestrial network in which a GBS communicates with a GU. A TRIS is used in the secondary network to manage the secondary network’s interference and prevent information leakage. More recently, a more complicated ARIS-assisted CR network was studied in [116]. In particular, the primary transmitter communicates with multiple primary GUs but is interfered by the secondary network, where the secondary transmitter transmits the information to multiple secondary GUs with the help of ARIS. An adaptive transmit power policy is proposed in this work. It aims at adjusting the transmit power of the secondary transmitter to keep the OP of the worst primary user’s channel below a given threshold.

Summary: This section opens up the vast possibility of using various promising technologies in RANTNs. The combination not only allows for boosting the performance of the considered RANTNs but also may help overcome several of the problems posed by these technologies. On the one hand, technologies, such as OWC enabled by FSO and VLC, mmWave, and THz, can be used to overcome spectrum scarcity in RANTNs. In the same vein, the wireless power transfer technology contributes toward extending the operation duration of aerial platforms, RISs, and even IoTDs, whereas the multiple access schemes and spectral sharing strategies offer opportunities to support massive users with limited channel resources. On the other hand, several challenges posed by the implementation of these advanced technologies may be mitigated when used in RANTNs. For example, RISs allow for better channel gains while relaying aerial platforms improves link reliability and thus are key to mitigating severe path loss experienced in high-frequency bands. Similarly, with the help of RIS and by a dynamic adjustment of aerial transmitters, EH devices can collect more energy without degrading the communication performance or requiring more transmitted power. Aerial platforms together with RIS can also mitigate interference caused by multiple users accessing the same spectrum. To sum up, the integration of advanced technologies into RANTNs is a topic of major interest, opening up possibilities for achieving higher performances and solving some of their limitations. Existing works provide several useful insights that can be of valuable importance for practical implementation. The summary of existing literature can be seen in Table 1.

IV. PERFORMANCE ANALYSIS
The various RANTN frameworks and the combination with other promising technologies lay a solid foundation for system design. However, how to evaluate the system performance becomes one of the leading research directions. Due to the randomness of the propagation environment between end devices, the quality of the received signal is also random, following particular distributions related to the properties of the wireless channel medium. Using tools such as probability theory, random matrix theory, or stochastic geometry, several essential metrics, such as OP, EC, and EE, can be theoretically analyzed in closed-form expression depending on the system’s parameters. The advantages of such theoretical studies are twofold: they avoid the need for extensive numerical simulations and assist in finding optimal allocation resources. Recently, a significant research effort from both industry and academia has been made to assess the benefits of the RANTNs, by carrying out several performance analysis studies.

A. Terrestrial RIS

1) UAV With Highly Directional Antenna: The recent work in [117] considered a UAV equipped with a small beamwidth highly directional antenna, that communicates with a GU through a TRIS. It is assumed that the UAV is directed toward the TRIS, and thus, due to the high directivity of the antenna, no direct link exists between the UAV and the GU. Moreover, only a fraction of the reflection elements of the TRIS lying inside the radiation footprint are illuminated. The authors delineated the footprint zone to characterize several metrics such as the average received SNR, OP, and the average outage duration, which is defined as the average time that the fading envelope remains below a specified level after crossing that level in a downward direction. The analyzed results show that the average received SNR increases as the distance between the UAV and the TRIS increases as a result of more reflecting elements being illuminated. Then, when the footprint becomes large covering all the reflecting elements, it decreases as the spillover losses continue to increase.

2) Relaying UAV: The work in [118] considered a UAV acting as a relay to forward signals transmitted by GU through a TRIS to a satellite in the presence of interfering users and hardware impairments. Specifically, the transmitted signals from the GU are first sent to a UAV under the assistance of the TRIS, and then, the UAV decodes and forwards them to the satellite. By analyzing the impact of the number of reflecting elements on the OP, the authors concluded that TRIS with a sufficient number of elements together with UAV is key to mitigating interference and severe path losses as well as combating hardware impairments. A similar relaying system was considered by Yang et al. [57], where a TRIS and UAV are cooperatively used to relay the communication between a ground pair. Under the assumption of a Rician distribution for TRIS-UAV and UAV-GU channels, and Rayleigh distribution for the TRIS-GU channel, the authors provided an approximation of the cascaded channel distribution.
Ye et al.: Non-Terrestrial Communications Assisted by Reconfigurable Intelligent Surfaces

Table 1 Summary of Existing Literature About the Integration of Other Promising Technologies Into RANTNs

<table>
<thead>
<tr>
<th>Reference</th>
<th>Technologies</th>
<th>RIS types</th>
<th>System model</th>
</tr>
</thead>
<tbody>
<tr>
<td>[41]</td>
<td>mmWave</td>
<td>ARIS</td>
<td>G2G MU MISO downlink transmission</td>
</tr>
<tr>
<td>[42]</td>
<td>mmWave, WPT</td>
<td>ARIS</td>
<td>G2G PDP MISO downlink transmission</td>
</tr>
<tr>
<td>[48]</td>
<td>WPT</td>
<td>TRIS</td>
<td>A2G MU SISO downlink transmission assisted by two RISs</td>
</tr>
<tr>
<td>[54]</td>
<td>NOMA</td>
<td>TRIS</td>
<td>A2G multi-group SISO system</td>
</tr>
<tr>
<td>[52]</td>
<td>mmWave</td>
<td>TRIS</td>
<td>A2G MU SISO downlink transmission assisted by multiple TRISs</td>
</tr>
<tr>
<td>[76]</td>
<td>OWC (VLC)</td>
<td>TRIS</td>
<td>A2G MU SISO downlink transmission served by multiple ABSs</td>
</tr>
<tr>
<td>[64]</td>
<td>OWC (FSO)</td>
<td>ARIS</td>
<td>G2G PDP SISO transmission</td>
</tr>
<tr>
<td>[65]</td>
<td>THz</td>
<td>ARIS</td>
<td>A2A PDP inter-satellite transmission</td>
</tr>
<tr>
<td>[81]</td>
<td>mmWave</td>
<td>TRIS</td>
<td>A2G MU SISO downlink transmission</td>
</tr>
<tr>
<td>[82]</td>
<td>mmWave</td>
<td>TRIS</td>
<td>A2G MU MISO downlink transmission</td>
</tr>
<tr>
<td>[83]</td>
<td>THz</td>
<td>TRIS</td>
<td>A2G MU SISO downlink transmission</td>
</tr>
<tr>
<td>[84]</td>
<td>mmWave</td>
<td>TRIS</td>
<td>A2G MU MISO downlink transmission assisted by multiple TRISs</td>
</tr>
<tr>
<td>[85]</td>
<td>THz</td>
<td>ARIS</td>
<td>G2G MU SISO downlink transmission</td>
</tr>
<tr>
<td>[86]</td>
<td>mmWave</td>
<td>ARIS</td>
<td>G2G MU MISO multicast communication</td>
</tr>
<tr>
<td>[87]</td>
<td>NOMA, mmWave</td>
<td>ARIS</td>
<td>G2G MISO heterogeneous network assisted by multiple ARISs</td>
</tr>
<tr>
<td>[88]</td>
<td>THz</td>
<td>ARIS</td>
<td>G2A PDP SISO downlink transmission</td>
</tr>
<tr>
<td>[99]</td>
<td>WPT</td>
<td>TRIS</td>
<td>A2G MU MISO downlink transmission for both information receivers and energy harvesting receivers</td>
</tr>
<tr>
<td>[97]</td>
<td>WPT</td>
<td>TRIS</td>
<td>A2G MU MISO energy transfer network served by multiple ABSs</td>
</tr>
<tr>
<td>[98]</td>
<td>WPT</td>
<td>TRIS and ARIS</td>
<td>G2G MISO downlink simultaneous wireless information and power transfer system</td>
</tr>
<tr>
<td>[100]</td>
<td>WPT</td>
<td>TRIS</td>
<td>A2G MU SISO wireless power transfer and G2A MU SISO information transmission</td>
</tr>
<tr>
<td>[101]</td>
<td>WPT</td>
<td>ARIS</td>
<td>A2G MU SISO downlink transmission</td>
</tr>
<tr>
<td>[108]</td>
<td>NOMA</td>
<td>TRIS</td>
<td>A2G SISO uplink transmission with with the coexistence of multiple ground users</td>
</tr>
<tr>
<td>[109]</td>
<td>NOMA</td>
<td>TRIS</td>
<td>A2G MU MISO downlink transmission</td>
</tr>
<tr>
<td>[110]</td>
<td>NOMA</td>
<td>TRIS</td>
<td>G2A MU SISO downlink transmission</td>
</tr>
<tr>
<td>[111]</td>
<td>NOMA</td>
<td>TRIS</td>
<td>G2A MU SISO uplink transmission</td>
</tr>
<tr>
<td>[112]</td>
<td>NOMA</td>
<td>ARIS</td>
<td>G2G MISO uplink transmission for multi-cell communication system</td>
</tr>
<tr>
<td>[113]</td>
<td>NOMA</td>
<td>ARIS</td>
<td>G2G MU SISO downlink transmission</td>
</tr>
<tr>
<td>[115]</td>
<td>CR</td>
<td>TRIS</td>
<td>A2G CR system with primary satellite communication and secondary terrestrial network</td>
</tr>
<tr>
<td>[116]</td>
<td>CR</td>
<td>ARIS</td>
<td>G2G MU SISO downlink transmission</td>
</tr>
</tbody>
</table>

and derived the SNR of the RIS-assisted G2A system. Considering the DF relaying protocol at UAV, the exact OP and BER, as well as the asymptotic approximation in high SNR regime, are obtained in closed form, while an upper bound of the capacity is derived by applying Jensen’s inequality. It has been shown that the OP performance is related to the experienced path loss and the probability of LoS transmission, both of which depend on the height of the UAV. More specifically, for small ranges of the UAV height, increasing the height increases the probability of LoS link and thus improves the OP. However, by continuing to increase the height further, a larger path loss is experienced, and thus, the OP increases. This explains why the OP decreases first with the height before increasing after a certain threshold. It has also been noted that the end-to-end OP in the considered system does not continue to improve with the number of reflecting elements when it is very large. The reason is given as follows. Since a DF protocol is considered, the OP depends on the minimum SNR of the two hops. It becomes thus dominated by the second hop for which no TRIS is employed.

3) Existence of Interferers: The work in [119] investigated the use of TRIS in the UAV enabled vehicular communication system with infinite and finite blocklength codes, where UAV communicates through a TRIS with a single-antenna ground vehicle in the presence of several interfering vehicles on the road. Assuming that all propagation channels experience Nakagami-m fading and considering nonidentical direct interference links between GUs and ground interfering vehicles, the authors derived several performance metrics, including the coverage probability, the BER, the block error rate, and the goodput defined as the number of successfully delivered information bits from the UAV to the vehicle receiver through RIS in a given time. The simulation results reveal that the system performance degrades when the number of interfering vehicles increases or when the CSI is inaccurate. Considering finite blocklength codes composed of both training and information bits, the authors showed that the goodput increases first with respect to the number of training bits before decreasing. This shows that there is a tradeoff between the number of information bits and the number of training bits in a finite blocklength frame.

B. Aerial RIS

Several theoretical studies were performed to explore the potential of ARIS. Targeting the G2G system, Jia et al. [61] considered the communication between a light source transmitter and a photodetector receiver through a UAV. The cascaded FSO channel is then the product of the atmospheric turbulence and the geometric pointing loss. The authors opted for a Gamma–Gamma distribution to model the atmospheric turbulence, which is known to present a good fit for a wide range of turbulence strengths. As for the geometric pointing error loss, it is a function of the movement of the aerial platform UAV and the misalignment vector between the centers of the photodetector and the beam footprint. Its distribution is then derived by assuming that the misalignment distance follows a Hoyt distribution. Under this model, the asymptotic
EC at high SNR was obtained, providing insights into the relationship between the capacity and the atmospheric conditions, the configurations of laser devices, as well as the steady ability of UAV. Targeting the same ARIS-assisted G2G communication network, Mahmoud et al. [59] provided a closed-form expression for the probability density function of a tight upper bound on the instantaneous SNR under the assumption of Rayleigh fading. The calculations hinge on the Cauchy–Schwarz–Buniakowsky inequality, which allows for obtaining the MGF of the SNR upper bound and its associated SER, EC, OP, and outage capacity. For the sake of comparison, the authors asymptotically analyzed these performance indexes based on the central limit theorem and showed that while the CLT-based approach is in agreement with the simulation in the low SNR regime, it presents a gap for high SNR values. Considering a communication link between a GU to a BS through a UAV carried RIS over Rician fading conditions between all communicating devices, Iacovelli et al. [120] proposed to approximate the cascaded channel by a complex Gaussian channel. Such an approximation was based on the presence of a weak LoS link between GU and a BS and accounts for frequency- and spatial-selective fading. They showed that it leads to computing a data rate lower bound and suggested to use it in practice to assess the performances and guide the system’s design. Focusing on RANTNs to assist G2G communication, Alfattani et al. [62] developed a complete link budget analysis by dividing the transmission into two different regimes: the specular reflection paradigm and the scattering paradigm. These regimes are determined by the geometrical size of the ARIS units, the communication frequency, and the link distances from ARIS to the transmitter and the receiver. In particular, when the ARIS is within relatively short distances from the transmitter and receiver or the ARIS units are electrically large, e.g., their dimensions are ten times larger than the wavelength, the path loss undergoes the specular reflection paradigm. In contrast, the path loss undergoes the scattering paradigm when both dimensions of the ARIS units are ten times greater than the wavelength or the ARIS dimensions are very small. In other words, the scattering paradigm can be designated as the far field, whereas the specular reflection can be designated as the near field. The biggest difference between the two regimes is that, while the number of reflecting elements plays the dominant role in the received power in the specular reflection paradigm, the impact of the altitude of the aerial platform and the wavelength becomes more important in the scattering paradigm. Under both paradigms, the authors derived the minimum feasible number of reflecting elements and the maximum number of them that can be deployed on a given aerial platform’s surface. They also developed a link budget analysis for both reflection paradigms using the 3GPP and the log-distance channels models, based on which they derive the optimal horizontal location of the aerial platform at a given fixed altitude for both reflection paradigms. It has been shown that regardless of the altitude of the aerial platform, the optimal horizontal placement of the aerial platform in the specular reflection regime is over the perpendicular bisector of the segment from the transmitter to the receiver. However, in the scattering reflection regime, the optimal position becomes dependent on the height of the aerial platform. Particularizing the budget analysis to three different aerial platforms, namely UAV, HAP, and LEO satellites, the authors deduced that the specular reflection paradigm is not feasible for UAV and LEO satellites. The main reasons lie in the limited available surface on UAVs and the high altitude of satellites. However, the specular reflection paradigm can be realized using RIS-equipped HAPs, in which case they outperform TRIS. Under the same number of reflecting elements, and for both reflection regimes, the UAV achieves the best received power performance due to its closeness to the terrestrial users. However, HAPs become the best one if each platform can be equipped with as many reflectors as it can support. Furthermore, the authors showed that UAVs are insensitive to frequency, as increasing the frequency, while leading to higher attenuation, enables the use of more reflecting elements. On the other hand, the performances of HAPs and LEO satellites are affected by the attenuation caused by water droplets or molecules at specific frequency ranges. Nevertheless, their performances become insensitive to frequency over other spectrum regions, which can then be exploited to provide high-capacity communications.

1) Integrated UAV-ARIS Network: Focusing again on G2G systems, Shafique et al. [63] considered an integrated UAV-ARIS network in which a UAV carries a large array of reflecting elements to assist communication between two GUs. Comparison between three operating modes was investigated, namely, the UAV-only mode, the ARIS-only mode, and the UAV-ARIS mode, in which for the UAV-ARIS mode, both the UAV and ARIS forward the data and the receiving GU combines the data through selection combining. For each of these modes, the authors derived approximations for the OP as well as upper bounds on the EC and the EE. The results given in Fig. 8 directly showed that the integrated UAV-ARIS mode outperforms the other two modes in terms of EC and OP, which is due to the opportunistic selection between UAV-only and ARIS-only modes. Compared to the UAV-only mode, the ARIS-only mode presents better outage performance when equipped with enough reflecting elements; otherwise, the UAV-only mode will be a better choice.

2) Existence of Eavesdroppers: The work in [121] considered a UAV-carried RIS system to assist the communication between a BS and a single-antenna legitimate user in the presence of multiple ground single-antenna eavesdroppers, spatially distributed as a homogeneous 2-D Poisson point process. Focusing on the secrecy OP, the authors considered the cases of cooperative and independent eavesdroppers; in cooperative eavesdroppers, it is assumed that they apply maximum ratio combining to the

---

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
received signals before detection, while in independent eavesdroppers, the eavesdropper with the highest SNR is the most detrimental one and thus is supposed to be used to wiretap transmitted messages. Under the assumption that the eavesdropper has partial imperfect knowledge of the channel with the UAV, the authors derived approximations of the secrecy OPs for both cases. The accuracy of these results is assessed by simulation and clearly illustrates the ARIS's benefits of improving the secrecy performance of wireless communication systems.

3) Discrete Phase Shifts: The recent work in [50] derived the error probability expression for the direct communication between a satellite and ground IoT devices under Rician channels assisted by an ARIS deployed near the satellite. Taking the realistic discrete-phase ARIS into account, the error probability shown in Fig. 9 decreases as the number of phase resolution bits increases. In particular, the 3-bit ARIS design almost approaches that of the continuous-phase ARIS, which indicates that the 3-bit resolution level is cost- and energy-efficient for the satellite IoT devices communications. Another interesting experiment that has been conducted in this work is the error performance comparison when the phase shift at ARIS belongs to discrete finite phases uniformly distributed over the intervals $[-\pi, \pi)$, $(-\pi, -\pi/2)$ $\cup$ $[(\pi/2), \pi)$, and $[-\pi/2, \pi/2)$. The results showed that smaller phase intervals lead to obvious worse performance because it cannot compensate for the phase information of the cascaded channel.

4) Misalignment: In [65], the misalignment fading caused by the high relative velocity of LEO satellites and the sharp beam of THz antennas was considered when calculating the error probability of the intersatellite communication assisted by ARIS. Under the circular beam assumption, the misalignment coefficient is determined as a function of the Rayleigh-distributed beam's radial distance and the receiver's jitter variance. Similar to [63], the gain of the sum of the product channels was approximated by a Gaussian distribution based on the central limit theorem. Then, the closed-form expressions for the probability error rate under the assistance of multiple independent ARISs were also derived.

5) Imperfect Phase Compensation: The gains provided by ARIS-assisted systems require optimal phase compensation by reflecting elements. However, optimal phase compensation cannot be achieved in practice because of the quantization error and the channel variation. As far as UAV systems are concerned, channel variation occurs even when UAVs are hovering, making optimal phase compensation unrealistic. Al-Dweik et al. [122] and Al-Jarrah et al. [123], [124] studied the effect of the imperfect phase compensation error on the performances of ARIS-assisted systems. Under the assumption of a Von Mises distributed phase error, Al-Dweik et al. [122] and Al-Jarrah et al. [124] derived closed-form expressions for the SER and OP for less than three reflecting elements and obtained accurate approximations of them for a larger number of reflecting elements. The obtained results showed that a large number of reflecting elements confer robustness toward phase error. In contrast, the degradation caused by phase error may surpass the RIS gain when the number of reflecting elements is small. Under a similar setting, Al-Jarrah et al. [123] analyzed the capacity performance for the constant channel fading model. The obtained results showed that the capacity degradation due to phase errors is inversely proportional to the SNR, becoming negligible at high SNR. The SER and OP derivations are extended in [124] to the case of a multilayer UAV network, in which each layer is composed of one ARIS and a DF relaying UAV. In this case, increasing the number of reflecting elements at each layer or using more transmit does not improve significantly the performance at high SNR. The reason lies in that a DF relaying is used,...
and hence, the system performance is dominated by the worst hop, with the phase compensation error becoming the most dominant factor.

Summary: Table 2 summarizes the existing performance analysis works reviewed above. From a theoretical standpoint, the analysis of a communication link assisted by an RIS is less mathematically tractable as the system model involves the product of channels between the RIS and the transmission ends. To get insights into the performances, most of the works use either asymptotic or nonasymptotic approximations. The analytical complexity becomes even more involved when complicated system models accounting for sophisticated channel models or imperfection in the estimation of the noise and RIS’s phases are considered. Undoubtedly, the point-to-point NTN system performance is improved by using RISs. However, the reaped gain may be affected by the experienced channel conditions. Current theoretical works rely on the far-field assumption and consider unrealistic channel models such as Rayleigh, Nakagami-m, and Rician distribution. All these show that there is still room for further analytic developments based on accurate channel models that fit with experimental data.

V. OPTIMIZATION OF RANTNs

Based on the performance analysis results, the controllability of RIS and the flexibility of aerial platforms have generated excitement in the wireless community to fully explore the potential of RANTNs to optimize various performance metrics, such as SNR, data rate, EE, and power/time consumption. All these works serve as a solid foundation and provide valuable insights for future research of RANTNs.

A. Received Power/SNR

1) Terrestrial RIS: Several recent works considered the analysis of the performance gains brought by TRIS for G2A/A2G communications. Ma et al. [60] studied the signal gain defined as the ratio of the received signal powers with TRIS and without TRIS under the practical path-loss model for the urban macro scenario described in [125]. As GBS antennas are down-tilted, UAVs flying above GBS are only supported through sidelobes resulting in the UAVs receiving low received powers. By deploying TRIS within the coverage of the main lobe of the GBS, it is possible to gather more energy from the GBS and reflect it to the UAV, allowing for a significant improvement of the received power. Similar to terrestrial networks, the signal gain was found to increase quadratically with the number of reflecting elements. However, the UAV height also heavily influences the received power, whose impact becomes more significant when the UAV flies well above the GBS.

a) Multiple TRISs: On the other hand, the received power by a GU served by ABS is investigated in [51], where multiple TRISs are deployed to assist the transmission. Targeting the average received power maximization within multiple time slots, the authors proposed a joint optimization of the active beamforming at ABSs, the passive beamforming at TRISs, and the ABS’s trajectory. Interestingly, the results showed that the trajectory adaptation plays a more important role than the passive/active beamforming, which is reflected by the fact that the received power gain brought by trajectory optimization is much larger than that of passive beamforming and active beamforming. Moreover, because of the performance gain brought by TRISs, it was found that over the optimized trajectory, ABS tries to fly as close as possible to the locations of TRISs to enhance the received signal strength, especially when the number of reflecting elements is large.

b) Predictable movement: The recent work in [126] focused on a TRIS-assisted system when the aerial receiver position can be predictable. A typical system falling within this setting is that of uplink transmissions to an LEO satellite since the positions of LEO satellites are known a priori. The authors developed continuous-time multipath channels and derived the optimal configurations of TRIS that optimize the received power, the Doppler spread, and the delay spread. It has been shown that optimal received power can be achieved without incurring Doppler spread, while the delay spread cannot be optimized without increasing the Doppler spread and decreasing the received power. Moreover, simulations showed that TRIS with planar elements can provide SNR gain and resilience to visibility outages, provided that it has a favorable orientation to the satellite. Rotating the RIS can achieve such a favorable orientation, yet at the cost of an increasing delay spread.

2) Aerial RIS: The preliminary contribution regarding the received SNR of ARIS-assisted G2G system appeared in [40] and [127], where an ARIS is considered to assist the signal transmission between a source node and users located in a certain coverage area. Both works aim to design the transmit beamforming, the passive beamforming, and the ARIS location so as to maximize the worst received SNR over a coverage area of interest for a ULA at the ARIS in [127] and a UPA at the RIS in [40]. Considering a fixed location for the receiver, the authors derived closed-form expressions for the optimal transmit and passive beamforming, as well as the optimal placement of the ARIS. In particular, depending on the altitude of the ARIS, the authors showed that the optimal horizontal placement for a given location of the receiver could be either at the midpoint between the source and the destination or at one of two positions that are symmetric with respect to the midpoint. ARIS is further utilized optimally adjusted in [128] to enhance the signal transmission from HAP to the ground target area with the consideration of the unknown movement of HAP which is caused by the changes in the stratospheric wind and air density. Also, Lee et al. [129] proposed to deploy ARIS on a satellite to enhance the communication between another satellite and a GU, where both satellites are assumed to rotate on the same orbit.
Table 2 Summary of Existing Performance Studies of RANTNs

<table>
<thead>
<tr>
<th>Reference</th>
<th>Performance Metrics</th>
<th>RIS types</th>
<th>NTN types</th>
<th>Characterization</th>
</tr>
</thead>
<tbody>
<tr>
<td>[57]</td>
<td>OP, average BER, and average capacity</td>
<td>TRIS</td>
<td>G2G</td>
<td>Approximation</td>
</tr>
<tr>
<td>[59]</td>
<td>OP, SER, and EC</td>
<td>ARIS</td>
<td>G2G</td>
<td>Upper bound &amp; Asymptotic</td>
</tr>
<tr>
<td>[61]</td>
<td>E2</td>
<td>ARIS</td>
<td>G2G</td>
<td>Asymptotic</td>
</tr>
<tr>
<td>[62]</td>
<td>Link budget</td>
<td>ARIS</td>
<td>G2G</td>
<td>Approximation</td>
</tr>
<tr>
<td>[63]</td>
<td>OP, EC, and EE</td>
<td>ARIS</td>
<td>G2G</td>
<td>Approximation</td>
</tr>
<tr>
<td>[65]</td>
<td>BER</td>
<td>ARIS</td>
<td>A2A</td>
<td>Approximation</td>
</tr>
<tr>
<td>[117]</td>
<td>Link budget, OP and average outage duration</td>
<td>TRIS</td>
<td>A2G</td>
<td>Approximation</td>
</tr>
<tr>
<td>[118]</td>
<td>OP</td>
<td>TRIS</td>
<td>G2A</td>
<td>Exact &amp; Asymptotic</td>
</tr>
<tr>
<td>[119]</td>
<td>Coverage probability, BER, block error rate, goodput</td>
<td>TRIS</td>
<td>A2G</td>
<td>Approximation</td>
</tr>
<tr>
<td>[120]</td>
<td>OP, capacity</td>
<td>ARIS</td>
<td>G2G</td>
<td>Approximation</td>
</tr>
<tr>
<td>[121]</td>
<td>Secrecy outage probability</td>
<td>ARIS</td>
<td>G2G</td>
<td>Exact &amp; Asymptotic</td>
</tr>
<tr>
<td>[122]</td>
<td>OP and SER</td>
<td>ARIS</td>
<td>G2A</td>
<td>Exact &amp; Approximation</td>
</tr>
<tr>
<td>[123]</td>
<td>Capacity</td>
<td>ARIS</td>
<td>G2A</td>
<td>Exact &amp; Approximation</td>
</tr>
</tbody>
</table>

a) UAV perturbations: Mursia et al. [130] considered the application in which a BS transmits a signal through a hovering UAV-carried RIS to cover a target region over which users are distributed according to a 2-D Gaussian distribution. Unlike previous works studying similar system models, the perturbations of the UAV, typically caused by wind or other meteorological phenomena, are considered. Such perturbations are modeled by assuming the ARIS surface deployed on the UAV to experience undesired rotations on the 3-D axes, the angles of which are assumed to be mutually independent and normally distributed with zero mean and certain variances. Considering this model, the authors proposed to adjust the reflection coefficients of the ARIS so as to maximize the minimum SINR experienced by users located in the target area. Its gain becomes more important when the altitude of the ARIS is high and the number of reflecting elements is large, a setting that is seemingly more prone to misalignment since orientation perturbation is higher.

b) UAV swarm: In practice, the number of reflecting elements mounted on a UAV is limited by its payload and its battery capacity. According to [131], one solution to overcome this limitation is to deploy multiple UAVs that form a UAV swarm to enable cooperation between UAVs. Such a system was designated in [131] as a swarm-enabled ARIS-assisted G2G system. Contrary to TRIS, for which it is desirable to deploy RIS near the transmitter or the user so as to reduce the path loss of the cascaded channel, swarm-enabled ARIS systems may not present the same behavior since it needs to jointly consider the effect of LoS connection and propagation path loss. In fact, there exists an optimal position for ARISs that strikes a favorable tradeoff between path loss and NLoS probability.

3) TRIS and ARIS: To better compensate for the severe propagation path loss in a satellite communication system, Zheng et al. [132] proposed to deploy RISs both at satellite and GU sides. The proposed two-sided RISs-assisted system outperforms the one-sided RIS-assisted system with only one RIS deployed at the GU side or at the satellite side and the system without the assistance of RIS. An important insight concluded in this work is that the asymptotic performance gain achieved by the proposed two-sided RIS-assisted system scaling with the number of the reflecting element $M$ with order $O(M^4)$, which is higher than the widely considered one-sided RIS-assisted system with scaling order $O(M^2)$.

B. Interference

Besides enhancing the received power, RIS has been investigated to mitigate interference. In this line, the work in [55] considered the scenario of a downlink G2A communication system in which multiple GBSs communicate with aerial users. To extend the communication coverage, TRISs are deployed in each cell boundary. Accordingly, the aerial area is divided into two regions: GBS coverage area and TRIS coverage area. Aerial users in the GBS coverage area are directly served by the GBS, while GBS and TRIS collaborate to serve aerial users in the TRIS coverage area. The interference power in the neighbor cells is determined by the elevation angle of the interfering aerial user and the positions of the TRISs. Based on this observation, the work in [55] proposes to find the optimal TRISs’ positions to minimize the impact of interference to neighbor cells. This was shown to help prevent interference from spreading over a wider area.

C. Achievable Rate Maximization

The other important performance index widely considered in the literature is the achievable rate. According to where the RIS is integrated, the related works concerning achievable rate maximization can be classified into two categories: TRIS-assisted performance optimization and ARIS-assisted performance optimization.

1) Terrestrial RIS: The achievable rate of A2G communication system assisted by TRIS is first studied in [133], where a rotary-wing UAV serves a GU within a finite time slot. Assuming the existence of a direct link between UAV and GU, the work in [133] proposed a joint design of the UAV trajectory and the TRIS beamforming that targets
the maximization of the achievable rate. The joint design allows the GU to enjoy the most benefit of the channel gains from both the UAV and the TRIS, thereby obtaining the largest average achievable rate.

a) Existence of jammer: The TRIS is further being utilized in [134] to enhance the antijamming performance of a UAV communication system, in which a TRIS is deployed to assist the transmission from a GU to the UAV in the presence of a ground jammer. Targeting the maximization of the achievable rate, the authors proposed a joint design of the GU's transmit power, the TRIS's beamforming, and the UAV's trajectory. The simulation results indicate that deploying the TRIS near the ground jammer is more effective than deploying it near the GU when the jamming power is high since the received signal at the UAV is dominated by the jammers' interference in case of high jamming power. On the other hand, when the number of reflecting elements at the TRIS is large and the jammer's power is well reduced, it is more advisable to deploy the TRIS near the GU. This is because the reception at the UAV is no longer dominated by interference from the jammer and thus can be significantly improved by deploying a larger number of reflecting elements.

b) Existence of both GUs and aerial user: The recent work in [108] considered the communication between a UAV acting as an aerial user and the GBS through a TRIS in the presence of other GUs sharing the same spectrum by employing NOMA. Assuming a dynamic urban environment with possibly unexpected surrounding obstacles in low-altitude airspace, the authors proposed a joint design of the UAV trajectory design, the TRIS configuration, and the uploading power control to maximize the network sum rate while ensuring the UAV's flight safety through a collision avoidance mechanism. Specifically, the UAV is not allowed to fly over a forbidden zone around obstacles. This is achieved by constraining the distance between the UAV and any surrounding obstacle to be bigger than a certain threshold. The authors observed that the fixed phase shift algorithm presents a higher sum rate compared to the random phase shift one. This is because the random phase shifts at TRIS do not allow for signal alignment.

c) Multiple users: The above-reviewed literature analyzes and optimizes the achievable rate of point-to-point wireless communications by deploying TRIS in the propagation environment. The rate maximization problems can be naturally extended to MU scenarios [49], [83], [99], [100], [135], [136]. Consequently, the solutions will be more involved due to the interference and resource competition among different users. In this context, Li and Liu [135] considered an A2G MU downlink system in which a UAV serves simultaneously single-antenna users through a TRIS within a limited time. Under this setting, the authors proposed a joint design of the UAV trajectory and the phase shifts of the TRIS with the aim of maximizing the achievable sum rate. It has been found that the behavior of the UAV trajectory differs according to the considered flight period. In particular, for short flight periods, the UAV was found to follow a relatively direct path to each user and the TRIS and then remains stationary over those points as long as possible. Meanwhile, the faster the UAV approaches the hovering points, the larger the average sum rate. On the other hand, under a long flight period, the UAV bypasses the GUs and flies almost directly toward TRIS in an arc path for close connection. The same system model is studied in [83] but in THz bands, where the minimum average achievable rate is maximized. It has been shown that the deployment of TRIS allows for shortening the trajectory of the UAV which in turn implies less consumed energy. Targeting MU communications, the work in [49] investigated a TRIS-assisted UAV OFDMA system in which the phases of the TRIS are aligned with respect to one selected user in each time slot, while the remaining users are solely served by the UAV directly over other subcarriers. Under this setting, this work focused on the joint optimization of the UAV trajectory, the user scheduling, and the resource allocation to maximize the system sum rate over QoS requirements for each user. It was shown that the obtained optimal power allocation follows a multilevel water-filling principle, while the optimized user scheduling approach indicates that a user with a higher composite channel power gain or a more stringent data rate requirement has a higher chance to be scheduled as a TRIS-assisted user. Furthermore, with a fixed data rate requirement, employing a TRIS and appropriate power allocation scheme can scale down the transmit power of the UAV by the inverse of the number of reflecting elements, which is consistent with the power scaling law in [137]. As a major outcome of the authors' work, it has been shown that the size of the TRIS significantly affects the trajectory of the UAV. Assuming a not too large size of TRIS, the UAV directly flies toward centroid-formed users for maximizing the system sum rate. This is because the minimum data rate constraint of the TRIS-assisted user can still be satisfied even if the UAV is far away from it. However, as the number of reflecting elements increases, the UAV in the proposed scheme would first detour to the TRIS and TRIS-assisted user at the beginning before flying to other users. In fact, equipping more reflecting elements allows the TRIS to reflect the radiated signal more efficiently, and thus, approaching the TRIS and the TRIS-assisted user becomes more beneficial to the system sum-rate performance. Moreover, an additional practical assumption in [49] is that the composite channels from the UAV to GUs are frequency and spatial selective, which is different from channels in narrowband TRIS-assisted communications. More specifically, for each GU, the channel gain presents a cosine pattern with respect to the subcarrier index. Moreover, on each subcarrier, the users' channel gains fluctuate with a cosine pattern depending on the differences in propagation distances of the UAV-to-GU and UAV-TRIS-GU links. The simulation results also showed that the adopted OFDMA scheme outperforms the TDMA scheme, and the corresponding performance gain...
increases with the total transmit power. This is because OFDMA enables more efficient utilization of the power budget by exploiting the inherent MU diversity with flexible subcarrier allocation.

\( d) \) Multiple aerial platforms: More recently, the achievable rate has been considered to optimize the trajectory and the phase shifts of TRIS in multi-UAV NOMA networks [54] in which multiple UAVs are deployed to serve user groups through a TRIS. The optimized trajectory has revealed that UAVs are likely to be deployed near the served users to ensure the communication quality while remaining considerably far away from the unserved users to reduce the intergroup interference. Comparisons with scenarios without TRIS have shown that the deployment of TRIS enhances the channel power gains between UAVs and their served users, especially those located close to TRIS, while mitigating the interference between UAVs and their unserved users. Cao \( et \ al. \) [53] maximized the overall capacity of a system involving multiple UAV-GU pairs, where the reflecting elements on a single TRIS are divided into multiple groups to serve different pairs of UAVs and users. It was found that increasing the number of the TRIS's groups resulted in worse performances in terms of achievable rate and power consumption, as this implies fewer available TRIS elements in each group. Moreover, assuming a fixed number of TRIS groups, the system throughput decreases as the number of UAV-GU pairs increases because more bandwidth is consumed with poor communication quality.

e) Multiple TRISs: On the lookout for better performances, researchers started investigating the use of systems assisted by multiple RISs. In this context, the work in [52] and [84] considered multiple RISs mounted on several buildings to enhance the communication quality between the downlink transmissions of one UAV to several GUs. Wang \( et \ al. \) [52] proposed to optimize the trajectory and the reflection coefficients over a set of discrete values for the phase shifts. Although the discretization causes a loss in the performances compared to optimization over continuous variables, it is often preferable for implementation due to the hardware limitations in practice.

2) Aerial RIS: A recent wave of research contributed to investigating the data rate of ARIS-assisted networks. In this context, Zhang \( et \ al. \) [42] considered an ARIS carried by UAV to assist the communication between a GBS and a GU. Once the GU receives an SNR lower than the threshold yielding a blockage, the ARIS will move to a new place to rebuild an LoS link for downlink transmission. The authors proposed a method based on RL that jointly optimizes the reflecting elements of the ARIS and its location to maximize the average sum rate. The obtained average rate increases faster with the transmit power than for a static ARIS, while it decreases as the altitude increases due to an increasing path loss. Comparisons with benchmark algorithms showed that the proposed method yields an LoS probability over 90%, while the static ARIS and the ARIS)

![LoS probability and date rate of direct transmission scheme, TRIS scheme, ARIS scheme without trajectory optimization, and ARIS scheme with trajectory optimization. Please refer to [41] for simulation parameters.](image_url)

Fig. 10. LoS probability and date rate of direct transmission scheme, TRIS scheme, ARIS scheme without trajectory optimization, and ARIS scheme with trajectory optimization. Please refer to [41] for simulation parameters.
process in [41] is divided into two processes named communication process and movement process. In the communication process, the position of the ARIS is assumed to be fixed, while the transmit beamforming and the ARIS phase shifts are optimized to maximize the sum rate. In case the average rate per GU becomes below a certain threshold, the communication stage ends, and a movement process starts to optimize the location of the ARIS so as to maximize the long-term downlink communication capacity. Compared with the direct transmission scheme, the TRIS scheme, and the ARIS scheme whose trajectory is not optimized, the ARIS with optimized trajectory yields gains in both LoS probability and achievable sum rate. As shown in Fig. 10, the TRIS, ARIS without trajectory optimization, and the ARIS with trajectory optimization improve the LOS probability of direct transmission system from almost zero to 50%, 60%, and 90%, respectively, and also enhance the achievable data rate by over 60%, 95%, and 150%, respectively. The recent work in [101] described in Section III-C considered the maximization of the throughput of an ARIS-assisted A2G communication system to serve multiple users. More specifically, an ABS equipped with a high-capacity battery and a UAV-carried RIS with a low-capacity battery leave the charging station to provide communication services to multiple GUs, before returning once their battery is empty. Under this setting, the authors proposed an optimized solution of the optimal collaboration between the ABS and the ARIS from the perspective of power control, UAV trajectories, and task termination. The authors showed that when the ABS is not required to charge ARIS, it is likely to keep quasi-stationary over the served GU until the task termination. However, in the charging scenario, both ABS and ARIS tend to terminate the task near the charging station since the farther away they are from the charging station at the termination step, the more energy they require to return.

c) Multiple ARISs: The ARIS-assisted multicast communication scenario was studied in [86], where the GBS broadcasts a common signal to multiple GU clusters, each served by its associated ARIS. The authors considered the optimization of the beamforming at the GBS, the placement of the ARISs, and their respective beamforming to maximize the minimum users’ rate. They showed that the optimized deployment of each ARIS tends to balance the link between GBS-ARIS and ARIS to the centroid of each GU cluster to ensure that all users in the cluster can be served effectively.

3) TRIS and AIRS: Yu et al. [98] investigated the use of both TRIS and two ARISs carried by UAVs to maximize the achievable rate of an SWIPT system in which the BS communicates with an IoT device. Under Rician channel conditions between all devices and assuming both perfect CSI and statistical CSI, the authors proposed an optimized solution to design the transmit beamforming at the GBS, the phase shifts at the TRIS, and the flying ARISs and their associated trajectories. They noted that the optimized achievable rate under perfect CSI is higher than the one under the statistical CSI scheme. However, the gap between them decreases as the Rician factor of the considered channels increases. Moreover, the ARISs tend to fly close to the multiple-antenna GBS and spend more flying time hovering over it than over the IoT device. However, their takeoff and final location heavily influence the flying trajectories of ARISs.

D. Energy Consumption

It is another important issue that limits the superior performance of RANTNs. Indeed, due to the fact that the devices, such as IoTs and aerial platforms, are battery-powered, energy consumption is one of the most important challenges in commercial and civilian applications. The limited endurance of batteries hampers the practical implementation of the RANTNs. The advantages of battery-limited device-enabled wireless systems could not be reaped without appropriate solutions of how to overcome the energy-hungry issue.

1) Terrestrial RIS: As stated in [109], the total energy dissipated by the UAV is mainly used for supporting the hovering and mobility of the UAV (usually more than 95%). In order to further reduce the energy consumption of the UAV, it is better to keep the UAV aloft without consuming its own energy. Based on this property, TRISs with low-cost passive elements, which can extend the coverage while keeping UAVs aloft, are introduced in [109] to assist aerial communications and reduce energy consumption. This work proposed to control the phase shift of the TRIS instead of changing the position of the UAV to establish LoS wireless links between continuously roaming mobile users. By invoking this protocol, the UAV can maintain hovering status except when concatenated virtual LoS links cannot be formed even with the aid of the TRIS, which minimizes the total energy dissipation of UAV and maximizes the endurance of the UAV. Intuitively, invoking more reflecting elements leads to the reduction of energy consumption, while serving more GUs results in an increase of energy consumption. Moreover, when the altitude of the UAV is high, it consumes more energy due to the increased path loss. Referring to the work in [48], described in Section III-C, where the authors mount TRIS on two buildings and PDBs to relay energy and data between a UAV and IoTs, the authors noted that the optimal UAV position that minimizes the energy consumption depends on the considered scenario. In particular, if one PDB is considered, the optimal UAV’s hovering position coincides with that of the PDB. However, if we account for the UAV’s energy to reach the optimal position, the best hovering position occurs before the UAV arrives at the same position as the PDB. This is because there is a tradeoff occurring between the UAV’s energy flying, the charging energy, and the data transmission energy. Cai et al. [47] considered the case in which a multiantenna UAV serves multiple GUs adopting...
TDMA scheme through either a direct transmission mode or a reflecting mode using TRIS. Under this setting, the authors proposed a model for the consumed energy that accounts for the energy consumed by the TRIS whenever it is used to reflect the incoming signal. With this model at hand, the problem of interest is to optimize the transmit beamforming, the phase shifts of TRIS, the UAV trajectory, and the transmission mode for each GU in order to minimize the transmit power under minimum rate constraints. It was noted that with the help of TRIS, the UAV is not required to hover at the positions of the served users to satisfy the minimum rate constraints, as in the case without TRIS. In other words, TRIS can improve the flexibility of designing the UAV’s trajectory, which can lead to a substantial reduction in power consumption.

a) Multiple TRISs: The transmit power optimization becomes more critical when the VLC technology is used. In this case, it is necessary to provide wireless services and illumination for users simultaneously. In this context, Cang et al. [76] considered a VLC-enabled UAV network consisting of a group of UAVs to provide communication and illumination to GUs through multiple RISs. The problem of interest is to minimize the transmit power of UAVs via adjusting four variables: the UAV deployment, the phase shift of RISs, the user association, and the RIS association under the constraints of data rate and illumination demand. The simulation results show that the user association plays a more important role than the other three parameters, which indicates only optimizing user association when an urgent deployment is required. Moreover, the application of the proposed algorithms revealed that UAVs should be deployed to be surrounded by their associated users and RISs. On the other hand, a small number of users distant from the majority of them can be served by one UAV, resulting in reduced transmit power. Furthermore, RIS should be associated with the closest UAV, as this reduces the path loss and thus enhances the channel gains between UAV and GUs.

b) Integrated HAP and satellite communications: The work in [140] considered downlink communications between a satellite sending a common signal to ground satellite users and an HAP serving multiple HAP GUs over Rician fading channels. Both systems operate on the same frequency band and are aided by a TRIS. The authors formulated the problem of jointly designing the beamforming at the HAP and the TRIS to minimize the transmit power at the HAP under constraints of minimum SINR OP at each SAT and HAP user and taking into account Gaussian CSI estimation errors and correlation between the reflecting elements of the TRIS. The authors found that the transmit power at the HAP decreases when a larger number of reflecting elements at the TRIS is employed or the constraints on the minimum SINR OP are more relaxed. It also gets decreased with the increase of the Rician factor of the HAP users’ channels. However, the transmit power grows when the channel variance error increases or when the correlation factor between the reflecting elements of the TRIS increases.

2) Aerial RIS: The work in [63] considered the optimization of the power consumption of an ARIS-assisted system in which a UAV equipped with RIS assists communication between two GUs. A model for the total power consumption is provided, which accounts for the hovering power consumption as well as the power for data transmission and that consumed by hardware. The power consumption is then minimized subject to rate constraints by optimizing the number of active RIS elements. It is noteworthy that optimizing the power consumption with respect to the number of RIS elements in an ARIS system is crucial for two reasons. First, the number of RIS elements that can be deployed on a UAV is limited by its size. Second, each RIS element consumes energy, which while being low, may lead to significant power consumption depending on the considered phase resolution for each element. Focusing on multicell systems, the recent work in [141] proposed a design based on ARIS to reduce the power consumption of a free-cell massive MIMO system composed of multiple-antenna GBs, each serving a group of multiple single-antenna ground sensors.

a) Multiple ARISs: The prospect of better performance in reducing the total transmit power has led researchers to investigate the use of multiple ARISs. In this context, the work in [87] described in Section III-D2 investigated the deployments of ARISs to improve the performance of downlink transmissions of a macro GBs’ network overlaid with a small GBs’ networks and operating over different frequency bands to serve multiple users uniformly distributed over the coverage area of the macro BS. More specifically, ARISs are deployed to operate on the same frequency band as the macro GBs’ network with the goal of sustaining a strong connection with its associated GUs. Under this setting, the authors proposed a joint design of the ARISs’ and the BSs’ beamforming, the trajectory/velocity of each ARIS, and the subcarrier allocation for both macro and small BSs to minimize the total transmit power under minimum data rate requirements. Simulation results illustrate the effectiveness of the proposed solutions to reduce the transmit power. They also showed that using a higher number of ARISs is better than increasing the number of reflecting elements at each ARIS. The reason lies in the availability of more DoFs in the design of the trajectory of the ARISs.

E. Energy Efficiency

It cannot be denied that designs based on minimizing the power consumption sacrifice the achievable data rate. To strike a tradeoff between the consumed power and the achieved data rate, the maximization of the EE is more advisable. In short, EE advocates for smaller energy consumption in regard to performing higher.

1) Terrestrial RIS: Referring to the work in [63] described in Section V-D, the authors showed that the
EE first increases and then decreases as the number of TRIS elements increases. The reason is that the effect of capacity dominates that of power consumption for a small number of TRIS elements. However, as the number of TRIS elements increases, the power consumption prevails over the capacity. An optimal number of RIS elements that maximize the EE exist. However, its value depends on the power consumed by each TRIS element. In particular, suppose that the power consumed by each element is small. In that case, the EE continues to increase for a wide range of values of the number of elements because the increase in the number of TRIS elements does not significantly increase the power consumption, whereas the capacity keeps increasing. On the other hand, when a high resolution is used, the power consumed by each element increases. As such, the impact of power consumption of TRIS elements dominates that of the gain brought by using more of them, leading to a decrease in the EE beyond a specific number of TRIS elements. It can be concluded that if the bit resolution power is very small, using the maximum number of TRIS elements is optimal, whereas when the bit resolution power is significantly large, using a small number of TRIS elements is optimal.

a) Multiple UAVs: Targeting the optimization of the EE, the work in [142] considered a multi-UAV wireless network in which each UAV serves a specific cluster of GUs through a common TRIS. The deployed TRIS would be required to enhance the received signal’s quality at the GUs from the associated UAV while mitigating the interference from other UAV transmitters. The authors proposed a joint design of the power allocation at the UAVs and the TRIS’ beamforming.

b) Integrated access and backhaul links: The recent work in [111] investigated the use of TRIS to improve the performance of a UAV-assisted integrated access and backhaul network, in which the UAV collects through an access link aided by the TRIS data coming from GUs before forwarding them to the micro GBS through the backhaul link. Both backhaul and access links are assumed to operate over different frequency bands. Under this setting, the authors developed a dynamic resource management framework that aims to optimize the end-to-end EE with respect to the RIS elements’ phase shifts, the system bandwidth split between access and backhaul, and the transmission powers of both the UAV and the users. The simulation results demonstrate that the use of TRIS can provide almost 1.5 orders of magnitude higher sum users’ EE, for 300 reflecting elements at the TRIS, compared to the case without TRIS.

2) Aerial RIS: The EE maximization problem of an ARIS-assisted MU multicell uplink and MU downlink communication system is studied in [112] and [143], respectively, in which the communication between GBS and GUs is assisted by a hovering UAV equipped with ARIS. The total dissipated power consists of the transmit power at the ground transmitter, the hardware-related power that is consumed by the ground equipment, the RIS power consumption depending on the resolution of the reflecting elements, and the UAV power consumption. Both works considered the optimization of the beamforming at the GBS and the ARIS, as well as the power allocation resource. The work [143] shows that the achieved EE increases as the transmit power at the GBS increases. On the GBS side, increasing the number of antennas clearly yields a higher EE. Moreover, if ARIS is closer to the users, higher EE can be achieved than when the ARIS is close to the GBS. On the other hand, the ARIS located between the GBS and the GU achieves lower EE since the NLoS link becomes dominant and the reflected signal becomes weaker due to a higher path loss. Similarly, for the uplink multicell system studied in [112], the authors showed the EE performance deteriorates once the ARIS is not deployed in an appropriate position, typically when getting closer to either of the GBSs. Despite considering different systems, both works in [112] concluded that ARIS yields higher EE compared to the scenario with conventional UAV operating in AF mode. This is the consequence of the AF relaying being an active device and consuming power to amplify the incoming signal, thus having higher energy consumption compared to the RIS. On the other hand, an increase in the rate requirement reduces the achievable EE as the GBS will need higher transmit power to meet the users’ demands. Finally, it is important to note the paper in [58], which unlike all aforementioned research works focusing on MU downlink communication systems, considered the use of ARIS to allow for high data rate yet energy-efficient backhaul links. In particular, the system model of interest is the downlink backhaul link between a GBS and multiple ABSs through the assistance of an HAP carrying an ARIS. In this work, the authors proposed partitioning ARIS into parts; each serves a group of ABSs with a channel gain above the half-power beamwidth of the beamforming gain, which is shown to enjoy a high EE by ensuring a source-nearby ARIS placement.

F. Physical Layer Security

The broadcast and shared nature of the wireless medium makes it challenging to guarantee the secrecy of wireless communication networks. Physical layer security has emerged as a new technique to significantly improve the communication security of wireless networks by exploiting the communication channel with unauthorized users. In theory, any achievable secrecy rate representing the quantity of information transmitted from the source to the destination that unauthorized users cannot decode could not exceed an information-theoretic limit known as the secrecy capacity. This latter depends on the channels between all communication nodes, and thus, if all nodes remain static, the achievable secure rate remains below the secrecy capacity, regardless of the deployed techniques. To further boost the security performance, a potential solution is to leverage UAVs, which, due to their high
mobility, offer opportunities of creating stronger LoS links between legitimate users to boost the secrecy capacity. However, such LoS links can be maliciously leveraged by eavesdroppers to increase their SNR, posing a potential security risk in UAV communications. Recently, several works demonstrated that the use of intelligent reflecting surfaces is key to solving all these issues by opening up possibilities of reshaping the propagation channels to improve the secrecy rate of wireless communication systems.

1) Terrestrial RIS: Sun et al. [81], Wang et al. [144], Pang et al. [145], [146], and Li et al. [147] studied the problem of maximizing the secrecy rate of a TRIS-assisted UAV network, where an ABS transmits information to a legitimate receiver while a passive eavesdropper tries to intercept this information. For this system, the authors proposed different solutions for which they assessed the improvement in performance due to the deployment of TRIS and the use of more reflecting elements. In this context, Wang et al. [144] showed that the improvement due to TRIS depends on the distance between the receiver and the TRIS. They concluded that when TRIS is far away from the receiver, the scheme without TRIS achieves a similar performance to that with TRIS. The gain of the TRIS becomes more significant when it is near the receiver, allowing for full exploitation of the passive beamforming to enhance the desired signal at the legitimate receiver. Other factors were also studied in the aforementioned papers, including the impact of increasing the number of transmitting antennas or the transmit power at the ABS. Liu et al. [148] considered a more involved TRIS-assisted system in which multiple legitimate GUs are served by the ABS through the TDMA mode in the presence of one eavesdropper. Under this setting, the authors investigated the maximization of the minimum average secrecy rate among users with respect to the users’ scheduling scheme, the reflecting beamforming, the UAV’s trajectory, and the transmit power. The authors showed that as the number of reflecting elements increases, the ABS tends to hover closer to the TRIS to provide a better reflection. The solution provided in the aforementioned work [147] was also extended to the case of a TRIS-assisted A2G system with MUs and multi-eavesdroppers. Li et al. [147] showed that in this case, the minimum secrecy rate of the considered system presents a decreasing trend with a larger number of legitimate users because more legitimate users would share the beamforming and the array gains of the system.

a) Artificial noise: All the works mentioned above are based on leveraging the impinging signal on the RIS from the aerial platform to degrade the quality link with the eavesdropper. This approach can be impractical if the link with the RIS undergoes a high path loss. A possible solution to address this issue is to rather produce a strong interfering signal to the eavesdropper. Such a solution was adopted by Sun et al. [81] where they studied the secure transmission problem in a TRIS-assisted UAV network in the presence of an eavesdropper. Unlike previous works, the RIS is primarily deployed here to overcome the blockage between the ABS and the GU, while the artificial noise is used to degrade the reception at the eavesdropper. The authors showed that the use of artificial noise is an efficient technique, especially when the eavesdropper is close to the legitimate GU. The use of artificial noise was also used to help protect satellite downlink communications. In this context, Xu et al. [115] considered a satellite and a GBS to serve a satellite user and a terrestrial user, respectively. To improve the secrecy of satellite communications, a TRIS is used to reflect the interference caused by the terrestrial communications in a way that degrades the communication between the eavesdropper and the satellite. The authors observed from the simulation results that increasing the number of reflecting elements or the transmit power of the GBS yields lower SNR at the eavesdropper. However, increasing the number of antennas at the GBS does not necessarily decrease the SINR at the eavesdropper despite the increasing spatial DoFs. This is because the spatial DoFs are rather used to satisfy the constraints on the SNR of the terrestrial and satellite users.

b) Imperfect CSI: In practice, the eavesdropper usually makes it difficult for legitimate users to detect or track its location. Hence, a more plausible assumption is to assume that the estimated eavesdropping channels are inaccurate. In this context, Li et al. [149] proposed a deterministic uncertainty model, according to which the true eavesdropper’s channel is within a ball centered at their associated estimates with a given uncertainty radius. Compared to the system model described in [144] and [150], the work in [149] assumed that each time slot is divided into downlink communication and uplink communication. Under this setting, the problem of interest is to jointly optimize the trajectory of the UAV, the phase shifts of the TRIS, and the power of the UAV in both downlink and uplink so as to maximize an average of the worst weighted sum of the uplink and downlink achievable secrecy rate over the eavesdropper’s uncertainty region. As expected, the average worst case secrecy rate decreases as the CSI uncertainty of the wiretap channels increases since the high uncertainty in the eavesdropper’s channel makes it difficult to achieve a robust design and can also lead to the failure of RIS’s passive beamforming. In the same vein, Guo et al. [82] studied the impact of outdated CSI on the secrecy performance of a system composed of one ABS communicating with multiple legitimate GUs through a TRIS in the presence of multiple ground eavesdroppers. The authors noted that a loss in performance is incurred under outdated CSI, thereby showing its negative effect on hindering system performance improvement. A different CSI model is considered in [150], according to which the position of the eavesdropper is known, but its small-scale fading with respect to the TRIS could not be acquired. Under this setting, the authors derived a lower bound on the secrecy rate. Next, they went on optimizing it with respect to the UAV’s trajectory, the TRIS’s phase shifts, and the power transmitted at each slot under total...
and peak power constraints, as well as given fixed start and end positions.

The obtained results showed that for long flight duration, the trajectory of the UAV is such that it flies to a hovering position before it flies again to the destination, which agrees with the results obtained in [145] and [147]. In particular, the UAV hovers over a position close to the TRIS to leverage the reflect beamforming gain, while it hovers over a position near the receiver in case of without TRIS.

2) Aerial RIS: Besides improving communication links, ARIS can also be used to enhance the secrecy of RANTN networks [116], [151]–[154]. In this context, the works in [151] and [152] considered the problem of secure transmissions between a legitimate pair of source and destination ground nodes communicating through an ARIS under the presence of a ground eavesdropper. Under this setting, the problem of interest is the maximization of the secrecy rate with respect to the deployment and the passive beamforming of the ARIS. A similar system was also considered in [153] with the main difference that it additionally optimizes the transmitting beamforming power at the GBS. In [153], the eavesdropper is set closer to the GBS than to the legitimate GU, which leads to the ARIS gradually flying to the eavesdropper to reduce its received power. In addition, the authors noted that the joint design they proposed ensures an improvement of approximately 81% in the secure transmission rate compared to that under the random reflecting beamforming. Long et al. [154] considered an uplink wireless communication network in which legitimate users communicate through an ARIS-assisted UAV with a GBS in the presence of an eavesdropper. Under this setting, the authors focused on maximizing the secure EE of the system defined as the ratio between the secrecy rate and the total consumed power via jointly optimizing the UAV’s trajectory, the ARIS’s phase shifts, the user association, and the power allocation under maximum per-user power constraints, and UAV’s trajectory-related constraints. Simulation results demonstrated that the designed system could enhance the secure EE by up to 38% gains, compared to the traditional schemes with an AF relay. As for the ARIS, it tends to establish communication with the users away from eavesdroppers to achieve high secrecy rates. Moreover, as the considered ARIS is a totally passive reflecting structure and does not cost any specific energy, the secure EE increases as the number of reflecting elements becomes large. On the other hand, with the ARIS’s getting at higher positions, the secure EE decreases because of the increase in the communication distance. Furthermore, the UAV follows a closed-loop trajectory over which it sequentially visits all users. However, if the maximal flying speed decreases, the UAV is found to fly closer to GBS to secure confidential information transmission.

Interestingly, the authors noted that secure EE improves with the increase in the maximum transmit power but not always at the same pace, increasing fast at small powers but only slowly for high powers. This can be explained by the fact that for high transmit powers, not all the power would be used. As a result, using high powers would not significantly improve the secure EE.

a) Existence of jammer: The work in [85] considered the communication between a GBS-GU pair to serve multiple GUs, where at most one user is scheduled at a given slot, while the other GUs act as potential eavesdroppers. Contrary to previous works focusing on securing the transmissions, we were interested in ensuring covert communications whose goal is to ensure the transmissions to be undetectable at the eavesdroppers. For that, they proposed to use one UAV-carried RIS to ensure reliable transmissions and another UAV to send a jamming signal in order to degrade the detection performance at the eavesdropper. For this setting, they derived the minimum detection error rate at the eavesdropper and studied the problem of maximizing the average EE under the constraint that the minimum detection error rate is greater than a certain threshold. The authors noted that the minimum average efficiency improves with the increase in the number of reflecting elements before saturating for very large values, due to the considered minimum detection error rate constraint.

b) Collaborative beamforming: To further enhance the secure achievable data rate, the collaboration between the transmit users can be envisioned. In this context, the work in [155] considered an ARIS-assisted multisensor IoT system in which several nodes collaboratively beamform the transmitted signal to be reflected by an ARIS to a GBS in the presence of an eavesdropper. The problem of interest is to maximize the achievable secrecy rate via jointly optimizing the collaborative beamforming weights of the sensor nodes, the trajectory of the ARIS, and the ARIS’s phase shifts. For that, the cases of availability and nonavailability of the eavesdropper’s channel are considered. Simulations showed that the UAV tries to be far from the eavesdropper while keeping a reasonable distance with the receiving GBS. Moreover, if the channel with the GBS and the eavesdropper is strong, the beamforming design accounting for the eavesdropper’s channel outperforms the one ignoring it. In contrast, both schemes present similar performances when these channels are weak. This suggests that, in noisy environments, it is a viable strategy to use the scheme discarding the channel with the eavesdropper when the location of the eavesdropper could not be determined exactly. On the other hand, the average secrecy rate improves when the sensors become scattered over a larger area. This can be explained by the fact that the channel matrix aggregating all channels with the sensors becomes better conditioned, offering larger DoFs to protect the signal from the eavesdropper.

G. Time Delay

Several emerging smart city applications are based on the freshness of sensory data (i.e., status updates), which is being monitored and generated by a plethora of IoTDs.
Examples of such applications include smart environment monitoring, industrial control systems, and intelligent transportation systems, which all require reliability and timeliness in delivering status-update information. Outdated updates may be inconsistent with the current status of the physical process being monitored and controlled, which may lead to erroneous decisions. Several works in the literature considered the time delay as the metric of interest. For instance, the work in [48], described in Section III-C, aimed to design the UAV’s trajectory so as to minimize the total time delay. Considering the scenario in which multiple IoTDs are scheduled using TDMA to transmit through an ARIS-equipped UAV and their status updates to a GBS, the authors used the expected sum AoI as their metric of interest, which represents the elapsed time since the generated/sampled of the most recently received status update. Under this setting, joint optimization of the altitude of the UAV, the communication schedule, and the ARIS’s phase shift was proposed so as to minimize the AoI under SNR and ARIS altitude constraints. The use of ARIS is motivated by the fact that ARISs operate in a full-duplex relaying mode, which should lead to reducing the AoI. Comparisons with other schemes considering either UAVs following a random walk policy with random selection of IoTDs or hovering at a fixed position that satisfies reliability constraints for all IoTDs have shown that the proposed scheme presents a considerable gain in terms of AoI.

H. Error Probability

Ultrareliable and lower latency communications (URLLC) is one of the three diverse features to be offered by upcoming 5G networks next to enhanced mobile broadband and massive machine-type communication. URLLC is indispensable to support important use cases of mission-critical applications, including UAVs control information delivery, vehicle-to-vehicle communications, self-driving cars, intelligent transportation, tactile Internet, E-health, and industrial automation. Such URLLC applications require deterministic communications with ultrahigh reliability where the packet decoding error rates are $10^{-9}$ or even lower, depending on the considered mission-critical application. Therefore, Ranjha and Kaddoum [56] and Li et al. [156] studied and formulated the problem of ultrahigh reliability in URLLC assisted by the TRIS in a short packet transmissions communication scenario. In particular, the work [56] considered the setting in which a ground transmitter sends short packets to a UAV that reflect them through a TRIS to a ground receiver. Under this setting, the SNR between the ground transmitter and the UAV depends on the UAV position, whereas the SNR between the UAV and the ground receiver depends on the UAV position and the TRIS’s phase shifts. For each hop, the transmitted packet uses a given blocklength. Based on that, the authors derived an expression for the total decoding error rate that involves the packets’ blocklengths and the SNR in each hop. Targeting the minimization of the decoding error rate, they proposed a joint design of the optimal UAV’s position and blocklength in each hop under a total blocklength constraint. It is observed that optimizing the UAV’s position is a crucial step that can pave the way for ultrahigh reliable transmission of short packets. Moreover, when increasing the blocklength, the probability of decoding error rate decreases, but this comes at the cost of decreasing the message transmission rate. Similarly, the goal of minimizing the decoding error rate in URLLC scenarios was recently considered in [156] but for a system involving multiple ARISs to assist communication between a GBS and multiple GUs. Both works [56], [156] concluded that when the number of passive reflecting elements increases, the decoding error rate decreases dramatically. Such a behavior results from the improvement in spatial diversity brought about by the increase in the number of TRIS’s/ARIS’s elements leading to a higher total channel gain with the ground receiver. Beyond URLLC communications, the minimization of metrics involving error probabilities was recently considered for a TRIS-assisted P2P satellite communication system in which a satellite serves one user, both equipped with dual-polarized antennas [157]. Under this setting, the authors minimized the pairwise error probability with respect to the TRIS’s phase shifts and the transmit power. They noted that the performance gain introduced by TRIS is heavily influenced by the accuracy of the CSI.

In addition to assisting transmissions of existing systems, RIS can be used to enable simultaneous passive beamforming and information transmission. To achieve this functionality known as symbiotic transmission, RIS uses existing radio to carry information bits encoded by the ON/OFF states of its reflecting elements while adjusting their associated phase shifts to enable passive beamforming. In this context, Hua et al. [45] and Hua and Wu [158] proposed a UAV-assisted RIS symbiotic radio system in which a UAV played the role of a primary transmitter for multiple TRIS to help them transmit their signals to a GBS. For the sake of simplicity, a wake-up communication scheduling approach is adopted, according to which the UAV can assist at most one TRIS at each time slot. Under this setting, a metric involving the BER is optimized under minimum primary rate constraints with respect to the UAV’s trajectory and the TRIS’s beamforming and association. More specifically, in a first scenario, the work in [45] considered the maximization of the weighted sum BER of all TRIS, where a higher value weighting factor represents a higher priority over other TRISs. In this case, the UAV tends to sequentially visit all TRIS when all TRIS has the same weight factors to obtain the maximal sum rate, while, for unequal priority weights, it is likely to fly by the TRIS with a lower weight factor than hovering over it. In a second scenario, they considered in [45] and in another related work [158] the minimization of the maximum BER among all TRISs over all time slots. In this case, they showed that the UAV flying with maximum speed becomes
closer to each TRIS and then remains stationary above it for a certain amount of time.

I. Number of Served Devices

Al-Hilo et al. [103] considered the scenario in which an ABS tries to gather data sent by a set of IoTDs scattered in an urban environment, using FDMA. Each device alternates between an active and passive activation mode and has a period during which its information should be collected before it becomes of no value. To enhance the communication link between the IoTDs and ABS, a TRIS is used. The problem of interest is to maximize the number of IoTDs that were served under limited available resources and trajectory constraints. It has been observed that under the assistance of TRIS, ABS tends to have a complex routing to serve more IoTDs around TRIS, while the IoTDs located far away from the TRIS cannot be served due to the restricted speed at ABS and their poor indirect LoS. The performance in terms of percentage of served IoTDs degrades with adding more IoTDs because of the limited available resources. It also degrades when increasing the transmitted data size, due to the fact that smaller data can be uploaded within shorter periods, and as such, the ABS can have more free resources to serve more devices. However, both limitations can be tackled by increasing the number of reflecting elements at the TRIS. Comparisons with schemes using nonoptimized trajectories adopting random trajectory, random TRIS configuration, and stationary ABS revealed the importance of trajectory planning and TRIS optimization to adapt to more complex scenarios. A major important fact is that for short data sizes, the trajectory optimized scheme using random phases at the TRIS is more efficient than the same scheme but without TRIS.

VI. METHODOLOGY

The above-reviewed optimization works on RANTNs revealed that the system performance optimization is always one of the key issues in RANTNs. Not only because of the reconfigurable reflecting elements and the controllable communication environment but also due to the flexible and dynamic deployment of aerial platforms, which is different from the traditional terrestrial network. The NTNs introduce an important adjustable parameter, that is, the position/trajectory of aerial platforms. The main research effort that has been carried out is to develop efficient algorithms to jointly design the position/trajectory of aerial platforms, RIS configuration, and other system parameters such as transmit beamforming, resource allocation, and user scheduling. In fact, the formulated optimization problems in RANTNs are mostly involved with nonlinear and nonconvex objective functions and constraints, and the involved multiple optimization variables are often intricately coupled with each other, which makes them hard to solve. As a solution, many efficient methodologies are proposed, which can be mainly divided into two types, i.e., alternating optimization algorithms and machine learning-based algorithms.

A. Alternating Optimization Algorithm

To obtain a favorable performance with an efficient algorithm, the alternating optimization algorithm is adopted. It consists in decomposing the original problem into several subproblems regarding different optimization variables and solving them iteratively. Specifically, in each subproblem, one or two variables are optimized, while the other variables are assumed to be fixed. This procedure is repeated in turn until a specific convergence criterion is reached. For instance, Ge et al. [51] aimed at maximizing the received power at the GU in a TRIS-assisted UAV communication system by jointly optimizing the active beamforming at ABS, the passive beamforming at the TRIS, and the trajectory of ABS. With the alternating optimization framework, the formulated optimization problem is decomposed into three subproblems, and solutions of all variables are derived iteratively. Specifically, the optimal beamforming at ABS is solved by fixing the phase shift matrix and the ABS trajectory. Then, the optimal reflecting elements of passive beamforming are derived based on the fixed trajectory and the given optimal active beamforming. Finally, the ABS trajectory is optimized based on the given active and passive beamforming.

In recent works using alternative optimization algorithms, some of the subproblems admit a closed-form solution. For example, the optimal solution of the involved active beamforming design at the transmitter is verified to be the maximum ratio transmission precoding [23], [38], [40], [51], [143], [155] once other variables are determined. Also, in most of the subproblems, the power allocations problems [38], [49], [58], [113], [144], [149], [150], [155], RIS configurations [38], [40], [45], [51], [56], [83], [127], [133], [150], [155], and aerial platforms positions [40], [113], [127] have closed-form solutions. Optimization variables in decoupled convex subproblems without closed-form solutions can also be solved by existing algorithms or tools efficiently, such as the Lagrange dual method, CVX, interior-point algorithm, and linear search.

Sometimes, the reformulated subproblems are still nonconvex and cannot be solved by traditional convex optimization methods. Depending on the considered objectives and their associated constraints, researchers propose to adopt different algorithms for solving optimization problems with particular forms.

1) SCA: The SCA algorithm acts as one of the most popular methods for solving nonconvex optimization problem such as power control [54], [148], [154], RIS configuration [54], [98], [113], [147], [149], [151], and position/trajectory optimization of aerial platforms [49], [51], [54], [76], [83], [98], [100], [122], [133]–[135], [138], [144]–[150], [152], [153], [158]. Specifically, at each iteration, the original nonconvex functions are approximated by some convex upper bounds of them with the same first order behavior. Then, the approximate solution...
of the original problems can be obtained by iteratively solving these convex problems. Once the convex upper bounds are properly selected, SCA has been shown to achieve a favorable convergence behavior.

2) Semidefinite relaxation: The semidefinite relaxation technique constitutes an efficient tool to solve nonconvex quadratically constrained quadratic programs (QCQPs) in an almost mechanical fashion [159]. It is generally used in RANTNs to transform the problems into a standard SDP by relaxing constraints [76], [81], [113], [115], [130], [134], [149]. The relaxed problem is convex and can be easily solved by existing toolboxes.

3) Riemannian manifold: The unit modulus constraints usually required for phase shift optimization are one of the main obstacles toward solving the optimization problem. As a solution, Yu et al. [160] showed that it can be efficiently solved by manifold optimization, as the unit modulus constraints define a Riemannian manifold. Therefore, the Riemannian conjugate gradient algorithm is adopted in [135], [136], [138], and [153], thereby enabling an efficient optimization for the RIS configuration.

4) Fractional programming: For the problems whose objective function in a fractional program is a ratio of two functions that are in general nonlinear, the fractional programming method is preferred to be invoked. This method is especially useful in the scenarios concerning about the sum rate [41], [134], secrecy rate [144]–[146], power minimization [76], and EE [111], [112], [154].

5) Difference-of-convex: In addition, the difference-of-convex programming and its difference-of-convex algorithm also adopted in [143] and [144] to address the problem of optimizing an objective composed by the difference of functions. It is able to quite often gave global solutions to a lot of different and various nondifferentiable nonconvex optimization problems based on local optimality conditions and difference-of-convex duality. This method is proven to be robust and efficient, especially in the large-scale setting.

6) Quadratic transform: There are also some objectives in the form of ratios of concave and convex functions with reference to the same variable. Fortunately, due to the structure of the problem, the globally optimal solution can be obtained by applying quadratic transform proposed. The quadratic transform converts the ratio of concave and convex function to the convex form by introducing an auxiliary variable [63].

7) Binary constraint relaxation: Consideration of multiple users or multiple RIS leads to user/RIS scheduling problems with binary constraints. The nonconvexity of this kind of problems can be handled by relaxing the binary constraint to a continuous one [45], [47], [87], [148].

8) S-procedure: Xu et al. [140] and Li et al. [149] also adopted a Lagrange relaxation technique called S-procedure to transfer the quadratic constraints into linear matrix inequalities to ease the optimization process.

9) Penalty method: This method is widely used for solving constrained optimization problems through replacing a constrained optimization problem by a series of unconstrained problems whose solutions ideally converge to the solution of the original constrained problem [54], [86], [87], [158]. The unconstrained problems are formed by adding a penalty function to the objective function. The penalty function consists of a penalty parameter multiplied with a measure of violation of the constraints. The measure of violation is nonzero when the constraints are violated, while it is zero in the region where constraints are not violated.

In addition to these existing methods, several works also proposed other approaches to solve some specific optimization problems regarding a certain RANTN system. For instance, to ensure that all users in the particular area lie within the main lobe of the ARIS and connect to the transmitter successfully, the reflecting elements were divided into subarrays in [127]. The use of subarrays produces larger main lobes but comes at the cost of smaller peak gains. The complete and improved ARIS design regarding the subarrays division based on a novel 3-D beam broadening and flattening technique is provided in its extension work [40]. The phase shifts of the subarrays are designed to form a flattened beam pattern with adjustable beamwidth, catering to the size of the coverage area.

Summary: The alternating optimization is the most general method to get a favorable solution for the jointly coupled optimization variables in different system settings. As reviewed above, although the original problems can be divided into several subproblems, it is necessary to choose an efficient method for solving each subproblem whose objectives and constraints have certain properties.

B. Machine Learning-Based Algorithm

Alternating algorithms are able to obtain near-optimal performance, but most of them are highly complex and time-consuming. On the other hand, the time-varying and highly dynamic nature of wireless NTNs requires the proposed solutions to be easily implemented with low complexity. Such a goal sometimes could not be achieved by conventional optimization methods, especially when the considered communication system is complicated. A recent research activity attempts to investigate the use of machine learning techniques, known for their suitability to tackle nonconvex and sophisticated optimization problems with near-optimal solutions. As far as RANTNs are considered, machine learning stands as an efficient tool to handle highly dynamic wireless environment.

1) Reinforcement Learning: RL gathers a group of methods aimed to make decisions in a given environment.
Such an environment is typically modeled through a Markov decision process, which involves four-tuple $S, A, P, r$, where $S$ is the state space grouping all possible states, $A$ is the action space containing all possible actions, $P_a(s, s')$ is the probability that action $a$ in state $s$ at time $t$ will lead to state $s'$ at time $t + 1$, and the reward $r(s, a)$ is the immediate reward (or expected immediate reward) received due to action $a$ while being at state $s$. A reinforcement problem interacts with the MDP through three elements.

1) **Agent**: An entity that takes actions and receive accordingly a reward.

2) **Policy**: It is the mapping of each state $s$ to an action $a$ and is denoted usually by $\pi$.

3) **Value functions**: These include the state-value function returning the total expected future discounted rewards starting from a given state, and the state–action value function, known also as Q value function, which returns these rewards starting from a given state and a given action and following a policy $\pi$ or considering optimal policy.

RL involves different approaches, which can be decomposed into three main categories: model-based techniques, value-based techniques, and policy-based techniques. In model-based techniques, the agent alternates between two processes: estimating the model of the underlying environment and determining the optimal decision from the estimated model. In value-based techniques, the agent estimates the value function and finds the policy that optimizes it, while in policy-based techniques, the agent directly targets the estimation of the optimal policy without estimating the value function.

In the sequel, we will briefly describe popular algorithms from the categories of value- and policy-based techniques and explain how they are applied to solve trajectory planning problems in RANTNs.

a) **Q-learning**: It is one of the most popular algorithms in RL that belongs to the category of value-based techniques. It is based on learning a Q value function that can be exploited to find the optimal action while being at a given state. The Q value function is a quantity denoted by $Q^*(s, a)$, which measures the total expected value of the cumulative discounted reward of choosing action $a$ when being at state $s$ and then following the optimal policy. Formally, $Q^*(s, a)$ writes as

$$Q^*(s, a) = \mathbb{E}_{s' \sim P_a}(r(s, a) + \gamma \max_{a'} Q^*(s', a'))$$

where $\gamma$ is a discount factor. In practice, the agent performs either an exploration step or an exploitation step. In the exploration step, the agent tests new actions to update the Q-table storing the Q value estimates for every state and every possible action. In this step, the Q values estimates are updated based on the following rule:

$$Q^*(s, a) \leftarrow (1-\alpha)Q^*(s, a) + \alpha \left[ r(s, a) + \gamma \max_a Q^*(s', a) \right]$$

where $\alpha$ is the learning rate and $s'$ is the new state after performing action $a$. In the exploitation step, the agent exploits the Q-table and performs the action that presents the highest Q value for the current state. To find a good balance between exploration and exploitation steps, one frequently used method is the $\epsilon$-greedy exploration method in which the exploration step is performed with a decaying probability $\epsilon$. In this way, the agent will start by exploring the environment. As it acquired more accurate estimates for the Q values, it can start exploiting them to perform optimal actions.

b) **Deep Q-learning**: When the number of states is too high, Q-learning faces two major problems. First, the amount of memory required to update new states can be prohibitively high. Second, the exploration step to create the required Q-table would require a lot of time, making it impractical. One solution to this problem is to use a neural network to approximate the Q value function for any possible state and action. For that, the value function is parameterized by a parameter $\theta$, hence denoted by $Q^*(s, a; \theta)$, where $\theta$ are the weights of the used neural network. The aim of the neural network is to minimize the following loss function:

$$L(\theta) = \mathbb{E} \left[ (\text{target}(s, a) - Q^*(s, a; \theta))^2 \right]$$

where target$(s, a)$ is the target Q value, which represents a refined estimate for the expected future reward from taking an action $a$ while being in state $s$ and is given by

$$\text{target}(s, a) = r(s, a) + \gamma \max_a Q^*(s', a').$$

The main difference with classical deep learning is that here, the target changes constantly during the process, which may lead to instability issues in practice. One solution to this problem is to employ two neural networks, where the first one is trained to optimize the loss in (2) and the second one termed “target network” is a copy of an old version of the former used to provide the unknown target and is updated less frequently. Such a method is known as DQN and applies only when the set of actions is discrete.

c) **Policy gradient methods**: To handle the case of continuous state and action spaces, policy gradient methods have been proposed. Central to these methods is to represent the policy by a parameterized function of the parameter vector $\theta$ and optimize it via gradient ascent to maximize the expected return. Such a policy may be either deterministic ($a = \pi_a(s)$) or stochastic, representing the probability distribution of the action given a
state \( s \) \((a \sim \pi(\theta|s))\). One issue that policy gradient methods should address is how to estimate the gradient and particularly the \( Q \) value function that arises in its expression. The most natural way to estimate this function is by Monte Carlo averaging where states and actions are sampled at each episode according to the previous policy to update the \( Q \) value function and the policy parameter vector \( \theta \) for each sampled state and action. Such an approach is known as the REINFORCE algorithm and, while being simple, suffers from high variance gradient estimates and slow learning rate. As a solution to the limitations of the REINFORCE algorithm, hybrid architectures combining policy and value-based techniques have been proposed. They are often termed actor–critic architectures, and as their name suggests, they involve an actor to update by stochastic gradient the parameter \( \theta \) parametrizing the policy, and a critic to estimate the \( Q \) value function and feed it to the actor. Several reinforcement algorithms adopt the actor–critic architecture, of which we distinguish the deep deterministic policy gradient (DDPG), the trust region policy optimization (TRPO), and the PPO. While DDPG can be thought of as an extension of deep Q-learning for continuous action spaces, PPO and TRPO aim to pre-

2) Distributional RL: As seen above, RL algorithms such as Q-learning involve estimating the \( Q \) value function through the update rule in (1). Such a value represents the expectation of the cumulative discounted rewards. A different alternative to conventional RL that has shown impressive performance improvements is to rather estimate the distribution of the sum of the discounted rewards instead of their expectation. Thus, instead of updating the \( Q \) value function as in (1), the distribution of the cumulative discounted rewards is modeled by a parameterized distribution, the parameters of which are estimated so as to minimize a certain distance to the estimated target distribution. Based on the estimated parameterized distribution, the \( Q \) value function is then computed and the action associated with the maximum \( Q \) value function is taken. Building on this approach, distributional versions of Q-learning and DQN have been proposed and were shown to outperform their conventional counterparts. As far as RANTNs are considered, the distributional RL algorithm was applied in [41] where an RIS-equipped UAV is employed to enhance the downlink communication between the GBS and multiple moving GUs. Targeting the maximization of the transmitted data volume under a limited energy constraint, the authors proposed an alternating optimization algorithm to optimize the transmit beamforming and the ARIS’s phase shifts, while a distributional RL is used to update the location of the UAV once a blockage occurs. Targeting a TRIS-assisted A2G NOMA communication framework, Zhao et al. [108] also designed a distributional robust DRL algorithm based on the soft actor–critic framework to jointly optimize the UAV trajectory, RIS configuration, and power control. The proposed distributional RL network captures environment uncertainties brought by the unknown locations of obstacles by
integrating the partial distribution information, thereby guaranteeing the worst case performance, which is proven to outperform the conventional RL network in terms of learning efficiency and robustness.

3) Deep Learning: Deep learning has been proven to be a powerful tool for the development of data-driven algorithms. As it does not require knowledge of the data model, it has naturally been applied to predict information from data that cannot be easily modeled like photographs or audio recordings. Although in wireless communication, the propagation channel model is easy to model and the transmitted signal is man-made, deep learning can still play an important role in the development of future communication systems. As shown in [161], deep learning can be used for problems where an efficient algorithm is known but suffers from a prohibitively high complexity, making it impractical for real-time implementation. In this case, a DNN can be used to approximate the output of the high-complexity algorithm based on an offline training phase and then used instead for real-time implementation. Doing so, the complexity of the online computation is moved to offline training, which reduces the computational complexity and facilitates the implementation. In practice, the training dataset can be obtained by running the high-complexity algorithm on a multitude of settings. In this context, Cao et al. [53] applied this approach to the scenario in which multiple UAVs serve through a multigroup RIS multiple users on orthogonal subcarriers. Of interest is the problem of optimizing the RIS group allocation to each UAV–user pair as well as the RIS phase shifts to optimize the total throughput. The authors proposed a solution based on deep learning to approximate the outputs of the exhaustive search-based algorithm. The proposed solution develops a two-task learning model that performs a classification task to predict an integer allocation vector describing the RIS group allocation and a regression task to predict a real-valued vector representing the RIS phase shifts. Another application of deep learning in wireless communication is when a theoretical model describing the relationship between physical quantities, such as the propagation environment and some information of interest, could not be obtained. In this case, deep learning can help provide such a model. One example of such an application of deep learning to RANTNs is represented by Abuzainab et al. [88] where they considered a THz drone network in which a mobile drone user is served by a GBS and a flying RIS. Of interest is the problem of proactively predicting the best beamforming vector at the GBS and the best communication link between direct and RIS-assisted links. To account for the drone mobility, at each time, the prediction task is required to learn the temporal correlation in the learning sequence, so that it can predict the values of the best communication link and the GBS beamforming vectors from their past values. The authors have thus opted for a recurrent neural network based on GRUs, which has been proven to be effective in learning sequence dependency, especially long sequences. This architecture is trained over a training dataset built using the DeepMIMO generation framework in [162].

Another neural network architecture that has been used in RANTNs is the graph neural network, which, unlike conventional DL methods that exhibit a grid-like structure, can produce state-of-the-art solutions for problems involving data in irregular domains. Such a property can be useful for channel estimation tasks since the observed data frequently changes because of the dynamic nature of the propagation channel. In this context, Tekbiyk et al. [64] first used graph attention networks (GATs) to estimate the channels of RIS-equipped HAP assisted full-duplex communications, and the channels in the LEO-satellite enabled IoT communications with RIS deployed near the satellite [50]. The numerical results show that for the full-duplex channel estimation, the performance of the GAT estimator is better than the least squares. Contrary to the previously studied method, GAT has the ability to estimate the concatenated channel coefficients at each node separately. As such, there is no need to use the time-division duplex (TDD) mode during the pilot signaling in the full-duplex communication. Moreover, the numerical results also show that the GAT estimator is robust to hardware impairments and small-scale fading characteristics changes, event when the training data do not include these changes.

Summary: Further enhancements are achievable for RANTNs when leveraging machine learning algorithms to empower the integrated technologies with RIS and aerial platforms. The RIS/UAV will provide autonomous decision-making, knowledge extraction and prediction, and near-optimal optimization performance. Machine learning algorithms can be used for enhancing the channel estimation and spectral efficiency and balancing different tradeoffs by automatically learning from the collected data, the propagation environment, and their past experience. All the machine learning-based algorithms share the general property that more extra iterations are needed for convergence when the number of reflecting elements increases. The adopted machine learning algorithms in the existing literature aiming to find the optimal solutions for different control variables are summarized in Table 3. Despite the important number of tools that have been thus far applied, some existing tools have not been explored in the context of RANTNs, such as semisupervised learning or unsupervised learning without labeled data.

VII. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Although the existing literature we reviewed has made a great contribution to RANTNs, there are some critical unsolved problems undermining the full potential of this cutting-edge technology in practical implementation. Since the studies on the integration of RISs into NTNs are still
Table 3 Summary of Existing Literature Adopting Machine Learning-Based Algorithm

<table>
<thead>
<tr>
<th>Reference</th>
<th>Optimization variables</th>
<th>Optimization objectives</th>
<th>Machine learning algorithms</th>
<th>System model</th>
</tr>
</thead>
<tbody>
<tr>
<td>[39]</td>
<td>UAV’s altitudes, user scheduling</td>
<td>Sum age-of-information</td>
<td>DRL, PPO, MDP</td>
<td>ARIS-assisted G2G MU uplink system</td>
</tr>
<tr>
<td>[41]</td>
<td>CSI, ARIS’s deployment</td>
<td>Sum rate</td>
<td>Distributional RL, MDP</td>
<td>ARIS-assisted G2G MU downlink system</td>
</tr>
<tr>
<td>[42]</td>
<td>CSI, ARIS’s deployment</td>
<td>Capacity</td>
<td>Q-learning, MDP</td>
<td>ARIS-assisted G2G P2P downlink transmission</td>
</tr>
<tr>
<td>[50]</td>
<td>CSI</td>
<td>Mean square error</td>
<td>GAT</td>
<td>ARIS-assisted A2G/G2A satellite IoT SISO downlink/uplink communications</td>
</tr>
<tr>
<td>[52]</td>
<td>UAV’s trajectory</td>
<td>Weighted fairness for users and sum data rate</td>
<td>DQN</td>
<td>TRIS-assisted A2O MU downlink system assisted by multiple TRISs</td>
</tr>
<tr>
<td>[53]</td>
<td>TRIS’s phase shift and location, TRIS’s elements allocation</td>
<td>Throughput</td>
<td>Multi-task learning</td>
<td>TRIS-assisted G2A multiple UAV-GU pairs SISO system</td>
</tr>
<tr>
<td>[64]</td>
<td>CSI</td>
<td>Mean square error</td>
<td>GAT</td>
<td>ARIS-assisted G2G P2P full-duplex system</td>
</tr>
<tr>
<td>[82]</td>
<td>ARIS’s trajectory and phase shifts, transmitting beamforming</td>
<td>Sum secrecy rate</td>
<td>DRL</td>
<td>ARIS-assisted G2G MU secrecy communication system with multiple eavesdropper</td>
</tr>
<tr>
<td>[87]</td>
<td>ARIS’s trajectories and phase shifts, subcarrier allocation</td>
<td>Total transmit power</td>
<td>DQN</td>
<td>Multiple ARIS-assisted G2G heterogeneous network</td>
</tr>
<tr>
<td>[88]</td>
<td>Best communication link prediction, transmission beam prediction</td>
<td>Prediction accuracy</td>
<td>Recurrent neural network, ORU</td>
<td>ARIS-assisted G2A P2P downlink system</td>
</tr>
<tr>
<td>[101]</td>
<td>UAV’s trajectory, power allocation, wireless charging operation</td>
<td>Throughput</td>
<td>multi-agent DRL, MDP</td>
<td>ARIS-assisted G2A MU downlink system</td>
</tr>
<tr>
<td>[103]</td>
<td>UAV’s trajectory, user scheduling</td>
<td>The total number of served devices</td>
<td>MDP, PPO</td>
<td>TRIS-assisted G2A MU SISO uplink system</td>
</tr>
<tr>
<td>[108]</td>
<td>UAV trajectory and TRIS’s phase shifts, power allocation</td>
<td>Sum rate</td>
<td>DRL</td>
<td>TRIS-assisted A2O uplink system with the joint existence of a GU and an aerial user</td>
</tr>
<tr>
<td>[109]</td>
<td>UAV’s trajectory, TRIS’s phase shifts</td>
<td>Energy consumption</td>
<td>D-DQN, MDP</td>
<td>TRIS-assisted A2G MU downlink system</td>
</tr>
<tr>
<td>[128]</td>
<td>ARIS’s trajectory</td>
<td>SNR</td>
<td>Q-learning, MDP</td>
<td>ARIS-assisted G2G downlink system</td>
</tr>
<tr>
<td>[139]</td>
<td>ARIS’s trajectory and phase shifts</td>
<td>Achievable rate</td>
<td>DQN, MDP</td>
<td>ARIS-assisted G2G downlink anti-jamming system</td>
</tr>
<tr>
<td>[142]</td>
<td>TRIS’s phase shifts, power allocation</td>
<td>BE</td>
<td>DRL</td>
<td>TRIS-assisted A2G multi-cluster communication system with multiple UAV transmitters</td>
</tr>
<tr>
<td>[151]</td>
<td>ARIS’s deployment</td>
<td>Secrecy rate</td>
<td>Q-learning, MDP</td>
<td>ARIS-assisted G2G secrecy communication system</td>
</tr>
</tbody>
</table>

at initial stages, we briefly illustrate some potential but significant research directions in RANTNs.

A. Practical Channel Modeling

Channel modeling is vital to enable fundamental and applied research. Theoretical channel modeling on RANTN is beneficial to assist the advanced system design by paving the way for theoretical performance evaluation and optimization by means of simulations. It is therefore of paramount importance to build accurate channel models for NTNs as well as for RIS, as the current models adopted in most existing works are inaccurate. As a matter of fact, most measurement studies on NTNs focused on ideal rural, suburban, and open fields environments and are not suitable to model dense environments characterized by the existence of dense buildings, various weather conditions, streets, trees, and lake water. In addition, it is often that, for the sake of simplicity, they rely on channel models applied to low-frequency bands and rarely considered to model the channel in high-frequency bands, which suffer from severe attenuation and high probability to be blocked. Furthermore, in RANTNs, a Doppler frequency shift is generally experienced due to the mobility of the aerial platforms. While a large Doppler spread can be experienced when different signal paths are associated with largely different Doppler frequencies, their effect is in general ignored. Besides Doppler spread, the mobility of users makes it necessary to evolve current channel models based on static GUs to nonstationary models accounting for the users’ mobility. These nonstationary models are also needed to model channel variations at aerial platforms, which, even when hovering, experience unavoidable and undesired oscillations caused by meteorological phenomena [130]. A future important research direction is to propose accurate channel models to account for wave propagation at high frequencies, Doppler spread, mobility of users, and undesired oscillations.

For the channel modeling on RIS, most of the existing works are based on far-field channel models, which allows one to assume that the distances between an end terminal and all reflecting elements are the same. However, the near-field channel model is more accurate for some communication scenarios. Specifically, when the deployment of RIS is in the vicinity of end terminals or the RIS has large size capturing a large number of incident electromagnetic waves, the distances from the end terminal to all the reflecting elements are different. In addition to near-field communications, accurate models should account for other important effects, such as reflection.
These unsolved problems further complicate the channel modeling for RANTNs since there is a necessity to investigate new aerial models that consider RIS capabilities together with aerial platforms properties. Specifically, the nonstationary channels with excessive spatial and temporal variations are general in NTNs systems, which are caused by the rapid dynamic movements, rotations, and aerial shadowing of aerial platforms. The RIS adds complexity to defining appropriate channel models because of its passive and reflective behavior and the near-field propagation that needs to be considered. The above two main components together make the channel modeling sophisticated and challenging, whose unique characteristics are still unknown. Recently, an increasing amount of research attention has been dedicated to propose an accurate RANTN channel model. For instance, the very recent works [163], [164] adopt the elliptic cylinders to model the propagation environment among the HAP/UAV transmitter, TRIS, and the ground mobile receiver, while the work [165] considers a cluster structure to describe the scattering environment between the UAV transmitter and the TRIS, which indicates that the signals from the UAV will first impinge on the cluster before arriving at the TRIS. Moreover, Xiong et al. [166] proposed the TRIS-assisted A2G mmWave channel model by treating the RIS as a virtual cluster and RIS units as virtual scatterers. The proposed UAV channel model can be further characterized by the Rician fading under optimal and discrete reflection phase configurations, whereas by Rayleigh fading under random uniform reflection phase configuration. However, all the channel models are proposed under some impractical assumptions mentioned above and without the support of real experimental data. The realistic channel modeling should be derived from real-world implementations and experiments, which involves accurate data to confirm the existing theoretical results.

B. Channel Estimation

The configuration of RISs in RANTNs requires solving an optimization problem depending on the CSI. Such a step is key to achieving optimal beamforming and control of the radio channel. Since the gain brought by RIS depends on how much accurate is the CSI, channel acquisition in RANTNs is an important research topic that deserves particular attention. It poses several challenges due to the passive nature of the RIS and the mobility of aerial platforms.

1) Channel Estimation With Passive RISs: From the RIS perspective, the passive nature of transceiver chains makes it more difficult to estimate the channel since only the cascaded channels are observed. The number of channel coefficients to be estimated proportionally increases with the number of elements, resulting in huge estimation overhead, enormous energy consumption at transceivers, as well as heavy signal processing burdens and long delay at the RIS’s controller that has limited computational capabilities. Most channel estimation methods are no longer applicable for RIS-assisted full-duplex communications. For instance, the assumption of channel reciprocity used in typical TDD does not hold. As a matter of fact, several experimental results showed that the reflection coefficient at RIS is sensitive to the arrival angle of the impinging wave. As this arrival angle in the uplink differs from that in the downlink, the reciprocity of channels could not be met. The frequency-division duplex-based estimation protocol is also unrealistic as it involves high feedback overhead due to the large-dimensional channel matrices. On top of that, the mutual coupling between RIS elements adds an additional layer of complexity to an already complicated problem.

Several complexity-reduced methods are proposed by grouping the adjacent RIS elements with the same configuration but come at the cost of a performance loss. Another approach is to alter the passive nature of the RIS by incorporating few low-power active sensors, which enables sensing and channel estimation directly at the RIS. However, the use of active elements is against the attractive characteristic of RISs that is to control the channel using passive elements. Moreover, even with active elements, a control loop is still required to jointly adjust the RIS configuration and the beamforming at the transmitter/receiver.

2) Impact of Mobility on Channel Estimation: In high-frequency communications, the propagation channel can experience an important variation due to only a few millimeters of movement. This is the reason why in RANTNs, real-time channel estimation is more challenging due to mobile aerial platforms and becomes even more difficult when ARISs instead of TRIS are considered [132], [167]. Achieving real-time reconfigurability in such various mobility conditions is still an open question that requires the development of appropriate tools. In this respect, machine learning techniques, such as supervised learning and RL [41], [42], might be a promising way to facilitate the estimation process [64].

3) Potential Research Directions: Beyond the channel estimation accuracy of a given method, it is important that the energy cost and the long delay it entails are considered. Overall, acquiring the channel with low latency, affordable energy consumption, low signaling overhead, and reduced computational complexity in dynamic wireless environments is still an open problem due to the mobility of aerial platforms the passive reflection of RIS, and the massive number of reflecting elements. A potential method to reduce the latency of channel estimation procedures is to utilize the slowly varying long-term CSI depending on both angular and location information, which can be estimated by the direction of arrival/departure estimation. It dramatically reduces the burden on channel estimation and control signaling transmission because the phase shifts...
designed based on slowly varying, which yields lower computation complexity and lower requirement for the control links. However, there are still not enough works in the system design based on long-term CSI.

C. Tracking

Tracking mobile users/aerial platforming is challenging in future wireless communication systems, which are required to sense their surrounding environment and perform accurate localization. RISs thus are adopted to better track the movement of devices, especially when the direct links between the end devices are blocked [168], [169]. However, since RISs cannot send pilot signals, the tracking problem becomes much more complicated than the conventional one without the assistance of RISs. Moreover, the utilization of higher frequency bands to support applications also greatly reduces the accuracy of sensing and localization. The reason is that the number of propagation paths in high frequency is reduced, which is mainly due to large penetration losses, high values of path loss, and low scattering. Adding to that, the conventional sensing and localization models are based on far-field models and as such may not apply to the case of near-field models. A potential research direction is to develop near-field sensing and localization models that exploit the information in the wavefront curvature [170], [171].

D. Hardware Limitations

Existing analysis algorithms for RANTNs have been developed based on the assumption of perfect hardware. Yet, various types of hardware limitations existing in practical RANTNs must be accounted for.

Most of the existing RANTN works are based on the unrealistic assumption that RIS can perfectly manipulate the impinging electromagnetic waves and reflect them with the optimized ideal phase shifts. More specifically, a widely considered assumption is to assume that the RIS's phase shifts are represented by an infinite bit resolution, allowing them to take any continuous value. Obviously, such an assumption is unrealistic as it could not be supported by the controller or the control link in practice. This has motivated researchers to consider designs in which the phase shift is constrained in a set of discrete values [172]–[174]. In this case, one possible approach is to perform the exhaustive search [175]. However, this involves prohibitively high computational complexity, especially for IRSs with high-resolution phase shifts and a large number of reflecting elements. A widely adopted design method is to relax the discrete phase shift constraints and quantize the relaxed continuous solution to the nearest values in the discrete constraint set. Although this approach is generally able to reduce the computation complexity, it can lead to arbitrarily severe performance loss. More efficient discrete beamforming algorithms have been proposed for particular RIS-assisted systems [172], [176], [177], but none of them consider the communication scenarios with moving aerial platforms. It is thus of interest to study the impact of the discrete RIS on the RANTNs and propose a corresponding efficient design scheme for particular systems integrating both discrete RIS and aerial platforms. Actually, not only the continuous phase shift but also the high-resolution reflecting elements are difficult to implement in practice since the required cost and complex hardware design increases with higher phase resolution. In this respect, the 1-bit resolution with only two possible phases is appealing as it is more suitable for a high number of reflecting elements [178]. The performance loss caused by using practical low-resolution RISs is also an interesting question that deserves investigation.

Another aspect that is worth studying is to consider the RIS reflection model capturing the fact that part of the incident wave reflected by each reflecting element is consumed at the resistance of the reflecting circuit, which induces a loss that depends on the phase shifts of the reflecting element [179]–[181]. As a consequence, the amplitude of the reflection coefficients is less than or equal to one depending on the phase shifts of the incident wave. Other additional hardware constraints involving RIS include but are not limited to the consideration of discrete amplitudes studied in [182] and the phase noise considered in [183] and [184]. It is thus necessary to build practical models accounting for properties of the used physical materials, the manufacturing processes, the RIS's reliability and configuration capabilities for different communication frequencies, and diverse numbers and sizes of RIS units.

As far as aerial platforms are considered, designs based on optimizing their trajectories and positions may not perform as expected [185], due to the finite precision of electronic circuits and imperfect manufacturing of mechanical components. The imperfect transmitting/receiving hardware modules (such as power amplifier nonlinearity, nonlinear phase noise, frequency and phase offsets, and in-phase and quadrature imbalance) and quantization noise jointly degrade communication quality. Hence, it is still unclear whether the favorable performance can be guaranteed considering these hardware limitations.

E. Backhaul Control

Because of the limited signal processing and computing capabilities of the RIS, the configuration of RIS is performed through a smart controller that communicates with a computing node to achieve real-time adaptive beamforming. However, such an approach requires a fully synchronized and reliable control link between the computing node and the RIS. A reliable control can be achieved in static propagation environments, which is not the case of RANTNs. For instance, the control link between the computing node and the ARIS is likely to experience time-varying channel fading conditions with shadowing, thus affecting the process of uploading the phase shift modifications in real time. In addition, the number of elements in the RIS ranging from a few to hundreds or more
generates a tremendous amount of signaling overhead. More importantly, the energy consumption at powerful controller degrades the system performance, especially for ARISs, which are energy-limited devices. In advanced system involving multiple end nodes and aerial platforms, the centralized controller is required to collect all the complex-valued channel matrices and compute all passive beamforming, before sending them back to the corresponding nodes over the network. This results in a heavy feedback overhead and a prohibitively high computational complexity that impedes the RIS configuration.

Therefore, new solutions are needed to provide stable control links with low latency and limit the control signaling and processing overhead without compromising the performance. The control protocol via a separate wireless link or via dedicated time slots is a possible option that has not been investigated in the literature. Other options include using distributed algorithms that possess appealing advantages over centralized algorithms, such as low information exchange overhead, reduced computational complexity, and increased scalability. These algorithms can be enabled by using, for instance, UAV swarms that can offer distributed computing and communications capabilities.

**F Efficient Design and Optimization**

1) **Trajectory Optimization**: The development of robust and efficient algorithms for enhancing the RANTNs is still a practically challenging task, because of the mobility of aerial platforms, in particular the controllable highly mobile UAVs. Most important topics in NTNs revolve around the dynamic 3-D location adjustment to provide an additional DoF for improving the communication performance. They involve solving a trajectory optimization problem under two types of constraints, namely, constraints on the aerial platforms' flight such as minimum/maximum flying altitude/speed, initial/final locations, maximum acceleration, obstacle and collision avoidance, and no-fly zone, and communication-related constraints including transmit power, serving time, and frequency resource limitations. Despite the important research works on trajectory optimization of aerial platforms, most contributions are based on that aerial platforms are deployed in free space, ignoring the scenario that low-altitude aerial platforms such as UAVs can be deployed in closed and rooftop environments. The impact of the propagation environment, including the speed and direction of wind, the propulsive efficiency, and weather conditions, is always overlooked. In general, the trajectory optimization problem is usually nonconvex, and thus, accounting for all these constraints would make it even more difficult to solve.

2) **RIS Configuration With Moving Aerial Platforms**: In RANTNs, the configuration of RISs poses several challenges. From an optimization perspective, the aerial platform's trajectory is coupled with that of the RIS, leading to a more involved problem. From an implementation perspective, the jittering of aerial platforms and the uncertainty of user location increases channel fluctuations [185], while the movement of aerial platforms introduces signal alignment. All this can lead to the RIS passive beamforming becoming ineffective. It is noticeable that all the performance improvement in RANTNs is obtained at the price of solving a complex optimization problem with high hardware cost and design complexity. Hence, how to strike an optimal performance–complexity/cost tradeoff remains an open problem. To address this issue, the RIS grouping scheme was proposed [54], [186], [187], where adjacent RIS reflecting elements with high channel correlation are grouped into a subsurface and are assumed to have the same reflection elements. Moreover, several complexity-reduced algorithms were developed under the unrealistic trajectory discretization assumption, where the UAV is considered to be at a fixed sample location during each interval. However, in reality, the UAV is continually moving.

3) **Distributed Devices**: To enable data collection applications, advanced networks are supported by a large number of ground devices, aerial platforms, and RISs distributed in a certain area. This thus induces a higher design complexity because the computational resources are distributed everywhere across heterogeneous aerial and ground nodes with distinct communication and computation capabilities. How to match the time- and spatial-varying communication/computation/prediction demands with distributed communication/computation/data supplies in such highly dynamic 3-D networks is a challenging task. The interaerial platform cooperation among aerial platform swarms constitutes a promising solution [188], [189] but has not been studied in RANTNs. It is based on dynamically selecting a cluster head, whereas other aerial platforms and devices first collect data in a small area and then transmit them to the cluster head for further processing. Such an architecture has powerful computing capabilities, but it also has the drawback of requiring a transmission delay to coordinate between clusters, as well as a possibly heavy loading on the cluster head, especially when the number of devices becomes large.

4) **Integration of RISs: Challenges and Possible Solutions**: As far as RISs are concerned, the large-scale deployment of RISs may bring inter-RIS interference to the network, degrading its performance. Specifically, transmissions coming from one RIS can leak to others, which causes strong inter-RIS interference. Such impairment could be mitigated by letting RISs with high interference levels work in a cooperative mode rather than operating independently. However, the coordination of multiple RISs can trigger an explosion in signaling and processing overhead and potentially results in high communication latency. Apart from interference, safeguarding wireless communication with such massive aerial platforms, RISs, and ground devices is another vital issue that is worth studying. In this context, Shang et al. [131] presented several applications of swarm-enabled ARIS, including
assisting G2G/A2G/G2A networks to tackle the physical layer security problem, and mentioned the key challenges in the ARIS swarm network, such as beamforming design, channel estimation, deployment, and trajectory design. Although some potential solutions have been suggested to rise to these challenges, their implementation in practical scenarios has not been thoroughly assessed. More research is required to examine the most effective network typology and communication protocol that allows for both efficient sensing and high data rate communication.

More importantly, the potential of RIS needs to be further dug. For instance, the RIS’s full-duplex characteristics should be further incorporated with the existing networks by optimizing the fraction overlap between the uplink and downlink, which yields new design problems in terms of power allocation and beamforming. In addition, most existing studies focus on using RIS to assist communication between devices. With the emergence of the new paradigm of symbiotic communication, RIS can be envisioned to have the additional role of transmitting information. In such settings, RISs would be able to simultaneously enhance the communication quality of the primary link and transmit their own information that may include control signals to acknowledge their current status, and some environmental parameters. The most simplest way to achieve this is to equip each RIS with a dedicated transmitter, at the cost of extra power consumption. The other appealing approach to achieve this is to modulate the RIS information onto the reflected signals to establish passive information transfer.

Since different types of aerial platforms, including satellites, HAPs, and UAVs, have their own advantages and disadvantages such as cost, latency, persistence, and mobility, a hot research topic is the design and optimization of integrated networks consisting of various types of aerial platforms and various RISs. For the integrated network, new research topics emerge that concern the design of efficient and fault-tolerant network mechanisms, reliable transmission protocols, seamless information exchange among different layers, and dynamic network operation control mechanisms, all of which aim to provide continuous service, expanded coverage, rapid mission response, and reliable transmission.

G. Spectral Sharing

A critical issue in RANTNs is the limited spectrum available for simultaneous G2G, G2A, and A2A communications. Unlike terrestrial communication systems where GBs are connected to a data hub via high-speed fixed-line backhaul links, e.g., optical fibers, the implementation of RANTNs usually relies on dedicated wireless communication channels. In particular, RANTNs have a stringent demand for system resources due to the required support of high data rate backhauling and exchange of time-critical control signals of aerial platforms. Furthermore, multiple RISs and aerial platforms tend to be deployed simultaneously in next-generation networks, which puts a significant burden on the need for available spectrum.

In this context, it is generally believed that traditional multiple access schemes based on orthogonal spectrum partition cannot even support moderate numbers of UAVs and GU due to the rapid exhaustion of available resources. Actually, the scarce wireless spectrum is already congested by existing communication systems. Thus, efficient multiple access schemes based on nonorthogonal spectrum utilizations have to be developed for enabling RANTN systems.

As reviewed in Section VII-G, NOMA is a promising technology for multiplexing in RANTNs. However, since RISs are dynamically optimized based on the instantaneous realizations of combined channels, the users’ effective channel gains become coupled of the RIS coefficients. This results in a highly complex problem to determine the optimal user ordering in an MU NOMA scenario, especially when the number of users and the number of reflecting elements increase, posing major challenges for the practical integration of NOMA into RANTNs. Therefore, the environment-dependent NOMA scheme with limited gain facilitates researchers to introduce rate-splitting multiple access (RSM), which has better spectral efficiency for an integrated network. However, this scheme is just utilized in the context of integrated spatial–terrestrial networks, where RSM can be distributively or centrally controlled to be employed horizontally at one of the layers or vertically at each layer. Moreover, RISs can be enablers of CR systems to save spectral resources, where RISs are used to ensure quality satisfying communication in targeted primary or secondary networks. Nevertheless, the investigation of the mentioned and other efficient spectrum sharing schemes in RANTNs is missing in the literature and needs the researchers’ attention.

VIII. CONCLUSION

RIS provides efficient and cost-saving solutions to the existing challenges in NTNs, such as endurance, blockage, and information leakage. This article has presented a comprehensive survey on the RANTNs from perspectives of framework, methodology, design, and applications. First, we have introduced the key properties of RIS in NTNs and the two kinds of RIS utilized in NTNs, which followed by the overview of the structure of RANTNs, including A2G/G2A communications, G2G communications, and A2A communications. Then, we have focused on the integration of other promising technologies to further enhance RANTNs. Then, we have provided detailed reviews on the insights brought by the performance analysis and optimization of RANTNs for different performance indices, such as SNR, achievable data rate, energy and time consumption, and EE. Afterward, we have concluded the methodology and tools utilized in the existing literature, which provides readers with efficient methods to pursuing solutions for various optimization frameworks in RANTNs. Finally, we have outlined key challenges and future research directions.
Meaningful insights are revealed in the literature and summarized in this survey. The joint design of the deployment of ARIS/aerial platforms and RIS configurations is the key factor in reaping the full benefits of RANTNs. The AO framework and ML algorithms are widely used in the current literature to obtain feasible solutions. The well-designed RANTNs tend to be a key component of future networks, which can be applied in smart city construction, worldwide coverage realization, emergency network, and so on. However, the studies on RANTNs have not yet been supported by realistic data due to the lack of practical experiments, especially on the practical channel modeling. The majority of recent works ignored the practical limitations such as the RIS’s finite phase resolution, signal misalignment, imperfect CSI, hardware impairments, and limited control resources. There is still limited literature on advanced networks with multiple users, RISs, and aerial platforms, as well as the combination with other promising technologies, such as communications in high-frequency bands, spectral sharing, and wireless power supply. This work opens a significantly larger space for the RANTNs to explore than the current research scope.

REFERENCES


Authorized licensed use limited to: KAUST. Downloaded on May 12, 2022 at 05:55:12 UTC from IEEE Xplore. Restrictions apply.


Jia Ye (Student Member, IEEE) was born in Chongqing, China. She received the B.Sc. degree in communication engineering from Southwest University, Chongqing, in 2018, and the M.S. degree from the King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, in 2020, where she is currently working toward the Ph.D. degree.

Her main research interests include the performance analysis and modeling of wireless/wireless communication systems.

Jingping Qiao (Member, IEEE) received the B.E. degree from the School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, China, in 2012, and the Ph.D. degree from Shandong University, Jinan, China, in 2018.

Since September 2019, he has been a Visiting Scholar with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. He is currently a Lecturer with the School of Information Science and Engineering, Shandong Normal University, Jinan. His research interests include physical layer security, cooperative (relay) systems, terahertz communications, intelligent reflecting surface (IRS), and signal processing for wireless communications.
Abla Kammoun (Member, IEEE) was born in Sfax, Tunisia. She received the Engineering degree in signal and systems from the Tunisia Polytechnic School, La Marsa, Tunisia, in September 2005, and the master’s and Ph.D. degrees in digital communications from Télécom Paris Tech (formerly, École Nationale Supérieure des Télécommunications), Paris, France, in September 2006 and April 2010, respectively.

From 2010 to 2012, she was a Postdoctoral Researcher with the TSI Department, Télécom Paris Tech. She was with the Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif-sur-Yvette, France, until 2013. She is currently a Research Scientist with the King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. Her research interests include performance analysis of wireless communication systems, random matrix theory, and statistical signal processing.

Mohamed-Slim Alouini (Fellow, IEEE) was born in Tunis, Tunisia. He received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998.

He was a Faculty Member with the University of Minnesota, Minneapolis, MN, USA, and Texas A&M University at Qatar, Doha, Qatar, before joining the King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, as a Professor of electrical engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.