

# Reconfigurable Intelligent Surfaces: Principles and Opportunities

Yuanwei Liu<sup>ID</sup>, Senior Member, IEEE, Xiao Liu<sup>ID</sup>, Graduate Student Member, IEEE,  
 Xidong Mu<sup>ID</sup>, Graduate Student Member, IEEE, Tianwei Hou<sup>ID</sup>, Jiaqi Xu<sup>ID</sup>, Student Member, IEEE,  
 Marco Di Renzo<sup>ID</sup>, Fellow, IEEE, and Naofal Al-Dhahir, Fellow, IEEE

**Abstract**—Reconfigurable intelligent surfaces (RISs), also known as intelligent reflecting surfaces (IRSs), or large intelligent surfaces (LISs),<sup>1</sup> have received significant attention for their potential to enhance the capacity and coverage of wireless networks by smartly reconfiguring the wireless propagation environment. Therefore, RISs are considered a promising technology for the sixth-generation (6G) of communication networks. In this context, we provide a comprehensive overview of the state-of-the-art on RISs, with focus on their operating principles, performance evaluation, beamforming design and resource management, applications of machine learning to RIS-enhanced wireless networks, as well as the integration of RISs with other emerging technologies. We describe the basic principles of RISs both from physics and communications perspectives, based on which we present performance evaluation of multi-antenna assisted RIS systems. In addition, we systematically survey existing designs for RIS-enhanced wireless networks encompassing performance analysis, information theory, and performance optimization perspectives. Furthermore, we survey existing research contributions that apply machine learning for tackling challenges in dynamic scenarios, such as random fluctuations of wireless channels and user mobility in RIS-enhanced wireless networks. Last but not least, we identify major issues and research opportunities associated with the integration of RISs and other emerging technologies for applications to next-generation networks.

**Index Terms**—6G, intelligent reflecting surfaces (IRSs), large intelligent surfaces (LISs), machine learning, performance

Manuscript received November 29, 2020; revised March 5, 2021; accepted April 15, 2021. Date of publication May 5, 2021; date of current version August 23, 2021. The work of Marco Di Renzo was supported in part by the European Commission through the H2020 ARIADNE Project under Grant 871464 and through the H2020 RISE-6G Project under Grant 101017011. (*Corresponding author: Yuanwei Liu.*)

Yuanwei Liu, Xiao Liu, and Jiaqi Xu are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: yuanwei.liu@qmul.ac.uk; x.liu@qmul.ac.uk; jiaqi.xu@qmul.ac.uk).

Xidong Mu is with the School of Artificial Intelligence and the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: muxidong@bupt.edu.cn).

Tianwei Hou is with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China (e-mail: 16111019@bjtu.edu.cn).

Marco Di Renzo is with the Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91192 Gif-sur-Yvette, France (e-mail: marco.di-renzo@universite-paris-saclay.fr).

Naofal Al-Dhahir is with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: aldhahir@utdallas.edu).

Digital Object Identifier 10.1109/COMST.2021.3077737

<sup>1</sup>Without loss of generality, we use the name of RIS in the remainder of this paper.

optimization, reconfigurable intelligent surfaces (RISs), wireless networks.

## I. INTRODUCTION

THE UNPRECEDENTED demands for high quality and ubiquitous wireless services impose enormous challenges to existing cellular networks. Applications like rate-centric enhanced mobile broadband (eMBB), ultra-reliable, low latency communications (URLLC), and massive machine-type communications (mMTC) services are the targets for designing the fifth-generation (5G) of communication systems. However, the goals of the sixth-generation (6G) of wireless communication systems are expected to be transformative and revolutionary encompassing applications like data-driven, instantaneous, ultra-massive, and ubiquitous wireless connectivity, as well as connected intelligence [1], [2]. Therefore, new transmission technologies are needed in order to support these new applications and services. Reconfigurable intelligent surfaces (RISs), also called intelligent reflecting surfaces (IRSs) [3], [4] or large intelligent surfaces (LISs) [5], [6], comprise an array of reflecting elements for reconfiguring the incident signals. Owing to their capability of proactively modifying the wireless communication environment, RISs have become a focal point of research in wireless communications to mitigate a wide range of challenges encountered in diverse wireless networks [7], [8]. The advantages of RISs are listed as follows.

- **Easy to deploy:** RISs are nearly-passive devices, made of electromagnetic (EM) material. As illustrated in Fig. 1, RISs can be deployed on several structures, including but not limited to building facades, indoor walls [9], aerial platforms, roadside billboards, highway polls, vehicle windows, as well as pedestrians' clothes due to their low-cost.
- **Spectral efficiency enhancement:** RISs are capable of reconfiguring the wireless propagation environment by compensating for the power loss over long distances. Virtual line-of-sight (LoS) links between base stations (BSs) and mobile users can be formed via passively reflecting the impinging radio signals. The throughput enhancement becomes significant when the LoS link between BSs and users is blocked by obstacles, e.g., high-rise buildings. Due to the intelligent deployment and

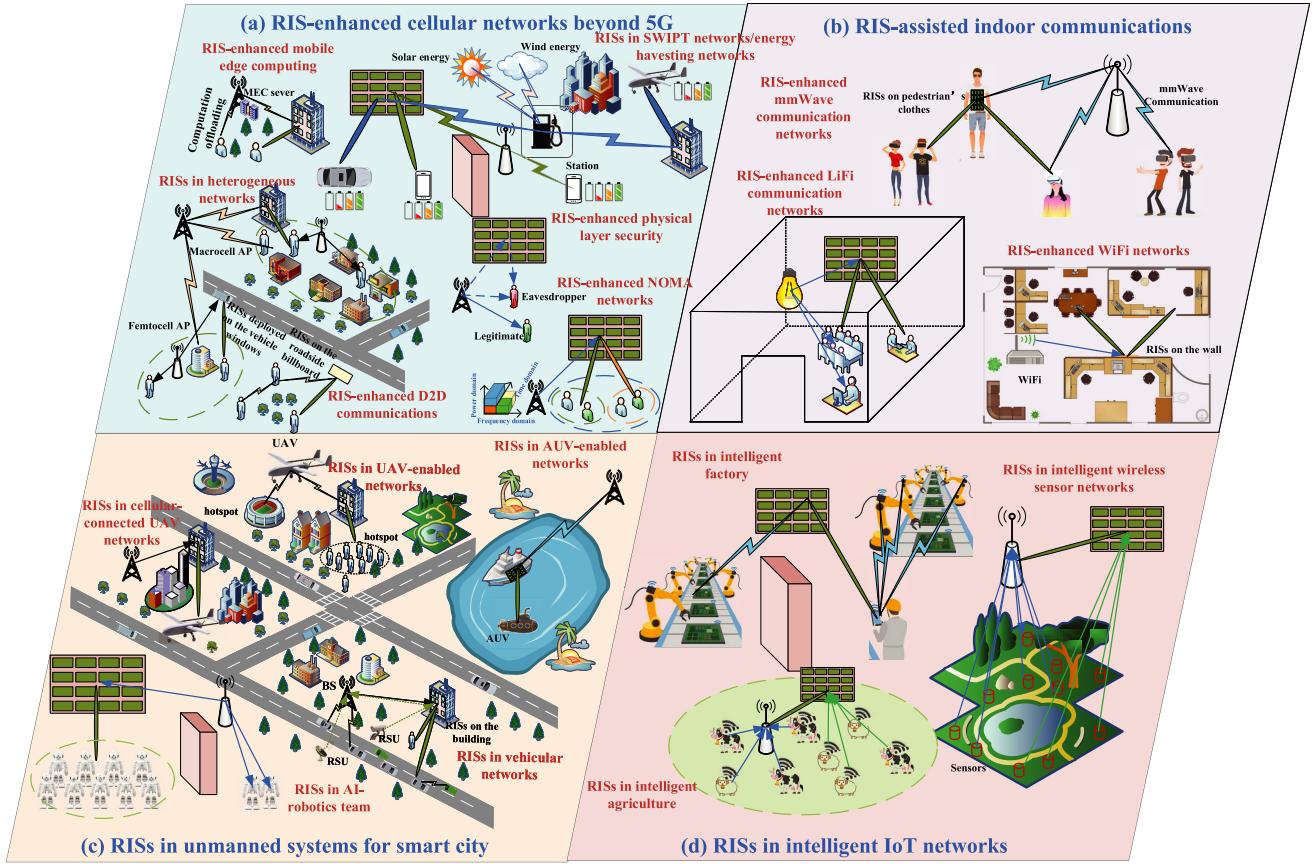


Fig. 1. RISs in wireless communication networks.

design of RISs, a software-defined wireless environment may be constructed, which, in turn, provides potential enhancements of the received signal-to-interference-plus-noise ratio (SINR).

- **Environment friendly:** In contrast to conventional relaying systems, e.g., amplify-and-forward (AF) and decode-and-forward (DF) [10], RISs are capable of shaping the incoming signal by controlling the phase shift of each reflecting element instead of employing a power amplifier [11], [12]. Thus, deploying RISs is more energy-efficient and environment friendly than conventional AF and DF systems.
- **Compatibility:** RISs support full-duplex (FD) and full-band transmission due to the fact that they only reflect the EM waves. Additionally, RIS-enhanced wireless networks are compatible with the standards and hardware of existing wireless networks [13].

Due to the aforementioned attractive characteristics, RISs are recognized as an effective solution for mitigating a wide range of challenges in commercial and civilian applications. There have been many recent studies on RISs and their contributions focus on several application scenarios under different assumptions. As a result, the system models proposed by these research contributions tend to be different. Thus, there is an urgent need to categorize the existing research contributions, which is one of the main goals of this paper.

Fig. 1 illustrates the applications of RISs in diverse wireless communication networks. In Fig. 1(a), RIS-enhanced cellular networks are illustrated, where RISs are deployed for bypassing the obstacles between BSs and users. Thus, the quality of service (QoS) in heterogeneous networks and the latency performance in mobile edge computing (MEC) networks are improved [14], [15]. On the other hand, RISs can act as a signal reflection hub to support massive connectivity via interference mitigation in device-to-device (D2D) communication networks [16], or RISs can cancel undesired signals by smartly designing the passive beamforming in the context of physical layer security (PLS) [17]. Additionally, RISs can be deployed to strengthen the received signal power of cell-edge users and mitigating the interference from neighbor cells [18], and the power loss over long distances can be compensated in simultaneous wireless information and power transfer (SWIPT) networks [19]–[22]. In Fig. 1(b), RIS-assisted indoor communications are illustrated, where RISs can be deployed on walls for enhancing the QoS in some rate-hungry indoor scenarios, such as virtual reality (VR) applications. Additionally, in order to guarantee no blind spots in the coverage area of some block-sensitive scenarios, such as visible light communications [23] and wireless fidelity (WiFi) networks, a concatenated virtual RIS-aided LoS link between the access points (APs) and the users can be formed with the aid of RISs, which indicates that both the propagation links between the APs and the RISs, as well as between the RISs and

TABLE I

COMPARISON OF THIS WORK WITH AVAILABLE MAGAZINES/SURVEYS/TUTORIALS. HERE, “DL” REFERS TO “DEEP LEARNING”, “RL” REFERS TO “REINFORCEMENT LEARNING”, “AV” REFERS TO “AUTONOMOUS VEHICLE”, AND “CV” REFERS TO “CONNECTED VEHICLE”

—	Classifications	Key Contents	Wu <i>et al.</i> [3]	Liang <i>et al.</i> [6]	Di Renzo <i>et al.</i> [7]	Huang <i>et al.</i> [10]	Liaskos <i>et al.</i> [28]	Gacanin <i>et al.</i> [29]	Di Renzo <i>et al.</i> [30]	Basar <i>et al.</i> [31]	Gong <i>et al.</i> [32] [33]	Wu <i>et al.</i>	This work
Physics-based	Electromagnetics	Distinguishing ray-optics & wave-optics perspective	✓						✓	✓			✓
		Macroscopic description of metasurface-based RIS		✓					✓	✓			
		Surface equivalence theorems		✓					✓				✓
		Distinguishing far-field & near-field			✓								✓
	RIS Modeling	Formulating the reflection coefficient	✓	✓						✓	✓	✓	✓
		Power conservation principle			✓								
Communication-based	Performance evaluation	RIS control mechanism	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Typical tunable functions		✓				✓	✓	✓	✓		✓
		RIS hardware and prototypes		✓	✓	✓	✓		✓	✓	✓	✓	✓
		RIS synthesis methods		✓	✓	✓	✓		✓				✓
	RIS-aided communication design	Path loss models		✓	✓					✓	✓	✓	✓
		Stochastic geometry			✓				✓	✓	✓		✓
		Discussion on stochastic analysis tools							✓				✓
		Discussion on small-scale fading channels											✓
		Information-theoretic capacity limits											✓
		Passive beamforming optimization	✓	✓	✓	✓				✓	✓	✓	✓
	ML-empowered RIS	Resource management				✓							✓
		Comparison of employed mathematical approaches											✓
		Channel estimation	✓	✓	✓	✓				✓	✓	✓	✓
		Deployment design	✓	✓						✓		✓	✓
		Modulation				✓			✓		✓		
		Localization and sensing			✓				✓				
	Compatibility	Architecture							✓				✓
		DL				✓	✓		✓	✓	✓		✓
		RL							✓				✓
		Comparison of RL-based algorithms											✓
		Supervised, unsupervised, and federated learning						✓					✓
		NOMA-RIS			✓								✓
		PLS-RIS	✓	✓		✓				✓	✓	✓	✓
		SWIPT-RIS	✓	✓		✓				✓	✓	✓	✓
		UAV-RIS											✓
		AV/CV-RIS											✓
		MEC-RIS			✓						✓	✓	
		mmWave-RIS			✓					✓	✓	✓	✓

the users can be in LoS. In Fig. 1(c), RIS-enhanced unmanned systems are illustrated. RISs can be leveraged for enhancing the performance of unmanned aerial vehicle (UAV) enabled wireless networks [24], cellular-connected UAV networks [25], autonomous vehicular networks, autonomous underwater vehicle (AUV) networks, and intelligent robotic networks by fully reaping the aforementioned RIS benefits. For instance, in RIS-enhanced UAV-aided wireless networks, one can adjust the phase shifts of RISs instead of controlling the movement of the UAVs in order to form concatenated virtual LoS links between the UAVs and the users. Therefore, the UAVs can maintain the hovering status only when the concatenated virtual LoS links cannot be formed even with the aid of RISs, which reduces the movement manipulations and the energy consumption of UAVs. In Fig. 1(d), RIS-enhanced Internet of Things (IoT) networks are illustrated, where RISs are exploited for assisting intelligent wireless sensor networks [26], intelligent agriculture, and intelligent factory [27].

There are some short magazine papers [3], [10], [28], [29], surveys and tutorials [6], [7], [30]–[33] in the literature that introduced RISs and their variants, but the focus of these papers is different from our work. More specifically, Wu and Zhang [3] provided an overview of the applications of RISs as reflectors in wireless communications, and identified some challenges and future research opportunities for implementing RIS-assisted wireless networks. Liang *et al.* [6] presented an overview of the reflective radio technology with a particular focus on the large intelligent surface/antennas. In [7], Di Renzo *et al.* provided a comprehensive overview of employing RISs for realizing smart radio environments in wireless networks, where an electromagnetic-based communication-theoretic framework for analyzing and optimizing metamaterial-based RISs is presented and a survey of recent research contributions on

RISs is given. Huang *et al.* [10] introduced the concept of holographic multiple-input and multiple-output (MIMO) surfaces (HMIMOS), and discussed both active and passive RISs, encompassing the hardware architectures, operation modes, and applications in communications. In [28], Liaskos *et al.* presented one kind of RIS prototype, namely the HyperSurface tile, for realizing programmable wireless environments. Gacanin and Di Renzo [29] gave an overview of employing artificial intelligence (AI) tools in RIS-assisted radio environments. Di Renzo *et al.* [30] introduced the concept of smart radio environments empowered by RISs, and discussed recent research progresses and future potential challenges. Basar *et al.* [31] reviewed recent research efforts on RIS-empowered wireless networks, identified the differences between RISs and other technologies, and presented future research challenges and opportunities. Gong *et al.* [32] surveyed recent research works on RIS-assisted wireless networks and discussed emerging applications and implementation challenges of RISs. From the perspective of enhancing the communication performance, Wu *et al.* [33] gave a tutorial on design issues in RIS-assisted wireless networks, including passive beamforming optimization, channel estimation, and deployment design.

Although the aforementioned magazines/surveys/tutorials presented either general concepts or specific aspects of RISs (e.g., from a physics-based or a communication-based perspective), the fundamental performance limits of RISs and some potential applications in wireless networks are not covered. The comparison between widely employed mathematical tools for performance evaluations and optimizations in RIS-enhanced wireless networks is also not discussed. Moreover, a detailed framework based on machine learning (ML) tools for designing RIS-enhanced wireless networks is missing, except for a short magazine paper [29]. Motivated by all the

aforementioned considerations, this paper provides a comprehensive discussion of RIS-enhanced wireless network principles, from physics to wireless communications, and discusses research opportunities for exploiting RISs in diverse applications, such as unmanned systems, non-orthogonal multiple access (NOMA), and ML. Table I illustrates the comparison of this treatise with the existing magazines/surveys/tutorials in the context of RISs.

Against the above observations, our main contributions are as follows.

- We overview the fundamental principles that govern the operation of RISs and their interaction with the EM signals. We also survey typical RIS functions and their corresponding principles. Specifically, we focus on patch-array based implementation and compare the ray-optics perspective with the wave-optics perspective.
- We develop performance evaluation techniques for multi-antenna assisted RIS systems. Research contributions are also summarized along with their advantages and limitations.
- We investigate RISs from the information-theoretic perspective, based on which we review the protocols and approaches for jointly designing beamforming and resource allocation schemes with different optimization objectives. Additionally, the major open research problems are discussed.
- We discuss the need of amalgamating ML and RISs. After reviewing the most recent research contributions, we propose a novel framework for optimizing RIS-enhanced intelligent wireless networks, where big data analytic and ML are leveraged for optimizing RIS-enhanced wireless networks.
- We identify major research opportunities associated with the integration of RISs into other emerging technologies and discuss potential solutions.

As illustrated in Fig. 2, this paper is structured as follows. Section II elaborates on the fundamental operating principles of RIS-enhanced wireless networks. Section III focuses on the performance evaluation of multi-antenna RIS-assisted systems and the main advantages of using RISs in wireless networks. In Section IV, the latest research activities on the joint design of beamforming and resource allocation are discussed. The framework of ML-empowered RIS-enhanced intelligent wireless networks is presented in Section V. Finally, Section VI investigates the integration of RISs with other emerging technologies towards the design and optimization of 6G wireless networks.

## II. RIS: FROM PHYSICS TO WIRELESS COMMUNICATIONS

An RIS is a two-dimensional (2D) material structure with programmable macroscopic physical characteristics. The most important characteristic of an RIS is that its EM wave response can be reconfigured. In contrast to conventional wireless communication networks, the channels between the transmitters and the receivers can be controlled in RIS-aided networks. Thus, the strength of the desired received signal can be enhanced at the terminal devices. In this section, we introduce

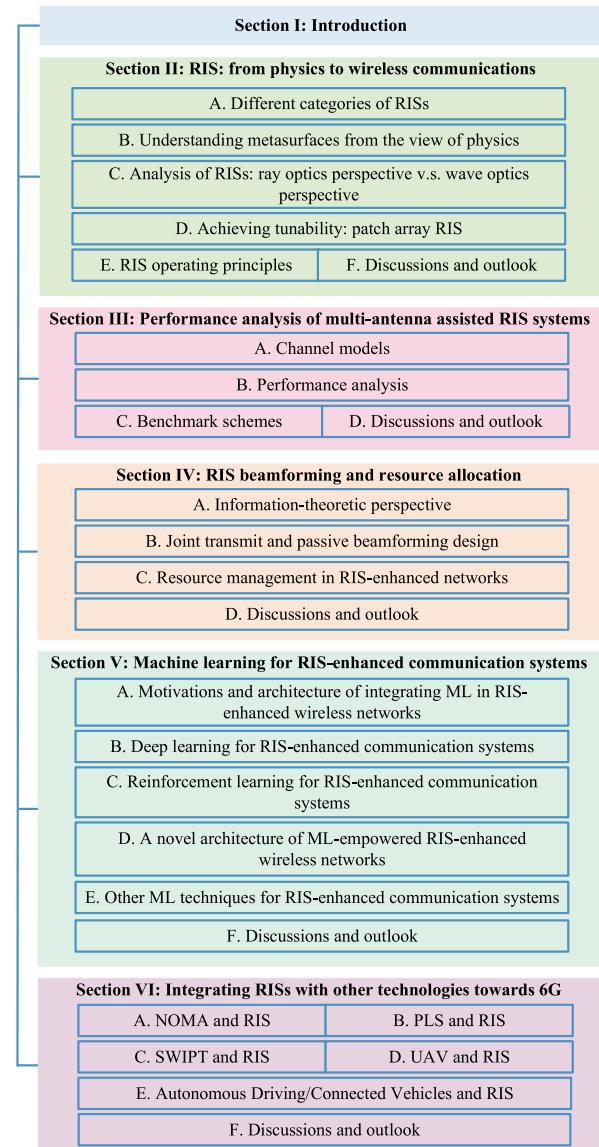


Fig. 2. Organization of the present paper.

the fundamental principles which govern the operation of RISs and their interaction with the EM signals. We also survey typical RIS functions and their corresponding principles.

### A. Different Categories of RISs

Considering their structures, RISs can be realized by using metamaterial or patch-array based technologies. Metamaterial-based RISs are referred to as metasurfaces. Deployed at different locations, RISs can be designed to work as reflecting/refracting surfaces between the BS and the user or waveguide surfaces operating at the BS. Considering the tuning mechanisms, RISs can be reconfigured electrically, mechanically, or thermally. Depending on their energy consumptions, RISs can be categorized as passive-lossy, passive-lossless, or active. The active or passive nature of RISs determines their ultimate performance capabilities. It is worth mentioning, however, that RISs cannot be completely passive because of their inherent property of being configurable. Here, we discuss three

TABLE II  
LIST OF ACRONYMS

AF	Amplify-and-Forward
AO	Alternating Optimization
AUV	Autonomous Underwater Vehicle
BC	Broadcast Channel
BS	Base Station
CSI	Channel State Information
D2D	Device-to-Device
DF	Decode-and-Forward
DL	Deep Learning
EE	Energy Efficiency
EM	Electromagnetic
FD	Full-Duplex
HD	Half-Duplex
IRS	Intelligent Reflecting Surface
IoT	Internet of Things
LIS	Large Intelligent Surface
LoS	Line-of-Sight
MEC	Mobile Edge Computing
MIMO	Multiple-Input and Multiple-Output
MISO	Multiple-Input and Single-Output
ML	Machine Learning
MOS	Mean Opinion Score
NP	Non-deterministic Polynomial-time
NOMA	Non-Orthogonal Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OMA	Orthogonal Multiple Access
PDF	Probability Density Function
PLS	Physical Layer Security
QoS	Quality of Service
RIS	Reconfigurable Intelligent Surface
RL	Reinforcement Learning
SCA	Successive Convex Approximation
SE	Spectral Efficiency
SG	Stochastic Geometry
SIC	Successive Interference Cancellation
SIMO	Single-Input and Multiple-Output
SINR	Signal-to-Interference-plus-Noise Ratio
SISO	Single-Input and Single-Output
SNR	Signal-Noise Ratio
SWIPT	Simultaneous Wireless Information and Power Transfer
UAV	Unmanned Aerial Vehicle
VLC	Visible Light Communication
VR	Virtual Reality
WiFi	Wireless Fidelity
5G	Fifth-Generation
6G	Sixth-Generation

important RIS working operations: waveguide [34], refraction [35], and reflection [36]. With the aid of Love's field equivalence principle [37], the reflected and refracted EM field can be studied by introducing equivalent surface electric and magnetic currents [35]. In the three working conditions, the RIS converts and radiates a wave (either induced by an incident wave or fed by a waveguide) into a desired propagating wave in free space. The surface equivalence principles (SEPs) including Love's field equivalence principle and Huygens' principle are introduced in Section II-B.

1) *Waveguide RIS*: Smith *et al.* [34] presented a theoretical study of waveguide-fed metasurfaces. The elements in the metasurface are modeled as uncoupled magnetic dipoles. The magnitude of each dipole element is proportional to the product of the reference wave and each element's polarizability. By tuning the polarizability, the metasurface antenna can perform beamforming. Each element on the metasurface serves as a micro-antenna. Compared to conventional antenna arrays, the compact waveguide metasurface occupies less space and can transmit towards wider angles.

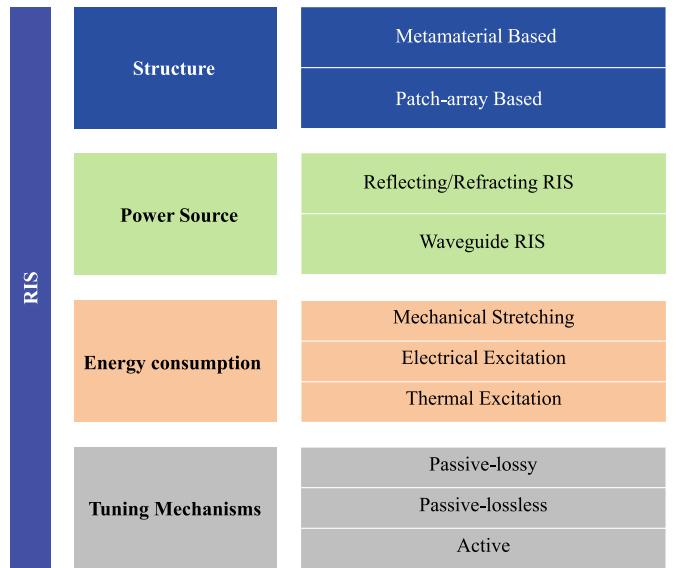


Fig. 3. Different types of metasurfaces.

2) *Refraacting RIS*: Asadchy *et al.* [38] proposed a theoretical design of a perfectly refracting and reflecting metasurfaces. The authors used an equivalent impedance matrix model so that the tangential field components at the two sides of the metasurface are appropriately optimized. Moreover, three possible device realizations are discussed: self-oscillating teleportation metasurfaces, non-local metasurfaces, and metasurfaces formed by only lossless components. The role of omega-type bianisotropy in the design of lossless-component realizations of perfectly refractive surfaces is discussed.

3) *Reflecting RIS*: Dai *et al.* [36] designed a digital coding reflective metasurface. The elements in the metasurface contain varactor diodes with a tunable biasing voltage. By pre-designing several digitized biasing voltage levels, each element can apply discrete phase shifts and achieve beamforming for the reflected wave.

The rest of this section is focused on the operating principles for RISs that operate as reflectors.

### B. Understanding Metasurfaces From the View of Physics

A wireless signal is essentially an EM wave propagating in a three-dimensional space. Attenuation or reduction of the signal strength occurs as the EM wave propagates through the space and interacts with the scattering objects. From basic principles of electromagnetism, the signal power per unit area is proportional to the square of the electric field of the corresponding wave in a given media. As far as reflective and refractive smart surfaces are concerned, this requires the understanding of how the EM waves interact with the surrounding objects. The equivalence principle, especially the SEP, is the building block for studying the EM wave transformations. Some authors also call it Love's field equivalence principle. The principle can be adopted for both external problems (source-free region) and internal problems. Love's field equivalence principle states that the EM field outside or inside a close surface can be uniquely determined by the electric and magnetic currents

TABLE III  
COMPARING DIFFERENT WAVE REPRESENTATION PERSPECTIVES

	<b>Ray-optics</b>	<b>Wave-optics</b>
Wave representation	Geometrical rays	Vector fields
Theoretical foundation	Snell's law	Maxwell's equations
Surface profile	Phase discontinuity	Surface impedance
Requirement of the reflected wave	Co-phase condition	Proper wavefront
Power flow	Not accurate	Accurate

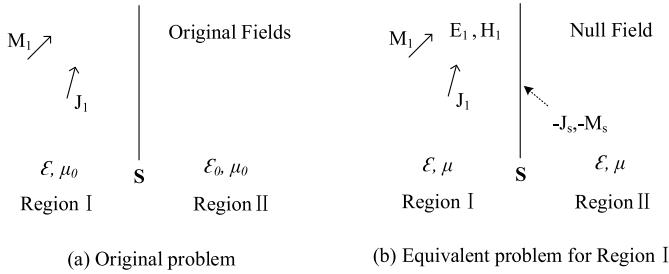


Fig. 4. Love's equivalence internal problem (for region I).

on the surface. As shown in Fig. 4, the equivalent problem for the region I can be reformulated by placing equivalent currents on S that satisfy the boundary condition for each particular case and filling the region II with the same medium of constitutive parameters  $\epsilon$  and  $\mu$ . Thus, the equivalent currents ( $-J_s, -M_s$ ), together with the original source currents ( $J_1, M_1$ ), radiate the correct fields in region I. The equivalent problem for region II can be formulated similarly.

Love's field equivalence principle is the theoretical foundation for analyzing the radiation pattern of RISs. However, the SEP does not specify how to calculate the EM field produced by the surface currents. To obtain the signal strength at an arbitrary field point, the Huygens-Fresnel principle can be employed. The Huygens-Fresnel principle is a method of analysis applied to problems of wave propagation, which states that every point on a wavefront is itself the source of spherical wavelets, and the secondary wavelets emanating from different points mutually interfere. The sum of these spherical wavelets forms the wavefront. Based on the Huygens-Fresnel principle, the EM field scattered by an RIS (in reflection and refraction) can be quantified analytically.

As far as waveguide-based RISs are concerned, the operating principle can be summarized as follows. In [39], the EM wave manipulation of the waveguide metasurface performs the coupling between three-dimensional free space waves and two-dimensional surface waves. As a result, the metasurface can be regarded as a hologram, which carries additional information about its radiated signal propagating in the 3D space. After being excited by the source, this pre-designed information is coupled into the radiated field. Fig. 5 conceptually represents a pre-designed holographic waveguide-based RIS.

### C. Analysis of RISs: Ray Optics Perspective v.s. Wave Optics Perspective

To characterize the interaction between an RIS and the impinging EM waves, one can adopt approximations and tools either from the perspective of ray-optics or wave-optics. These

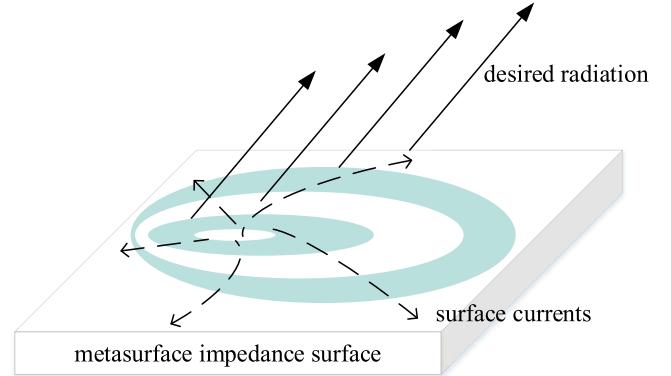


Fig. 5. Conceptual illustration of the holographic impedance smart surface.

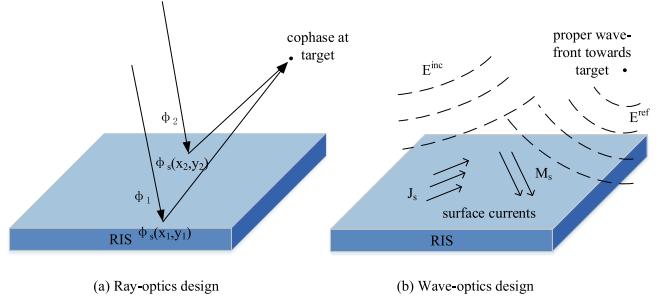


Fig. 6. Comparison between ray-optics and wave-optics perspective.

two perspectives have been used by physicists for a long time. Even though they are based on some approximations, these two methods of analysis are useful in order to obtain important insights into the interaction of light or radio waves with materials. In the RIS literature, both methods of analysis are often employed. However, the assumptions behind their use and their physical interpretations are intrinsically different. To shed light on their differences and similarities, we compare the two methods in this subsection. As shown in Fig. 6(a), from the ray-optics perspective, an EM wave is modeled as a collection of geometrical rays with varying phases. The phase of each ray increases linearly with the optical path length as it traverses through the vacuum or other media. As a result, at each location of the  $i$ -th ray, a phase (denoted by  $\phi_i$ ) can be defined. When the ray interacts with a material, the phenomenon is studied by determining the relationship between the change of the phase and the material refraction index. The desirable reflected wave is obtained if the ensemble of rays obey the proper co-phase condition. The wave-optics perspective is shown in Fig. 6(b), where an EM wave is represented by the corresponding electric field and magnetic field. At each position, each of these two vector fields can be characterized

by a time-varying complex-valued vector, with a direction, an amplitude, and a phase. From the wave-optics perspective, the interaction between the wave and the material can be studied using the equivalent principles discussed in the previous subsection. The points with equal phase values form a series of surfaces in space, which we refer to as wavefronts. As a result, the desired scattered waves (reflected or refracted) are obtained if proper wavefront transformations are performed by the RIS.

*1) Comparing Ray-Optics and Wave-Optics Perspectives:* Table III highlights some of the differences between the two methods of analysis. Compared with wave-optics, the ray-optics perspective is a stronger simplification of the real system. As a result, it is easier to be adopted and can produce a quick prediction about the RIS design. However, ray-optics methods fail when considering the RIS power flow. Wave-optics methods, on the other hand, predict the power flow by using the Poynting vector, which enables us to study the local and overall RIS power consumption. This is an important issue to consider when designing and manufacturing RISs. For example, the authors of [38] and [40] point out that it is impossible to realize lossless plane-wave beam steering with locally passive RISs. One has to adopt the wave-optics perspective to study the power flow of the system. Moreover, the authors of [40] state that, according to their simulation results, one can expect increasingly improved performance if the RISs are designed based on wave-optics approximation in comparison to those designed based on the ray-optics approximation. In conclusion, both perspectives have their advantages and limitations. However, adopting the wave-optics perspective is the most appropriate choice for most cases. A study of the differences, in terms of power flow and reflection coefficient, between ray optics and wave optics methods of analysis is reported in [7].

#### D. Achieving Tunability: Patch Array RIS

The EM characteristics of an RIS, such as the phase discontinuity, can be reconfigured by tuning the surface impedance, through various mechanisms. Apart from electrical voltage, other mechanisms can be applied, including thermal excitation, optical pump, and physical stretching. Among them, electrical control is the most convenient choice, since the electrical voltage is easier to be quantized and controlled by field-programmable gate array (FPGA) chips. The choice of RIS materials include semiconductors [41] and graphene [42].

Regardless of the tuning mechanisms, we focus our attention on patch-array smart surfaces in the following text. The general geometry layout of this type of RIS can be modeled as a periodic (or quasi-periodic in the most general case) collection of unit cells integrated on a substrate. For ease of description, we limit our discussion to RISs that are based on a local design, in which the cells do not interact with each other. A local design usually results in the design of sub-optimal RISs. A comprehensive discussion about local and non-local designs can be found in [7]. To characterize the tunability of the RIS, the method of equivalent lumped-element circuits can be adopted. As shown in Fig. 7, the unit cell is equivalent to a

lumped-element circuit with a load impedance  $Z_l$ . Particularly, the equivalent load impedance can be tuned by changing the bias voltage of the varactor diode. When modeling patch-array RISs in wireless communication systems, we can characterize, under a local design, each of its unit cells through an equivalent reflection coefficient. For example, the reflection coefficient of the  $i$ -th cell can be modeled as follows:

$$r_i = \beta_i \cdot e^{j\phi_i} \quad (1)$$

where  $\beta_i$  and  $\phi_i$  correspond to the amplitude response and the phase response, respectively. As shown in [43], the equivalent reflection coefficient depends on the tuning impedance of the lumped circuit that controls each unit cell, as well as the self and mutual impedances (if mutual coupling cannot be ignored) at the ports of the RIS. In particular, as shown in [43] and [44],  $\beta_i$  and  $\phi_i$  in (1) are usually not completely independent with each other, i.e.,  $\beta_i = f(\phi_i)$ . In the following sections,  $\Phi(\vec{r}_x)$  refers to the phase discontinuity introduced by the RIS as a function of the position on the RIS, and  $\phi_{mn}$  refers to the phase discontinuity of the  $(m, n)$ -th element of a patch-array RIS.

Existing designs of patch-array RISs can apply discrete phase control and, in some cases, amplitude control. Arun and Balakrishnan [45] designed *RFocus*, which is a two-dimensional surface with a rectangular array of passive antennas. The size of each passive unit is  $\lambda/4 \times \lambda/10$  and the EM waves are either reflected or refracted. The authors show that the *RFocus* surface can be manufactured at a low cost, and that it can improve the median signal strength by 9.5 times. Welkie *et al.* [46] developed a low-cost device embedded in the walls of a building to passively reflect or actively transmit radio signals. Dunna *et al.* [47] realized *ScatterMIMO*, which uses a smart surface to increase the scattering in the environment. In their hardware design, each reflector unit uses a patch antenna connected to four open-ended transmission lines. The transmission lines provide  $0, \pi/2, \pi$  or  $3\pi/2$  phase shifts. Based on measurements, it is shown that *ScatterMIMO* increases the throughput by factor of two and the signal-noise ratio (SNR) by 4.5 dB as compared with baselines schemes. It is worth mentioning that tunability can be achieved with implementations other than patch-array based surfaces. For example, PIVOTAL COMMWARE proposed a new technique called *Holographic Beam Forming (HBF)*. The proposed holographic beamformer has a low cost and power consumption, as compared with other transmission technologies, such as massive MIMO and phased arrays.

#### E. RIS Operating Principles

Considering single-beam reflection, a patch-array based RIS can be configured to serve a terminal device in the far-field and near-field regions. Among the many operating functions and configurations of RISs, *anomalous reflection* and *beamforming* are widely used in the context of wireless communications. Adopting the wave-optics perspective, anomalous reflection is a wavefront transformation from a plane wave to another plane wave, while beamforming is a wavefront transformation from a plane wave to a desired wavefront. Adopting

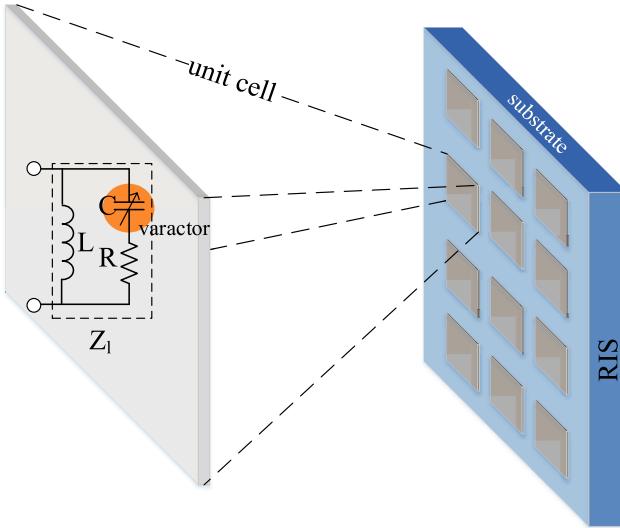


Fig. 7. Schematic diagram of the varactor RIS.

the ray-optics perspective, we present the operating principles of these two configurations in the following text. As far as anomalous reflection is concerned, the RIS is designed to reflect an incident beam to a far-field terminal, following the generalized laws of reflection [48]. As far as beamforming (also called focusing) is concerned, the incident wave is focused towards a targeted region, often referred to as the *focal point*. The required RIS configuration follows the co-phase condition [49]. The relation between these two operating principles is discussed in detail in [50]. Before presenting these two different principles, the physical distinction between the near-field region and the far-field region is clarified.

1) *Near Field v.s. Far Field*: In the spirit of dimensional analysis, the characteristics of a system can be represented by dimensionless numbers. In order to separate the near-field region from the far-field region, a proper dimensionless number is needed. Let  $L$  and  $R_F$  denote the antenna aperture size of the RIS and the focal distance, respectively. Assume that  $z$  is the distance of a particular field point to the RIS. Theoretically, the far-field and near-field regimes can be differentiated as follows: The distance of  $2L^2/\lambda$  is a commonly used criterion to decide the boundary between the near-field and far-field regions see [51, eq. (1)]. The position corresponding to  $z = 2L^2/\lambda$  is the boundary between the near-field region and the far-field region. This result comes from the inspection of the power density variation with the distance between a field point and the RIS. Within the near field where  $z < 2L^2/\lambda$ , the power density shows significant variations. The peak position of the power density in the near-field region, namely  $R_F$ , changes with different RIS configurations. Using proper co-phase conditions, beam focusing can be achieved within the near-field of the RIS. It is worth mentioning that, in general, the boundary between the near-field and far-field regimes depends on the specific configuration of the RIS, as it was recently remarked in [50].

In general, the essential difference between the near-field region and the far-field region is how the power density

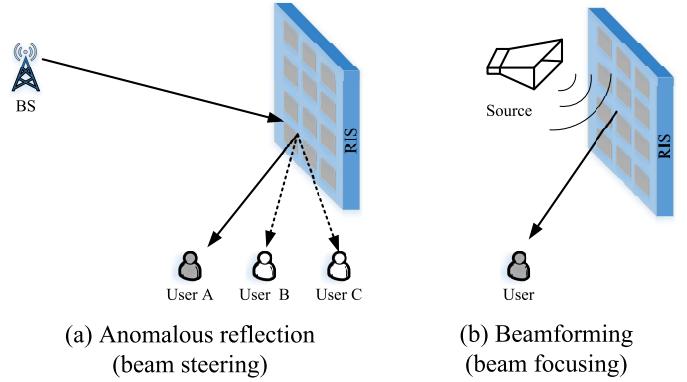


Fig. 8. Typical functions of reflecting surfaces.

changes with distance. Consider, for example, that the RIS focuses the wave within an area  $a$ . The total energy incident on the RIS is proportional to the solid angle,  $\Omega$  spanned by the surface area of the RIS with respect to the location of the transmitter. After the reflection, the transmitted energy is spread over the area  $a$ . Thus, the power density around the focal point is proportional to  $\Omega/a$ . Moreover, according to [45], the area  $a$  is proportional to  $\lambda^2(1 + 4z^2/L^2)$ , as a result of the Abbe diffraction limit. In the far-field region, the second term inside the brackets dominates and  $\Omega/a$  is proportional to  $L^2\Omega/z^2$ , which is the typical spherical dissipation of the signal power with the distance. In the near-field region, the first term dominates and the area  $a$  becomes very small. As a result, a high focusing gain can be achieved.

With the aid of an RIS, in general, the signal can be enhanced in both the near-field and far-field regimes. However, the rationale of this enhancement is different. For near-field applications, the RIS is expected to enhance the signal strength for users located at targeted locations with respect to the RIS, while reducing the signal at other locations. For far-field applications, the RIS is typically expected to enhance the signal strength for users located at targeted angles with respect to the RIS.

In the following text, adopting the ray-optics perspective, we discuss the generalized laws of refraction and reflection, as well as the corresponding co-phase condition.

2) *The Generalized Laws of Refraction and Reflection*: From a geometrical optics perspective, anomalous reflection and refraction from an RIS can be described by using the generalized laws of refraction and reflection [48], which is a natural derivation of both Fermat's principle and the boundary conditions governed by Maxwell's equations.

*Principle 1 (Achieving anomalous reflection [Fig. 8(a)]):* Suppose that the phase discontinuity at the boundary is a function of the position along the  $x$  direction  $\Phi(\vec{r}_x)$ , where  $\vec{r}_x$  is the position vector on the boundary. Moreover, suppose that the derivative of the phase discontinuity exists. Then, the angle of reflection ( $\theta_1$ ) and the angle of refraction ( $\theta_2$ ) are [52]:

$$\theta_1 = \sin^{-1} \left[ \sin \theta_i + \frac{\lambda}{2\pi n_1} \frac{d\Phi}{dx} \right], \quad (2)$$

$$\theta_2 = \sin^{-1} \left[ \frac{n_1 \sin \theta_i}{n_2} + \frac{\lambda}{2\pi n_2} \frac{d\Phi}{dx} \right], \quad (3)$$

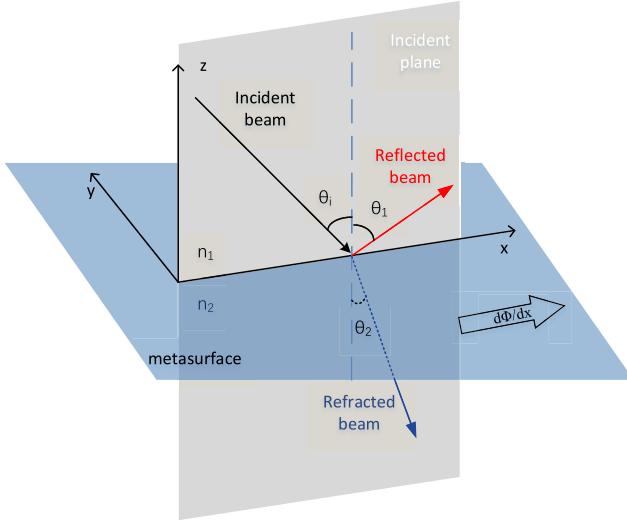


Fig. 9. Illustration of the generalized laws of refraction and reflection.

where  $\theta_i$  is the angle of incidence,  $\lambda$  is the wavelength of the transmitted signal in vacuum, and  $n_1, n_2$  are the refractive indexes, as shown in Fig. 9.

There are other results related to the generalized laws of refraction and reflection, including the critical angles for total internal reflection or refraction. The main result just presented here states that, when a phase discontinuity is introduced at the boundary surface, the angles of reflection and refraction depend not only on the angle of incidence but also on the wavelength, refractive indexes, and the gradient of the phase discontinuity. This gives extra controllable parameters to manipulate the reflected and refracted EM waves. As a result, anomalous reflection can be achieved by tuning the phase gradient ( $d\Phi/dx$ ) based on (2) or, in the discrete patch-array implementation, by tuning the length of the super-lattice. However, the assumption that the derivative of the phase discontinuity is constant ( $d\Phi/dx = \text{const.}$ ) does not necessarily hold if different wave transformations are needed.

3) *Co-Phase Condition*: Focusing is usually implemented when the RIS is within the near-field of the source or the terminal is close to the RIS. In these cases, the curvatures of the incident and reflected wavefront are non-negligible. The optimization of the surface aims to produce a pencil-beam pointing towards the direction of the terminal. When the link between the source and the RIS, as well as the link between the RIS and the terminal are in LoS, the following co-phase condition [49] can be applied.

*Principle 2 (Achieving beamforming (focusing) [Fig. 8(b)])*: Let  $r_{mn}$  denote the position of the  $(m, n)$ -th RIS element,  $r_s$  denote the position of the source, and  $\hat{u}$  denote the direction of the observer with respect to the surface. As shown in Fig. 10,  $\phi_{mn}$  can be chosen as follows [53]:

$$-k_0(|\vec{r}_{mn} - \vec{r}_s| - \vec{r}_{mn} \cdot \hat{u}) + \phi_{mn} = 2\pi \cdot t \quad (4)$$

where  $t = 1, 2, 3 \dots$  and  $k_0 = 2\pi/\lambda_c$ .

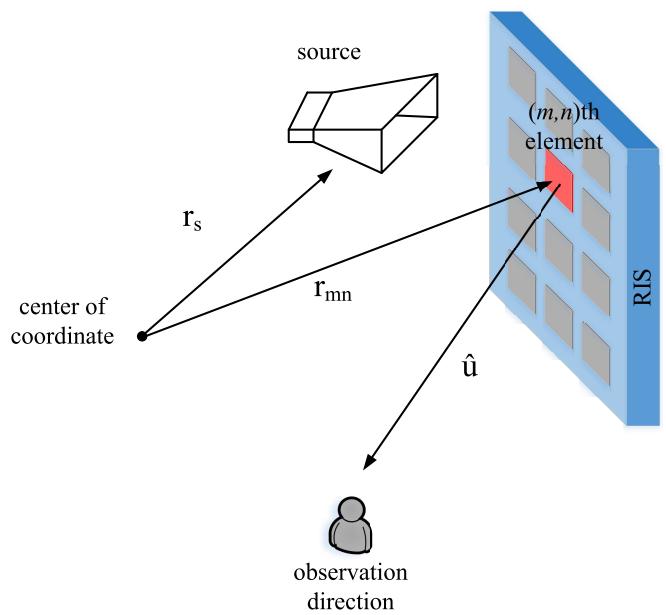


Fig. 10. Coordinate representation of co-phase condition.

The two design principles just discussed provide guidance on how to configure the RIS phase shift patterns for typical applications. In more complex wireless communication systems, the RIS role is more intricate, thus cannot be categorized by the two working functions described in Fig. 8. In these cases, to determine the RIS configuration, an optimization problem needs to be formulated. These issues are elaborated in the following sections.

#### F. Discussions and Outlook

In the design and configuration of RISs, both theoretical limitations and hardware implementation limitations affect the overall performance of the resulting system. Theoretical limitations are, e.g., the result of considering simplified assumptions or adopting naive perspectives during the modeling of the RIS and its interaction with the EM wave. Hardware implementation limitations come, e.g., from the discretization of an ideally continuous RIS profile. In the following text, we elaborate on three major points.

1) *Hardware Limitations*: In practical application scenarios, many hardware parameters significantly affect the achievable performance of the system. For example, the number of quantization levels of the RIS phase shifts, the maximum number of elements that is possible to integrate on the substrate, and the percentage of scattering environment that can be coated by an RIS. Existing research contributions studied the limitations and tradeoffs caused by these hardware constraints by analyzing their effect on the channel distribution [54], the power scaling law [55], and performance metrics such as the outage probability [56] and ergodic capacity [57]. At the time of writing, the analysis of the impact of hardware limitations and how an RIS performs compared to other available technologies are open research issues.

2) *System Design Simplifications*: The adoption of oversimplified models for RIS hardware or channel models result

in limitations on the system design. Because of the complex nature of RISs and their interaction with the environment, initial research contributions have adopted simple models. For example, using hardware models based on local designs, the ray-optics perspective for channel modeling, and decoupling reflection amplitude and phase shift of the RIS elements. Even though the ray-optics approximation can yield effective designs in some cases (as shown in Section II-E), it is preferable to adopt the wave-optics design in practical cases [50]. Further research efforts are needed to bridge complex physical models of different RIS implementations with widely used communication models [43].

*3) Optimization Limitations:* To reap the benefits of deploying RISs in wireless networks, the RIS parameters (e.g., the reflection coefficient and deployment location) and network resource allocation (e.g., transmit beamforming and user scheduling) should be jointly optimized. However, the resulting joint optimization problems are normally non-convex and involve highly-coupled variables, which make it challenging to derive a globally optimal solution. Though some efficient algorithms have been recently proposed to compute high-quality suboptimal solutions [58]–[60], the performance limits of RISs remain unknown. To fully understand the attainable performance limits, information-theoretic perspective investigations [61], [62] are important and sophisticated mathematical tools [63] are expected to be employed. Further details are given in Sections VI and V.

### III. PERFORMANCE ANALYSIS OF MULTI-ANTENNA ASSISTED RIS SYSTEMS

In Section II, we have discussed the fundamental physical properties of the RISs. However, how RISs affect the communication performance is still an open problem. To systematically survey existing designs for RIS-enhanced networks, we discuss the following topics: (1) channel models, (2) performance analysis, and (3) benchmark schemes.

#### A. Channel Models

*1) Path Loss Models:* Some research contributions on the path loss for RISs are available in [7], [50] and [64], which showed that the power scattered by an RIS is usually formulated in terms of an integral that accounts, by leveraging the Huygens principle, for the impact of the entire surface in the free-space scenario, where scattering, shadowing, and reflection are ignored. Closed-form expressions of the integral are, on the other hand, difficult to obtain, except for some asymptotic regimes, which correspond to viewing the RIS as electrically small and electrically large (with respect to the wavelength and the transmission distances). It is worth mentioning, in addition, that the path-loss model depends on the particular phase gradient applied by the RIS. Notably, the scaling laws can be different if the RIS operates as an anomalous reflector and as a focusing lens. In the following, we briefly discuss two scaling laws that have recently been reported for RISs that operate as anomalous reflectors (described in the previous section). Further information and details can be found in [7] and [50].

- **Electrically Small RISs:** In this asymptotic regime, the RIS is assumed to be relatively small in size compared with the transmission distances. In this regime, the RIS can be approximated as a small-size scatterer. In general, the path-loss scales with the reciprocal of the product of the distance between the transmitter and the center of the RIS and the distance between the center of the RIS and the receiver. In addition, the received power usually increases with the size of the RIS. The received power is usually maximized in the direction of anomalous reflection, where the path loss through the RIS follows the “product of distances” models, which can be formulated as:

$$L(d_{SR}, d_{RU}) \approx \lambda_S (d_{SR} d_{RU})^{-1}, \quad (5)$$

where  $\lambda_S$  denotes the coefficient of the electrically small scenario,  $d_{SR}$  and  $d_{RU}$  represent the distance of source-RIS and RIS-user links, respectively. A detailed discussion is given in [50, Secs. IV-A–IV-C].

- **Electrically Large RISs:** In this asymptotic regime, the RIS is assumed to be large (ideally infinitely large) in size compared with the transmission distances and the wavelength. In this regime, the RIS can be approximated as a large flat mirror. Let us denote by  $x_0$  the point of the RIS (if it exists) at which the first-order derivative of the total phase response of the combined incident signal, reflected signal, and the surface reflection coefficient of the RIS is equal to zero. In general, the path-loss asymptotically scales with the reciprocal of a weighted sum of the distance between the transmitter and  $x_0$  and the distance between  $x_0$  and the receiver. In addition, the received power is not dependent on the size of the RIS, which is viewed as asymptotically infinite. This result substantiates the fact that the power scaling law of the RIS is physically correct, since it does not grow to infinity as the size of the RIS goes to infinity. This is because the scaling law and the behavior of the RIS are different with respect to the electrically small regime. In this case, the path loss through the RIS follows the “sum of distances” models, which can be approximated as:

$$L(d_{SR}, d_{RU}) \approx \lambda_L (d_{SR} + d_{RU})^{-1}, \quad (6)$$

where  $\lambda_L$  denotes the coefficient of the electrically large scenario. A detailed discussion is given in [50, Secs. IV-A–IV-C].

- *2) Spatial Models:* Stochastic geometry (SG) tools are capable of capturing the location randomness of users thus enabling the derivation of computable or closed-form expressions of key performance metrics. Specifically, several spatial processes exist for modeling the locations of users in different wireless networks, i.e., the homogeneous Poisson point process (HPPP) [65], [66], the Poisson cluster process (PCP) [67], [68], the Binomial point process (BPP), as well as the Hard core point process (HCPP) [69]. We list some promising approaches for analyzing the performance of RIS-enhanced networks by using SG in Table IV. In [70], the RIS elements are employed on obstacles, and it is assumed that randomly distributed users are located in the serving area of the

TABLE IV  
SUMMARY OF RIS-ENHANCED SG NETWORKS

Approaches	Advantages	Disadvantages	Ref.
RIS-enhanced HPPP	Fairness-oriented design	Restricted user distribution	[5]
RIS-enhanced two layer HPPP	Coverage-hole enhancement	RISs are deployed at obstacles	[70]
HPPP conditioned on angle	Practical design	Not tractable	[73]
RIS-enhanced HPPP conditioned on angle	Practical design for RIS-enhanced networks	Complicated	-

RISs. In [71], the objects are modeled by a modified random line process of fixed length and with random orientations and locations. Therein, the probability that a randomly distributed object that is coated with an RIS acts as a reflector for a given pair of transmitter and receiver was investigated. In [72], PCP was invoked in the RIS-enhanced large-scale networks, where the angle of reflection is constrained by the angle of incidence. Therefore, the randomly distributed users and BSs are located at the same side of the RIS.

3) *Small-Scale Fading Models*: Currently, two main approaches have been used for analyzing the performance of RIS-aided systems in small-scale fading channels: (1) the central-limit-theorem-based (CLT-based) distribution and (2) the use of approximated distributions.

- **CLT-based Distribution:** Let us consider a single-antenna BS that communicates with a single-antenna user with the aid of an RIS of  $N$  elements. If the two received signals, from the BS and from the RIS, can be coherently combined, the effective channel power gain is given by

$$\begin{aligned} & \left| \mathbf{r}^H \Phi \mathbf{g} + h \right|^2, \\ & \text{s.t. } \beta_1, \dots, \beta_N = 1 \\ & \phi_1, \dots, \phi_N \in [0, 2\pi), \end{aligned} \quad (7)$$

where  $h \in \mathbb{C}^{1 \times 1}$ ,  $\mathbf{g} \in \mathbb{C}^{M \times 1}$ ,  $\mathbf{r} \in \mathbb{C}^{M \times 1}$  denote the channels of the BS-user, BS-RIS, and RIS-user links, respectively.  $\Phi = \text{diag}(\beta_1 e^{j\phi_1}, \beta_2 e^{j\phi_2}, \dots, \beta_N e^{j\phi_N})$  denotes the reflection-coefficient matrix of the RIS, where  $\{\beta_1, \beta_2, \dots, \beta_N\}$  and  $\{\phi_1, \phi_2, \dots, \phi_N\}$  represent the amplitude coefficients and phase shifts of the RIS elements, respectively. In this setup, the CLT-based technique stands as an approximation tool for analyzing the performance in the low-medium-SNR regime. This is due to the fact that the distribution of the probability density function (PDF) in the range 0 to 0+ is not precise by using the CLT [74]. In Rayleigh fading channels, the distribution of an RIS-enhanced link follows a modified Bessel function [74]. Since both transmitter and RISs are part of the infrastructure, and the RISs are typically positioned to exploit the LoS path with respect to the locations of the transmitters and the receivers for increasing the received signal power, Zhang *et al.* [56] studied Rician fading channels, and the analysis showed that the signal power follows a non-central chi-squared distribution with two degrees of freedom. Ding and Poor [75] proposed an RIS-enhanced network, where RISs are utilized for effectively aligning the directions of the users' channel gains. By utilizing the CLT-based technique, Cheng *et al.* [76] studied the multi-RIS network,

where the channel distributions were investigated with or without BS-user links.

- **Approximated Distribution:** The exact distribution of the received SNR of the signal reflected from an RIS is non-trivial to be obtained, and hence the use of approximated distributions is often necessary. Qian *et al.* [55] proposed a simple approximated distribution of the received SNR, and proved that the received SNR can be approximated by two (or one) Gamma random variables and the sum of two scaled non-central chi-square random variables. A prioritized signal enhancement design was proposed by Hou *et al.* [77], where both the outage performance and ergodic rate of the user with the best channel gain were calculated. Lyu and Zhang [78] proposed a single-input and single-output (SISO) network with multiple randomly deployed RISs, and showed that the exact distribution in terms of received signal power can be approximated by a Gamma distribution. Makarfi *et al.* [26] proposed an RIS-enhanced network, whose equivalent channel is modeled by the Fisher-Snedecor  $\mathcal{F}$  distribution.

Based on the above mentioned contributions [26], [56], [74], [75], [77], [78], where only approximated channel distributions are obtained, the exact channel distribution of RIS-enhanced networks is still an open problem. Based on recent research results, for example, a fundamental limitation lies in the calculation of the diversity order of RIS-enhanced networks under ideal operating conditions and in the presence of hardware limitations, e.g., quantized phase shifts. For example, the diversity order obtained by using the CLT-based distribution [74] is  $\frac{1}{2}$  in the high-SNR regime, whereas the diversity order is  $\frac{N}{3}$  if an approximation based on the Gamma distribution is used, where  $N$  denotes the number of RIS elements. However, the CLT-based and Gamma-based distributions are not exact, making the performance analysis of RIS-enhanced networks an interesting problem for future research. Furthermore, since the exact distribution contains higher-order components, which approach zero in the high-SNR regime, most of the previous contributions [4], [79] adopt the approximated distribution method for modeling small-scale fading channels, and the exact distribution of RIS-enhanced networks is still an open problem. For example, recent exact results on the impact of phase noise on the diversity order of RIS-enhanced transmission can be found in [80].

### B. Performance Analysis

In this subsection, we briefly discuss currently available papers on the performance analysis of RISs that are realized as large arrays of tiny and inexpensive antennas whose

TABLE V  
IMPORTANT CONTRIBUTIONS ON RIS-ENHANCED NETWORKS. “DL” AND “UL” REPRESENT DOWNLINK AND UPLINK, RESPECTIVELY. THE “SUM-RATE GAIN” IMPLIES THAT THE GAIN BROUGHT BY INVOKING RIS TECHNIQUE

Ref.	Scenarios	Direction	Users	Main Objectives	Techniques
[5]	MIMO	DL	Multiple users	OP and throughput	Fairness SEB
[12]	SISO	DL	Single user	sum-rate gain	Compare with relay
[26]	SISO	UL	Single user	OP and throughput	Effective channel gain
[74]	SISO	DL	Single user	Effective channel gain	Compare with random phase shifting
[56]	SISO	DL	Single user	Effective channel gain	SEB
[75]	SISO	DL	Single user	OP	SEB
[77]	SISO	DL	Multiple users	OP and throughput	Prioritized SEB
[78]	SISO	DL	Single user	Effective channel gain	SEB
[55]	MIMO	DL	Single users	Effective channel gain	Random matrix theory and CLT
[83]	MIMO	DL	Multiple users	sum-rate gain	SEB
[82]	MIMO	DL	Multiple users	Interference cancellation	SCB and less constraint at RAs
[84]	SISO	UL	Multiple users	Sum-rate	Minimum required finite resolution
[85]	MISO	DL	Multiple users	Sum-rate	Multi-RIS distribution
[86]	MISO	DL	Multiple users	Sum-rate	Discrete phase shifts

phase response is locally optimized. By offering extra diversity in the spatial domain, multi-antenna techniques are of significant importance. The application of multi-antenna enhanced RIS networks has attracted substantial interest from academia [5], [77], [78] and industry [12], [81], [82]. Given the increasing number of research contributions on RISs, its advantages are becoming more clear, especially in terms of spectral efficiency (SE) and energy efficiency (EE) enhancement. There are several key challenges for performance analysis in RIS-enhanced networks. One of the main challenges is to evaluate the exact distributions of the cascade channels between the BS and users through RISs. Another challenge is evaluating the effective channel gain after passive beamforming at the RIS. Table V summarizes the existing contributions on RISs with multiple antennas and illustrates their comparison. RIS-enhanced single user networks have been analyzed in Section III-A. Hence, we turn our attention to RIS-enhanced multi-user networks. A prioritized signal-enhancement-based (SEB) was proposed by Hou *et al.* [77], where passive beamforming is designed for the user with the best channel gain, and all the other users rely on RIS-enhanced beamforming.

- **RIS-enhanced Signal Enhancement Designs:** By assuming that multiple waves are co-phased at the users, the received signal can be significantly enhanced, which leads to the following optimization problem:

$$\begin{aligned} & \max \left| \mathbf{r}^H \Phi \mathbf{g} + h \right|^2 \\ & \text{s.t. } \beta_1, \dots, \beta_N = 1 \\ & \quad \phi_1, \dots, \phi_N \in [0, 2\pi]. \end{aligned} \quad (8)$$

In order to further enhance the SE of RIS-enhanced networks, multiple antenna techniques can be employed at both the BS and users. Yuan *et al.* [85] proposed a cognitive-radio-based RIS-enhanced multiple-input and single-output (MISO) network, where both perfect and imperfect channel state information (CSI) setups were considered. However, in many research works, continuous amplitude coefficients and phase shifts are assumed at the RISs, whilst in practice the phase shifts of RISs may not be continuous. Thus, You *et al.* [86] proposed a discrete phase shifts model for a MISO enhanced RIS network.

Zhang *et al.* [84] then evaluated the required number of bits for finite-resolution RISs in an uplink SISO network. Hou *et al.* [5] investigated an RIS-enhanced MIMO network, where a fairness oriented design was considered by applying SG tools for modeling the impact of the users’ locations.

- **RIS-enhanced Signal Cancellation Designs:** Another application of deploying RISs in wireless networks is signal cancellation, where the reflected signals and the direct signals can be destructively combined. The corresponding optimization problem can be formulated as follows:

$$\begin{aligned} & \min \left| \mathbf{r}^H \Phi \mathbf{g} + h_I \right|^2 \\ & \text{s.t. } \beta_1, \dots, \beta_N \leq 1 \\ & \quad \phi_1, \dots, \phi_N \in [0, 2\pi], \end{aligned} \quad (9)$$

where  $h_I$  denotes the aggregate interference signals from other-cell BSs. By assuming that both the inter-cell and intra-cell interferences are perfectly known, the optimal solution to (9) is to adjust both the signal phase and amplitude coefficients of the BS-RIS-user links to the opposite of the effective interference links with the same amplitude. By doing so, some promising applications can be realized, e.g., RIS-enhanced PLS and interference cancellation. On the one hand, by assuming that perfect CSI is available at the RIS controller, the inter-cell and intra-cell interferences can be eliminated. On the other hand, considering the PLS requirements, RISs also stand as a potential solution for cooperative jamming techniques, i.e., RISs act as artificial noise sources. By adopting this approach, several contributions have been made. Hou *et al.* [82] proposed an RIS-enhanced interference cancellation technique in a MIMO network, where the inter-cluster interference can be eliminated without active beamforming weights and detection vectors. Furthermore, this work can be adopted for application to coordinated multi-point (CoMP) networks for inter-cell interference cancellation in cellular networks. Shi *et al.* [87] investigated an RIS-enhanced secure beamforming technique, where the secrecy rate of the legitimate user was derived. Lyu *et al.* [88] investigated an RIS jamming scenario,

TABLE VI  
COMPARISON OF THE RIS-ENHANCED AND BENCHMARKS. THE “POWER” IMPLIES ADDITIONAL POWER SUPPLY AT THE RISs OR AT THE RELAY. THE “CT” DENOTES CONCURRENT TRANSMISSION

Mode	Pros	Cons	Delay	Power	Interference	CT
RIS-enhanced	High EE, simple device	CSI must be perfectly known	✗	✗	✗	✓
HD relaying under AF protocol	No decoding at relay	Noise is amplified	✓	✓	✗	✗
HD relaying under DF protocol	No self-interference	Latency is high	✓	✓	✗	✗
FD relaying under AF protocol	No decoding at relay	Noise and interference are amplified	✗	✓	Self-interference	✓
FD relaying under DF protocol	No latency	Rate ceiling occurs	✗	✓	Self-interference	✓
MIMO relay	High SE	High cost, difficult to realize at mmWave	✗	✓	✓	✓

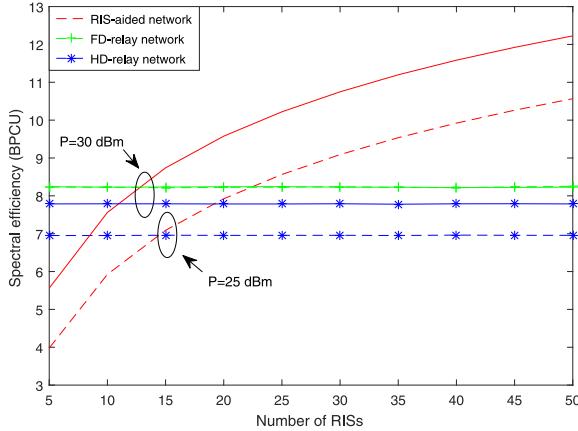


Fig. 11. Spectral efficiency of RIS-enhanced, FD-relay as well as HD-relay networks versus the number of RIS elements. Please refer to [77] for simulation parameters.

where RISs act as jammers for attacking legitimate communications without using any internal energy.

### C. Benchmark Schemes

In order to assess the advantages and limitations of RISs, two benchmark transmission technologies are usually considered: 1) surfaces with random phase shifts; and 2) relay networks.

- **Random Phase Shifts:** RISs are capable of shifting the phase of the incident signal, and hence multiple signals can be boosted or eliminated at the user side or at the BS side. Hence, a well-accepted benchmark scheme to quantify the performance enhancement by RIS elements is given by a surface that is not configurable and that can ideally be modeled by a surface with random phase shifts [74].

- **Relay Networks:** Generally speaking, relay-aided networks can be classified into two pairs of classic relaying protocols, which are 1) FD and HD relay networks; and 2) AF and DF relay networks. By assuming that the optimal power split strategies of both the AF and DF relays are employed, the performance gain between RIS-enhanced and relay-aided networks can be compared. Specifically, Björnson *et al.* [12] compared the achievable data rate of both RIS-enhanced and DF-relay-aided SISO network, where the BS-user links are blocked. It was pointed out that when the number of tunable elements of the RISs is large enough, an RIS-enhanced network is capable of outperforming a DF-relay-aided network. In an

effort to provide a comprehensive analysis for both RIS-enhanced and relay-aided networks, Ntontin *et al.* [83] compared the system performance of classic maximal ratio transmission (MRT) and maximal ratio combining (MRC) techniques. Fig. 11 illustrates the potential benefits of RIS-enhanced networks compared with both HD-relay and FD-relay networks in terms of network throughput [77]. Here, the performance of HD-relaying is obtained for an equal time-split ratio. We can see that the network throughput gap between the RIS-enhanced network and the other pair of relay aided networks becomes smaller, when the number of RIS elements increases. For example, when the number of RIS elements  $N = 23$  and the transmit power  $P = 25$  dBm, the proposed RIS-enhanced network is capable of outperforming both FD and HD relay aided networks, which indicates that the RIS-enhanced network becomes more competitive, when the number of RIS elements is large enough. Table VI provides a comparison between RIS-enhanced and relay aided networks in terms of advantages and limitations.

### D. Discussions and Outlook

Although previous research contributions have analyzed the approximated performance of RIS-enhanced networks, there are still three major open research problems.

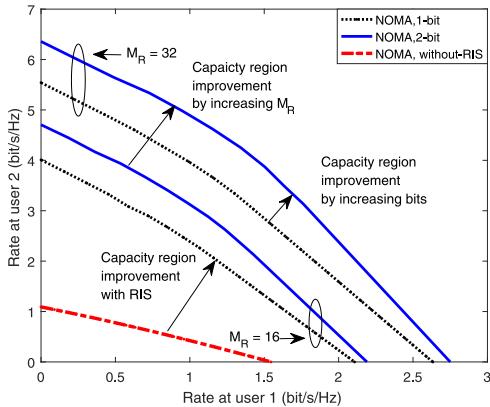
- 1) Path loss experiments for outdoor scenarios: Since only [64] reported experimental measurements of the path loss in free-space environments, the development of experimentally-validated path-loss models in outdoor scenarios is an open research issue, especially in the presence of reflecting and scattering objects.

- 2) Exact distributions: Current research methods for performance analysis are based on the CLT-based distribution and approximated distribution [26], [76], which are, however, accurate only in the high-SNR regime. More advanced and accurate analytical models are needed for analyzing the diversity order and the network performance in the low-SNR regime.

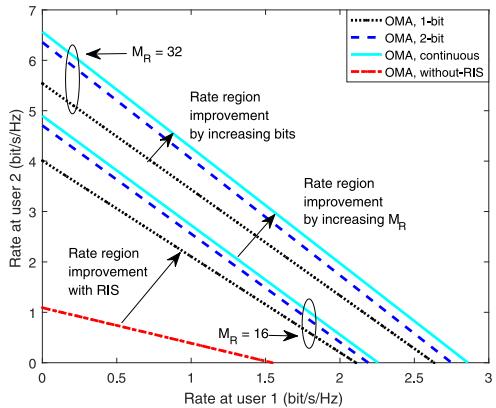
- 3) Integrated application scenarios: To serve the desired users in different scenarios, signal enhancement and signal cancellation designs were well investigated in recent research works [77], [82], [83]. However, the desired signals and interference signals can be simultaneously enhanced and mitigated, which constitutes an important future direction.

## IV. RIS BEAMFORMING AND RESOURCE ALLOCATION

As described in the previous section, deploying RISs enables high performance enhancements in wireless networks.



(a) Capacity regions with RIS-NOMA.



(b) Rate regions with RIS-OMA.

Fig. 12. Illustration of the capacity and rate regions for a random channel realization with different RIS phase resolutions.  $M_R$  denotes the number of RIS reflection elements. The full parameter settings can be found in [62].

Motivated by this benefit, how to jointly design the transmit and passive beamforming as well as the optimal allocation of the wireless resources has become an important task for RIS-enhanced wireless networks. In this section, we first discuss the information-theoretic performance limits of RISs. Then, we review recent research contributions with a particular focus on the joint beamforming optimization and the resource allocation design. Along the literature review, representative mathematical tools for facilitating the two types of design are also discussed along with their benefits and drawbacks.

#### A. Information-Theoretic Perspective

In order to unveil the fundamental performance limits of RISs, several research works [61], [62], [89] have been devoted to investigating the RIS performance gains from an information-theoretic perspective.

- *Capacity-achieving design:* In [61], Karasik *et al.* derived the capacity for an RIS-aided single-input and multiple-output (SIMO) communication system. With finite input signal constellations, it was proved that a joint information encoding scheme at both the transmitted signals and the RIS configurations is necessary for achieving the channel capacity [61]. Based on this insight, the authors further proposed a practical transmission strategy

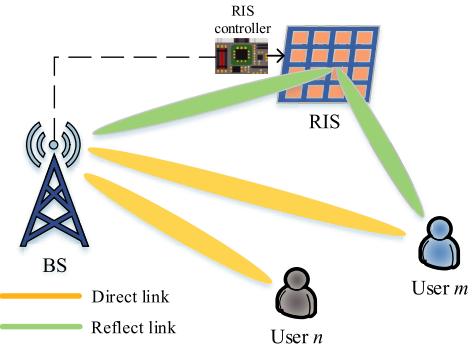


Fig. 13. Illustration of joint transmit and passive beamforming design.

by utilizing layered encoding and successive cancellation decoding techniques. Numerical examples showed that the proposed joint encoding scheme outperforms the conventional max-SNR scheme.

- *Capacity region characterization:* The capacity region of the fading SISO broadcast channel (BC) was proved to be achieved by invoking the superposition coding (SC) at the transmitter and the successive interference cancellation (SIC) at multiple receivers [90], i.e., employing the NOMA transmission. Inspired by these results, Mu *et al.* [62] investigated the capacity and rate regions of RIS-enhanced multi-user wireless communication systems achieved by NOMA and orthogonal multiple access (OMA), respectively. The Pareto boundary of each region was characterized by solving a series of sum rate maximization problems via the rate-profile technique. As shown in Fig. 12(a) and Fig. 12(b), by deploying an RIS, the NOMA capacity region and the OMA rate region can be improved. The capacity/rate regions are further enlarged by employing more precise phase resolutions and larger numbers of reflection elements. Furthermore, Zhang and Zhang [89] investigated the capacity region of the multiple access channel (MAC) with two users in both centralized and distributed RIS deployment strategies. The results demonstrated that the centralized RIS deployment strategy can achieve a higher capacity gain than the distributed strategy.

#### B. Joint Transmit and Passive Beamforming Design

- 1) *Optimization Objectives:* As shown in Fig. 13, an RIS is deployed to assist the transmission between the BS and the users by passively reflecting the signals. The RIS reflection coefficients can be adjusted by the BS through an RIS controller. Hence, the transmit beamforming at the BS and the passive beamforming at the RIS have to be jointly designed to improve the communication performance. In the following, we review the related research works in terms of their considered optimization objectives.

- *Transmit power minimization or EE maximization:* In [58], Wu and Zhang minimized the transmit power for an RIS-enhanced MISO system in both single-user and multi-user scenarios. Alternating optimization (AO) based algorithms were developed to find locally-optimal solutions. The passive beamforming was designed by

invoking the semidefinite relaxation (SDR) approach. It was revealed that an RIS can simultaneously enhance the desired signal strength and mitigate the interference for the multi-user scenario. The same problem was further investigated in [91] by taking discrete RIS phase shifts into consideration. The optimal solutions were derived by applying the branch-and-bound method and exhaustive search for single-user and multi-user scenarios, respectively. To reduce the computational complexity, efficient successive refinement algorithms were further designed. It was shown that the proposed low complexity algorithms are capable of achieving near-optimal performances. Han *et al.* [92] investigated physical-layer broadcasting in an RIS-aided network, where the total transmit power for satisfying QoS requirements of all users was minimized. Furthermore, Fu *et al.* [93] focused on an RIS-enhanced MISO downlink NOMA system, where the transmit power was minimized by jointly optimizing the transmit and passive beamforming vectors as well as user decoding orders. In order to overcome the drawbacks of the SDR approach, an alternating difference-of-convex (DC) method was proposed for handling the non-convex rank-one constraint. Zhu *et al.* [94] proposed an improved quasi-degradation condition for the RIS-enhanced MISO NOMA system to minimize the transmit power. Under this condition, NOMA is shown to be able to outperform the zero-forcing beamforming scheme. Zheng *et al.* [95] compared the minimum transmit power performance between OMA and NOMA in a discrete phase shift RIS-enhanced SISO system. A near-optimal solution was obtained by applying the linear approximation initialization and the AO method. The results showed that NOMA may perform worse than TDMA when the users have symmetric deployments and rate requirements, which revealed the importance of user pairing in the RIS-assisted NOMA system. Huang *et al.* [59] solved the EE maximization problem in an RIS-enhanced multi-user MISO system, where a realistic RIS power consumption model was proposed in terms of the number of reflection elements and the phase resolutions at the RIS. An AO-based algorithm was designed for addressing the formulated problem, where the RIS phase shifts and the BS transmit power allocation were optimized by invoking the gradient descent method and the fractional programming method. The results demonstrated that the RIS achieves significantly better EE performance than the traditional active relay-assisted communication. In contrast to the aforementioned works based on the perfect CSI assumption, Zhou *et al.* [96] investigated the robust beamforming design for an RIS-enhanced multi-user MISO system with imperfect CSI assumptions. The transmit power was minimized while satisfying QoS requirements of all users under all possible channel error realizations. The formulated non-convex problem was transformed into a sequence of semidefinite programming (SDP) subproblems, where the CSI uncertainties and the non-convex unit-modulus constraints were handled by applying approximation transformations

and the convex-concave procedure [97], respectively. In [98], the robust beamforming design was further studied under two channel error models, namely the bounded CSI error model and the statistical CSI error model. The S-procedure and the Bernstein-Type inequality were applied in each model. The results unveiled that the RIS may degrade the system performance when the channel error is high. Zappone *et al.* [99] modeled the overhead for carrying out the channel estimation and adjusting the RIS. Based on the proposed overhead model, the EE of an RIS-empowered MIMO communication network was maximized by jointly optimizing the RIS phase shifts as well as the transmitted and received filters.

- *SE or capacity maximization:* Yu *et al.* [100] considered the SE maximization problem in an RIS-enhanced MISO system. Since the SDR approach only provides an approximate solution [58], two efficient algorithms were proposed by invoking the fixed-point iteration method and the manifold optimization method for the passive beamforming design. It was demonstrated that the proposed algorithms can achieve a higher performance and consume a lower complexity than the SDR approach. To solve the same problem, a branch-and-bound algorithm was further proposed by Yu *et al.* [101], which is capable of obtaining a globally optimal solution. Though suffering from an extremely high computational complexity, the proposed branch-and-bound algorithm serves as a performance benchmark to verify the effectiveness of existing suboptimal algorithms. In [102], Ning *et al.* focused on an RIS-enhanced downlink MIMO system to maximize the SE. The passive beamforming was designed by using the sum of gains maximization principle and by utilizing the alternating direction method of multipliers. In [103], Ying *et al.* considered an RIS-enhanced mmWave hybrid MIMO system, where the phase shifts at the RIS were designed by leveraging the angle information of the LoS BS-RIS channel. Moreover, Perović *et al.* [9] investigated RIS-assisted indoor mmWave communications, where two schemes were developed to maximize the channel capacity. Zhang and Zhang [104] characterized the fundamental capacity limit of RIS-aided point-to-point MIMO communication systems, by jointly optimizing the RIS reflection coefficients and the MIMO transmit covariance matrix. The communication capacity was maximized in both the narrowband transmission with frequency-flat fading channels and the broadband orthogonal frequency division multiplexing (OFDM) transmission over frequency-selective fading channels. Yang *et al.* [8] proposed a practical transmission protocol by considering the channel estimation for an RIS-enhanced OFDM system under frequency-selective channels. To reduce the required training overhead, the RIS reflection elements were divided into multiple groups and only the combined channel of each group has to be estimated. Based on the proposed grouping scheme, the achievable rate was maximized by jointly optimizing the power allocation at the transmitter and the phase shifts at the RIS with the AO

TABLE VII  
CONTRIBUTIONS ON JOINT TRANSMIT AND PASSIVE BEAMFORMING DESIGN

Ref.	Scenarios	Phase shifts	CSI	Main Objectives	Techniques/Characteristics
[58]	SU/MU-DL-MISO	Continuous	Perfect	Transmit power	AO-SDR-based algorithm and two stage algorithm
[91]	SU/MU-DL-MISO	Discrete	Perfect	Transmit power	Near-optimal ZF-based successive refinement method
[92]	MU-DL-MISO	Continuous	Perfect	Transmit power	Physical-layer broadcasting
[93]	MU-DL-MISO NOMA	Continuous	Perfect	Transmit power	Alternating DC method
[94]	MU-DL-MISO NOMA	Continuous	Perfect	Transmit power	Improved quasi-degradation condition
[95]	MU-DL-SISO NOMA	Discrete	Perfect	Transmit power	Asymmetric and symmetric user pairing
[59]	MU-DL-MISO	Continuous	Perfect	EE	RIS power consumption model
[96]	MU-DL-MISO	Continuous	Imperfect	Transmit power	The worst-case robust beamforming design
[98]	MU-DL-MISO	Continuous	Imperfect	Transmit power	Imperfect cascaded channels at the transmitter
[99]	SU-DL-MIMO	Continuous	Estimated	EE	Overhead model for channel estimation and RIS configuration
[100]	SU-DL-MISO	Continuous	Perfect	SE	Fixed point iteration and manifold optimization methods
[101]	SU-DL-MISO	Continuous	Perfect	SE	Branch-and-bound algorithm
[102]	SU-DL-MIMO	Continuous	Perfect	SE	Sum of gains principle
[103]	SU-DL-MIMO mmWave	Continuous	Perfect	SE	Broadband hybrid beamforming
[9]	SU-DL-MIMO mmWave	Continuous	Perfect	Channel capacity	RIS-assisted indoor mmWave environments
[104]	SU-DL-MIMO	Continuous	Perfect	Channel capacity	Frequency-flat fading and frequency-flat selective channels
[8]	SU-DL-SISO OFDMA	Continuous	Estimated	Achievable rate	RIS element grouping scheme
[105]	SU-UL-SISO	Discrete	Estimated	Achievable rate	Near-orthogonal DFT-Hadamard based reflection patterns
[106]	MU-DL-MISO	Continuous	Perfect	Sum rate	Majorization-minimization approach
[107]	MU-DL-MISO	Continuous/Discrete	Perfect	Weighted sum rate	Iterative algorithms with closed-form expressions
[108]	MU-DL-MISO	Discrete	Perfect	Sum rate	Interference-free modulation scheme
[60]	MU-DL-MISO NOMA	Continuous/Discrete	Perfect	Sum rate	Sequential rank-one constraint relaxation approach
[109]	MU-DL-MISO	Discrete	Statistical CSI	Sum rate	Low channel training overhead
[110]	MU-DL-MISO	Continuous	Statistical CSI	Max-min SINR	Signal exchange overhead reduction
[111]	MU-DL-SISO/MISO NOMA	Continuous	Perfect	Max-min rate	Near-optimal NOMA user ordering scheme

method. In [105], You *et al.* designed a transmission protocol by considering the channel estimation with discrete phase shifts at the RIS. To reduce the channel estimation errors, a low complexity discrete fourier transform (DFT)-Hadamard based reflection pattern scheme was developed. The achievable data rate was further maximized based on the estimated channel by designing the RIS phase shifts using the proposed successive refinement algorithm.

- *Sum rate maximization:* In [106], Huang *et al.* maximized the sum rate in RIS-enhanced multi-user MISO downlink communications. By employing the zero-forcing precoding at the BS, the RIS reflection matrix and the power allocation were alternately optimized with the aid of the majorization-minimization approach. Moreover, the weighted sum rate maximization problem was investigated by Guo *et al.* [107]. Under the AO framework, the transmit beamforming was obtained using the fractional programming method, and three iterative algorithms were designed for optimizing the reflection coefficients in terms of different types of RIS reflection elements. In [108], the asymptotic optimal discrete passive beamforming solution was derived and a modulation scheme was proposed to maximize the achievable sum rate for the RIS-enhanced multi-user MISO transmission. To further enhance the performance, Jung *et al.* [108] designed a joint user scheduling and transmit power control scheme, which can strike the tradeoff between the rate fairness and the maximum sum rate among the users. Mu *et al.* [60] focused their attention on the sum rate maximization problem in an RIS-enhanced MISO NOMA system with both ideal and non-ideal assumptions of RIS elements. The non-convex rank-one constraint of the passive beamforming design was handled by invoking the sequential rank-one constraint relaxation approach, which is guaranteed to obtain a locally optimal rank-one solution.

Instead of optimizing the passive beamforming with the instantaneous CSI, Zhao *et al.* proposed a two-timescale transmission protocol for maximizing the achievable average sum rate in an RIS-enhanced multi-user system [109]. To reduce the channel training overhead and complexity, the RIS phase shifts were firstly optimized with the statistical CSI. Then, the transmit beamforming was designed with the instantaneous CSI and the optimized RIS phase shifts.

- *User fairness:* Nadeem *et al.* [110] maximized the minimum SINR of an RIS-enhanced MISO system, where the BS-RIS-user link was assumed to be a LoS channel. A deterministic approximations was developed for the minimum SINR performance under the optimal linear precoder by employing the random matrix theory. As a result, the RIS phase shifts can be optimized using the channel's large-scale statistics, which can significantly reduce the overhead of the signal exchange [110]. Yang *et al.* [111] investigated the max-min rate problem in an RIS-enhanced NOMA system in both single-antenna and multi-antenna cases. A combined-channel-strength based user ordering scheme was proposed for achieving a near-optimal performance.

All the aforementioned research contributions on the joint transmit and passive beamforming design are summarized in Table VII. “SU” and “MU” represent single-user and multi-user, respectively. “DL” and “UL” represent downlink and uplink, respectively.

2) *Approaches for Passive Beamforming Design:* An example of the joint transmit and passive beamforming design problem can be formulated as follows

$$\begin{aligned}
 & \min_{\mathbf{w}, \boldsymbol{\theta}} / \max_{\mathbf{w}, \boldsymbol{\theta}} f(\mathbf{w}, \boldsymbol{\theta} | \mathcal{H}) \\
 & \text{s.t. } \mathbf{w} \in \mathcal{T}, \\
 & \quad \boldsymbol{\theta} \in \mathcal{P},
 \end{aligned} \tag{10a}$$

TABLE VIII  
SUMMARY OF APPROACHES TO PASSIVE BEAMFORMING DESIGN

Approaches	Phase shifts	Advantages	Disadvantages	Ref.
SDR	Continuous	Relax to convex problem	Require rank-one solution construction	[58], [94], etc.
Quantization method	Discrete	Easy to implement	Substantial performance loss	[111]
Branch-and-bound	Continuous/Discrete	Optimal solution	Relatively high complexity	[91], [101]
Iterative algorithms	Continuous/Discrete	Good complexity-performance tradeoff	Performance depends on initialization	[91], [93], etc.

where  $\mathcal{H}$  denotes the set of given CSI,  $\mathbf{w}$  and  $\boldsymbol{\theta}$  denote the transmit beamforming vector and passive beamforming vector, respectively,  $\mathcal{T}$  and  $\mathcal{P}$  denote the corresponding feasible set for  $\mathbf{w}$  and  $\boldsymbol{\theta}$ , respectively. Here,  $f(\mathbf{w}, \boldsymbol{\theta} | \mathcal{H})$  denotes the objective function that depends on  $\mathbf{w}$  and  $\boldsymbol{\theta}$  given the CSI. Let  $\theta_i = \beta_i e^{j\phi_i}$  be the  $i$ th element of the passive beamforming vector  $\boldsymbol{\theta}$ . Depending on the specific implementation of the RIS, three case studies can be considered [60].

- **Continuous amplitude and phase shift:** In this case, it is assumed that the amplitude and phase shift of each RIS element can be adjusted continuously, which results in the following feasible set.

$$\mathcal{P}_1 = \{\beta_i, \phi_i | \beta_i \in [0, 1], \phi_i \in [0, 2\pi)\}. \quad (11)$$

- **Constant amplitude and continuous phase shift:** In this case, it is assumed that the amplitude and phase shift of each RIS element are fixed, e.g.,  $\beta_i = 1$ , and can be adjusted continuously, respectively. The corresponding feasible set is given by

$$\mathcal{P}_2 = \{\beta_i, \phi_i | \beta_i = 1, \phi_i \in [0, 2\pi)\}. \quad (12)$$

- **Constant amplitude and discrete phase shift:** In this case, it is assumed that the amplitude and phase shift of each RIS element are fixed, e.g.,  $\beta_i = 1$ , and can be adjusted based on a discrete set of values, respectively. The feasible set can be expressed as

$$\mathcal{P}_3 = \{\beta_i, \phi_i | \beta_i = 1, \phi_i \in \mathcal{D}\}, \quad (13)$$

where  $\mathcal{D} = \{0, \frac{2\pi}{N}, \dots, (N-1)\frac{2\pi}{N}\}$  and  $N$  denotes the number of candidate phase shifts.

It is worth noting that the first two case studies are difficult to realize in practice. Due to the hardware constraints in fact, it is quite challenging to realize continuous amplitude and phase shift control. However, the first two case studies can be used to characterize the theoretical performance upper bounds of RISs.

By inspection of the three case studies just considered, we evince that the joint beamforming optimization problem is generally a non-convex problem since  $\mathbf{w}$  and  $\boldsymbol{\theta}$  are coupled together. The existing algorithms for solving the non-convex joint beamforming optimization in RIS-assisted wireless networks are mainly based on the AO method. The advantage of this approach is that, given the passive beamforming vector, the transmit beamforming design becomes a conventional problem, which has been extensively investigated. However, the passive beamforming design under given transmit beamforming vectors is still a non-trivial task to tackle. The main challenges to solve the problem include the unit modulus constraint and the discrete nature of the feasible set. In the following, we list the approaches employed in

current research contributions for optimizing the passive beamforming. Table VIII summarizes the characteristics of those approaches.

- **SDR:** A common method for handling the non-convex unit-modulus constraint is to transform the passive beamforming vector into a rank-one and positive semi-definite matrix. By applying the SDR approach, which ignores the non-convex rank-one constraint, the original non-convex problem becomes a convex SDP problem that can be solved by using many efficient convex optimization tools. If the obtained matrix solution is not rank-one, Gaussian randomization methods are usually used. However, the constructed rank-one solution is, in general, suboptimal and it may be even infeasible for the original passive beamforming design problem, which not only causes some performance degradation but cannot guarantee the convergence of the AO-based iterative algorithm either.
- **Quantization method:** Under the assumption of finite resolution phase shifts, one method is to relax each discrete phase shift variable  $\phi_i \in \{0, \frac{2\pi}{N}, \dots, (N-1)\frac{2\pi}{N}\}$  into a continuous variable  $\phi_i \in [0, 2\pi)$ . After solving the relaxed problem, the obtained continuous solutions are quantized to their nearest discrete values. However, the quantization method may lead to a substantial performance loss, especially for low-resolution phase shifts. Additionally, it is worth mentioning that the non-convex unit-modulus constraints still exist after applying the continuous relaxations.
- **Branch-and-bound:** Due to the non-convex nature of the passive beamforming design problem, it is challenging to obtain the optimal solution with standard convex optimization techniques. The branch-and-bound approach has been applied for solving polynomial-time (NP)-hard discrete and combinatorial optimization problems and some specific continuous optimization problems. For example, the branch-and-bound approach was adopted for deriving the optimal solution of the discrete passive beamforming design [91] and the continuous passive beamforming design [101].
- **Iterative algorithms:** The main idea of iterative algorithms is to obtain a locally optimal or a high-quality suboptimal solution for the original problem at an acceptable computational complexity. Some iterative algorithms were developed for the passive beamforming design, such as the successive refinement algorithm [91], [104], [105], [107], the alternating DC algorithm [93], the conjugate gradient search [59], the fixed point iteration method and the manifold optimization method [100], and the sequential rank-one constraint relaxation approach [60]. It was shown

TABLE IX  
SUMMARY OF APPROACHES TO RESOURCE ALLOCATION PROBLEMS

Approaches	Advantages	Disadvantages	Ref.
Binary relaxation	Relax to convex feasible set	Existence of performance gap	-
Matching theory	Achieve near-optimal performance	Require predefined preference list	[113]
Heuristic algorithms	Flexible complexity-performance tradeoff	Unstable performance	[114]

that these proposed iterative algorithms can achieve a good tradeoff between performance and computational complexity.

### C. Resource Management in RIS-Enhanced Networks

1) *Resource Allocation Problems*: In Fig. 14, a large-scale RIS-assisted transmission scenario is considered, where multiple BSs serve multiple users with the aid of multiple RISs. In this context, several key issues need to be discussed.

- *Subchannel assignment*: The bandwidth efficiency can be improved by properly allocating users to different subchannels. If the RIS elements are not frequency-selective, one single common RIS reflection matrix needs to be shared among all subchannels, which makes the resulting optimization problems challenging to solve. To address this difficulty, Yang *et al.* [112] proposed a dynamic passive beamforming scheme, where the resource blocks are dynamically allocated to different user groups with varied RIS phase shifts over different time slots. In [113], Zuo *et al.* investigated the joint subchannel assignment, power allocation, and passive beamforming design problem in a multi-channel downlink RIS-NOMA network.

- *User-RIS association*: In multi-RIS assisted multi-user communications, how to associate the users to different RISs is an interesting problem. The user-RIS association schemes in general determine the overall network performance. Considering a multi-RIS assisted massive MIMO system, Li *et al.* [114] found that the automatic interference cancellation property holds for RISs with infinitely large sizes. Then, the considered max-min SINR problem can be transformed into a user-RIS association problem, which was efficiently solved by the proposed greedy search algorithm.

- *Multi-cell RIS association*: In multi-cell scenarios, the optimization problem becomes much sophisticated for jointly considering the user-BS association, the user-RIS association, and the subchannel assignment. In some initial works considering multi-cell scenarios [18], [115], RISs were deployed to enhance the performance of cell-edge users for simultaneously improving the received desired signal power and mitigating the received interference from other cells.

2) *Approaches to Resource Allocation Problems*: Scheduling different users with different subchannels/RISs/BSs is an NP-hard problem. Though the optimal solution can be obtained by exhaustively searching over all possible association combinations, it requires a prohibitively high computational complexity, especially for the large-scale networks in Fig. 14. Therefore, low complexity and efficient algorithms have to be developed for striking

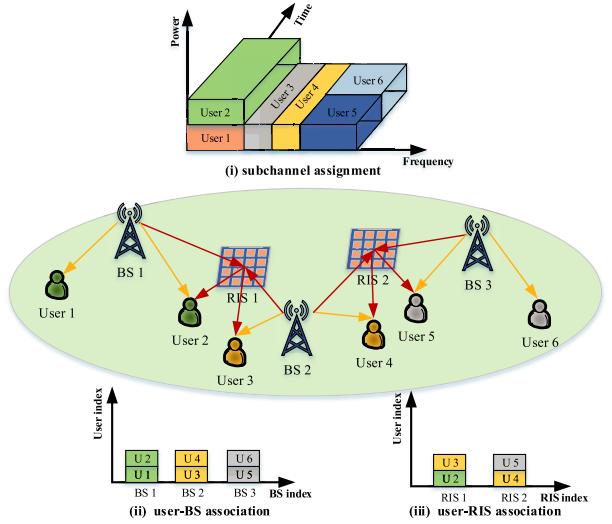


Fig. 14. Illustration of resource management in large-scale RIS-assisted networks.

a performance-versus-complexity tradeoff. In the following, we present some promising approaches, and discuss their advantages and disadvantages. Table IX summarizes the characteristics of those approaches.

- **Binary relaxation**: One idea is to relax the binary variable  $\alpha \in \{0, 1\}$  into the continuous variable  $\alpha \in [0, 1]$ , where  $\alpha$  represents the user association state. By doing so, the non-convex integer constraint is relaxed to be convex, and conventional convex optimization techniques can be applied for solving the relaxed problem. It is worth pointing out that the relaxed problem might still be non-convex especially when the optimization variables are highly-coupled. Additional efforts, such as utilizing the successive convex approximation (SCA) method, are required to obtain an approximate solution. Moreover, this kind of relaxation may result in a considerable performance loss between the original integer problem and the relaxed one.

- **Matching theory**: Matching theory is a powerful method developed for solving the combinatorial user association problems. The user association combinational optimization problem in RIS-enhanced networks can be modeled as a high dimensional Users-BSs-RISs-Subchannels matching problem. Though high dimensional matching is NP-hard, it can be decomposed into several 2D matching subproblems, which can be efficiently solved. For example, Zuo *et al.* [113] applied many-to-one matching theory for the subchannel assignment in an RIS-enhanced NOMA system, which can achieve a near-optimal performance. However, leveraging matching theory requires to establish a predefined

preference list for both users and resources. As the channel conditions always fluctuate in RIS-enhanced networks, these preference lists may need to be dynamically updated, which needs further investigations.

- **Heuristic algorithms:** For solving computationally complex problems, one commonly employed method is to develop heuristic algorithms, where approximate solutions for the original optimization problem can be obtained with an acceptable computational complexity. In the aforementioned works, the greedy search based heuristic algorithm was designed for solving the user-RIS association problem [114]. However, the performance of heuristic algorithms is sensitive to the designed strategies, which is not always stable.

#### D. Discussions and Outlook

With the growing number of research contributions on RIS-enhanced communications, the advantages of RISs have been verified in terms of SE, EE, and user fairness. However, most of the existing treatises solved the non-convex joint beamforming optimization problem employing the AO method, which decouples the joint transmit and passive beamforming design into two subproblems. Though a high-quality suboptimal solution can be obtained, advanced optimization techniques are required for solving this problem to characterize the optimal performance gain introduced by RISs, which also provides an important benchmark for verifying the optimality of any other low complexity suboptimal algorithms. Furthermore, in RIS-enhanced communication systems, it is known that obtaining accurate CSI is rather challenging due to the nearly passive working mode of RISs. Investigating the robust joint beamforming design and resource allocation [96], [98] constitutes an important research direction for practical RIS implementations.

Besides passive beamforming, RISs also introduce the following DoFs, which can be further exploited to reap the benefits of RISs in the future work.

- **RIS deployment design:** The reflection link via the RIS experiences a severer path loss than the direct link. Therefore, the deployment location of the RIS has to be carefully designed to achieve considerable performance enhancements. How to jointly optimize the passive beamforming and the deployment location at the RIS as well as the wireless resource allocation at the AP/BS is a non-trivial task, which deserves further research efforts. In particular, one of prominent challenges is that the deployment location of the RIS determines both the path loss and the LoS components of the reflection channels, which causes the optimization variables to be highly-coupled. Therefore, efficient algorithms need to be designed. An initial study [116] has investigated the optimal RIS deployment strategy for both NOMA and OMA transmissions, which showed that asymmetric and symmetric RIS deployment locations among users are preferable for NOMA and OMA, respectively. Additionally, the RIS is usually deployed to avoid signal blockage to achieve a LoS dominated channel, thus

having a small path loss. However, such a LoS channel based RIS deployment strategy may be ineffective, especially for RIS-assisted multi-user communications. This is because the LoS dominated channels are low-rank, the resulting ill-condition channel matrices significantly limit the achievable capacity even with a relatively small path loss. How to deploy the RIS to strike a balance between the path losses and the Non-LoS (NLoS) components of channels is another interesting problem, which deserves further research interests.

- **Dynamical RIS configuration:** In [61], [62], the authors revealed that, from an information-theoretic perspective, the capacity-achieving transmission schemes need to dynamically adjusting the RIS. However, most of the existing research contributions assumed that the RIS reflection coefficients can be only adjusted once for each channel coherence duration. Note that one of the most typical application scenarios of RISs is to assist transmission of the users, who are static or moving slowly in the vicinity of RISs. In this case, the duration of one channel coherence block is usually tens of milliseconds, which is much larger than the time duration for adjusting the RIS (e.g., 220 microseconds in [45]). Therefore, adjusting RIS multiple times in one channel coherence duration, namely the dynamic RIS configuration, is practically valid. This unique characteristic opens up new research opportunities, such as dynamical passive beamforming and resource allocation schemes, which merit further investigations.

## V. MACHINE LEARNING FOR RIS-ENHANCED COMMUNICATION SYSTEMS

ML techniques have gained remarkable interests in wireless communications due to their learning capability and large search-space [29], [117], [118]. We survey existing research contributions, which apply ML techniques for tackling challenges in RIS-enhanced wireless networks. Finally, potential research challenges and opportunities of ML-empowered RIS systems are presented.

### A. Motivations and Architecture for Integrating ML in RIS-Enhanced Wireless Networks

In this subsection, we first present the challenges of the conventional RIS-enhanced wireless networks and the motivations for integrating ML in these networks, followed by the system architecture of ML-empowered RIS-enhanced wireless networks.

To effectively exploit RISs for optimizing wireless networks, preliminary research contributions have studied a number of technical challenges that include channel estimation/modeling, joint transmit and passive beamforming design, as well as resource allocation from the BS to the users. Powerful optimization techniques, such as convex optimization [107], iterative algorithm [58], gradient descent approach [59], and alternating optimization algorithm [119] have been adopted for addressing the aforementioned fundamental challenges. Although important insights have been

gained by these research contributions, the following limitations still exist in conventional RIS-enhanced wireless networks:

- The users are generally assumed to be static for simplicity, i.e., the dynamic mobility of users is typically ignored. Another limitation in the existing literature is that the communication environment is assumed to be perfectly known, the differentiation of users' demand is always ignored as well.
- The RIS/BS are not capable of learning from the unknown environment or from the limited feedback of the users. In practical applications of RISs in wireless networks, the system parameters are treated as random variables, which naturally leads itself to the derivation of insightful joint probability distributions conditioned on the users' tele-traffic demand and mobility. However, this is a highly dynamic stochastic environment, which is difficult for employing conventional optimization approaches. Additionally, the feedback from the users is usually resource-hungry and limited, which aggravates the challenges for the conventional RIS-enhanced wireless networks.
- Finally, instantaneous CSI of all the channels are assumed to be available at the BS. However, CSI acquisition in RIS-enhanced wireless networks becomes more challenging than that in the conventional relay systems due to the passive nature of RISs, which also aggravates the challenge imposed on the conventional RIS-enhanced wireless networks.

#### B. Deep Learning for RIS-Enhanced Communication Systems

Deep learning (DL) has shown great potentials to revolutionize communication systems. It can be applied in diverse areas of RIS-enhanced wireless networks due to its powerful learning capabilities [120]–[122].

The acquisition of timely and accurate CSI plays a pivotal role in wireless systems, especially in MIMO networks. However, CSI acquisition becomes more challenging due to the large number of antennas in massive MIMO systems [123]. In order to tackle this challenge, a number of research contributions have adopted DL for estimating the CSI, especially for exploiting CSI structures beyond linear correlations.

In contrast to the conventional AF relay-aided wireless networks, in RIS-enhanced wireless networks, the RIS is a passive device, which is not capable of performing active transmission/reception and signal processing [124]. In an effort to estimate a large number of unknown parameters caused by RISs, Taha *et al.* [125] exploited the DL method for learning the RIS reflection matrices directly from the sampled channel knowledge without any knowledge of the RIS array geometry. Liu *et al.* [126] proposed a deep denoising neural network assisted compressive channel estimation for RIS-assisted mmWave systems with a low training overhead. Elbir *et al.* [127] presented a DL framework for channel estimation in the RIS-enhanced MIMO system. It was shown that the proposed convolutional neural networks (CNNs)-based approach achieves lower normalized mean-square-error (NMSE) and more robust performance than other benchmarks.

The data-driven DL approach has the advantage of model-free representation or function learning such that no explicit models of the complicated wireless channels are needed, at the expense of requiring large amounts of training data and corresponding computational power. Thus, the DL method can be adopted for estimating the CSI of RIS-enhanced wireless networks.

Apart from the aforementioned applications of DL in RIS-enhanced wireless networks, Huang *et al.* [128] leveraged a deep neural network (DNN)-based approach in the indoor communication environment for estimating the mapping between a user's position and the configuration of the RIS to maximize the received SNR. Additionally, DL can also be applied for learning the optimal RIS phase shift configuration. Gao *et al.* [129] proposed a DL-based algorithm for optimally designing the phase shift of the RIS by training the DL offline. It can be observed that the proposed unsupervised learning mechanism outperforms the conventional optimization approach in terms of computational complexity. Khan and Shin [130] investigated the signal estimation and detection in the RIS-enhanced wireless networks. A DL-based approach was proposed for estimating channels and phase angles from a reflected signal received by an RIS. With the aid of DL, the bit-error-rate (BER) performance of the system was improved.

#### C. Reinforcement Learning for RIS-Enhanced Communication Systems

Reinforcement learning (RL) is a powerful AI paradigm that can be used to empower agents by interacting with the environment. More explicitly, by exploiting the learning capability (e.g., learning from the environment, learning from the feedback of users, and learning from its mistakes) of the RL model, the challenges encountered in the conventional RIS-enhanced wireless networks may be mitigated, thus leading to improved performance.

The core idea of employing RL techniques in the RIS-enhanced wireless networks is that they allow the BS/RIS to improve their service quality by learning from the environment, from their historical experience, and from the feedback of the users [131]. More explicitly, RL models can be used for supporting the BS/RIS (agents) in their interactions with the environment (states), whilst finding the optimal behavior (actions) of the BS/RIS. Furthermore, the RL model can incorporate farsighted system evolution (long-term benefits) instead of only focusing on current states. Thus, it is applied for solving challenging problems in the RIS-enhanced wireless networks.

As illustrated in Fig. 15, the RL algorithms can be divided into three categories, namely, value-based algorithms, policy-based algorithms, and actor-critic algorithms. Both advantages and disadvantages exist in the RL algorithms. Since RISs have discrete phase shifts, the DQN algorithm is more suitable for tackling the corresponding phase shift design problem.

To fully reap the benefits of deploying RISs in wireless networks, the joint transmit and passive beamforming design of the RIS-enhanced system has been considered in MISO systems [142], [143], OFDM-based systems [144], wireless

RL algorithm	Key features	Pros	Cons
Q-learning	Utilize the Q-table to train Q-value offline	Can find the optimal policy without requiring knowledge about the environment	Only suitable for scenarios with small state space and action space
SARSA	Allows the agent to approach the optimal policy online	Allow the agent to choose optimal actions at each time step in a real-time fashion	Only suitable for scenarios with small state space and action space
DQN	Invoke NN to as an approximator of Q-function	Can be invoked in scenarios with large state space and action space	Over estimation on action values
Double DQN	Invoke two Q-function to select and evaluate Q-values, respectively	Can solve the problem of over estimation	Only discrete state space and action space
Dueling DQN	Invoke two NNs to estimate the action value and state value	Have a faster convergence rate than DQN model	High complexity
Noisy DQN	Adding Gaussian noise layer in the DQN model	Enhance the exploring performance	Only discrete state space and action space
Distributional DQN	Invoke a distribution function instead of Expectation to update Q-function	Higher accuracy than DQN in terms of evaluating Q-function	Require the distribution information of reward function
Asynchronous DQN	Multi-agent can train parallelly	Have a faster learning speed than DQN	High complexity
Retrace	Return-based off-policy algorithm	More efficient than Q-learning	Only suitable for scenarios with small state space and action space
PG	Approximate a stochastic policy directly using an independent function approximator with its own parameters	Easy for application;	Sensitive to the parameter setting
TRPO	Invoke a loss function to find the optimized parameters	Have a faster convergence rate than value-based approaches	Can only be invoked in special limiting cases; Have a low data efficiency
PPO	Alternate between sampling data through interaction with the environment	Easier and more general than TRPO	Have a tradeoff between sample complexity, simplicity and wall-time
AC	Utilize critic to update the value function parameters, while utilize actor to update the policy parameters	Have the advantages of both value-based algorithm and policy-based algorithm	Have a lower convergence rate
A3C	Multi-agent can train parallelly	Changes in the approximated function get propagated much more quickly	Have a higher variance
DDPG	Combination of AC and DQN	Have the advantages of both policy gradient and DQN	Higher complexity; Can not be invoked in a random scenario
TD3	Invoke two critics to update the value function parameters	Overcome the problem of over estimation	Higher complexity
SAC	Based on maximum entropy	More stable; Enhance the exploring performance	Higher complexity

Fig. 15. Key features, pros and cons of RL algorithms [131]–[141]. (PG represents Policy Gradient, TRPO denotes Trust Region Policy Optimization, PPO represents Proximal Policy Optimization, AC denotes Actor-Critic, A3C represents Asynchronous Advantage Actor-Critic, DDPG denotes Deep Deterministic Policy Gradient, TD3 represents Twin Delayed DDPG, SAC denotes Soft Actor-Critic).

security systems [145] and millimeter wave systems [146] with the aid of RL algorithms. In contrast to the AO method, which alternately optimizes the transmit beamforming at the BS and the passive beamforming at the RIS, the RL-based solution is capable of simultaneously designing them. More explicitly, Huang *et al.* [142] applied a deep deterministic policy gradient (DDPG) based algorithm for maximizing the throughput by utilizing the sum rate as instant rewards for training the DDPG model. In the proposed model, the continuous

transmit beamforming and RIS phase shift were jointly optimized with low complexity. Taha *et al.* [144] proposed a deep reinforcement learning (DRL) based algorithm for maximizing the achievable communication rate by directly optimizing interaction matrices from the sampled channel knowledge. In the proposed DRL model, only one beam was utilized for each training episode. Thus, the training overhead was avoided, while the dataset collection phase was not required. Zhang *et al.* [146] presented a DRL based algorithm for

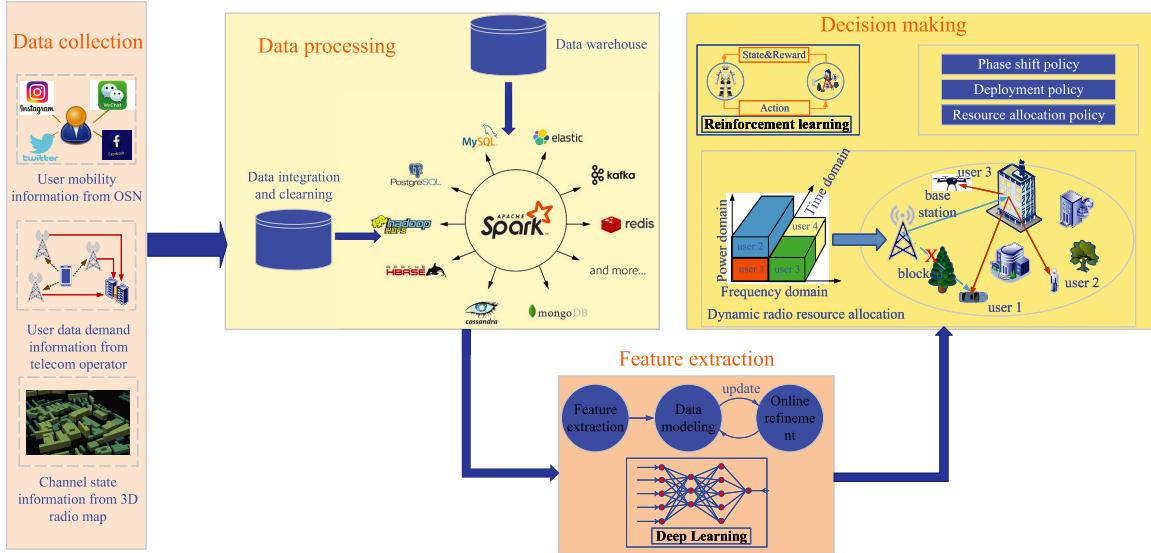


Fig. 16. Architecture of ML-empowered RIS-enhanced wireless networks.

maximizing the throughput with both perfect and imperfect CSI. A quantile regression method was applied for modeling a return distribution for each state-action pair, which modeled the intrinsic randomness in the MDP interaction between the RIS and communication environment. Helin *et al.* [145] considered the application of RISs to PLS. The system secrecy rate was maximized with the aid of the DRL model by jointly optimizing the beamforming and phase shift matrices under different users' QoS requirements and time-varying channel conditions. Additionally, post-decision state and prioritized experience replay schemes were utilized to enhance the learning efficiency and secrecy performance.

#### D. A Novel Architecture of ML-Empowered RIS-Enhanced Wireless Networks

As a benefit of the ML-based framework, many challenges in conventional wireless communication networks have been circumvented, leading to enhanced network performance, improved reliability, and agile adaptivity [117]. Fig. 16 illustrates a novel ML-empowered architecture for RIS-enhanced wireless networks. As shown in this figure, RISs are installed on the facade of a building for enhancing the wireless performance [3], [30]. The RIS is linked with a controller, which controls the reflecting elements for hosting the functionality of phase-shifting and amplitude absorption. A two-step approach is applied in the proposed ML-empowered RIS-enhanced wireless networks.

- As illustrated in the data collection, data processing, and feature extraction parts in Fig. 16. The associated user information (e.g., device type, position, data rate demand, mobility, caching demand, and computing ability) is collected, stored, and processed. Thus, the users' behaviors and requirements can be predicted for efficiently deploying and operating the RIS. Meanwhile, the predicted information can be modified online with the currently collected data as the input.

- Given the extracted features, adaptive schemes are leveraged for controlling the RISs, designing the phase shifts, resource allocation, and interference cancellation.

In the proposed ML-empowered RIS-enhanced wireless networks, RISs are capable of rapidly adapting to the dynamic environment by learning both from the environment and from the feedback of the users.

Research on the RIS deployment is fundamental and essential. However, there is a paucity of research on the problem of RIS position determination. Additionally, current research contributions mainly consider the performance optimization for both single-user and multi-user scenarios by optimizing the phase shift and/or precoding solutions of the RIS-enhanced communication systems [6], [18], [110], [147], [148].

Considering the RIS deployments based on the users' mobility information and particular data demand implicitly assumes that the long-term movement information and tele-traffic requirement of users are capable of being learned/predicted. With this proviso, the deployment and control method of RISs may be designed periodically for maximizing the long-term benefits and hence reducing the control overhead. By considering the long-term mobility and data demand of users, RIS-enhanced wireless networks become highly dynamic systems. Meanwhile, in an effort to maximize the service quality in an unknown environment, RISs are supposed to learn by interacting with the environment and adapting the control/deployment policy based on the limited feedback of the users to overcome the uncertainty of the environment.

In this subsection, an RL-based model is presented to jointly design the deployment policy and phase shift policy of RISs while considering the time-varying data demand of users. As illustrated in Fig. 17, in the RL-based model, the BS acts as an agent. Since a controller is installed, the BS can control both resource allocation policy for users and the RIS's position and phase shifts. At each timeslot, the BS periodically observes the state of the RIS-enhanced system. The state space

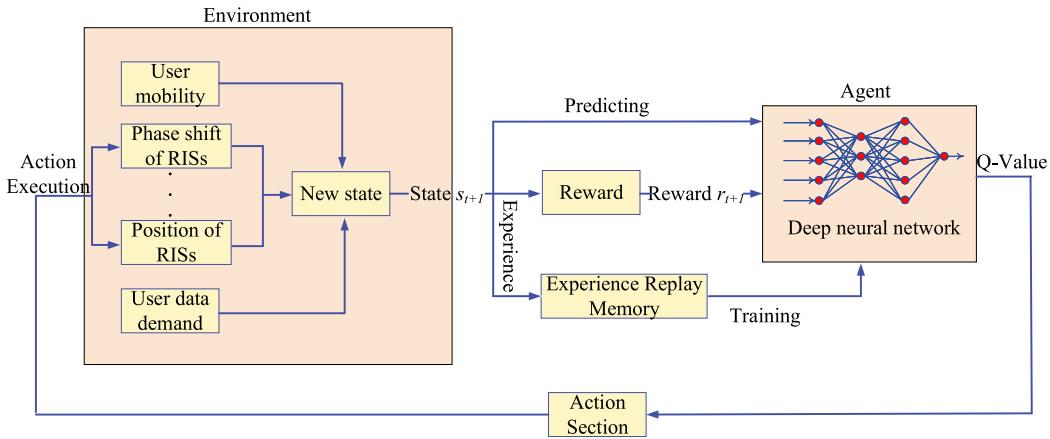


Fig. 17. DRL model for RIS networks.

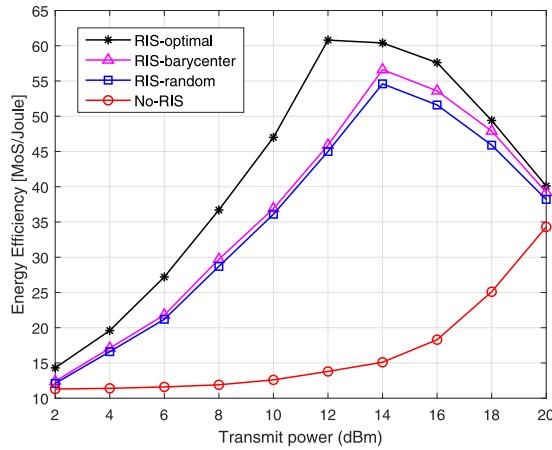


Fig. 18. EE with and without RIS [63].

consists of the RIS phase shifts, the allocated power to each user, as well as coordinates of both the RIS and users. An action is carried out by the BS for selecting the optimal control policy. The actions contain changing positions and varying phase shifts of the RIS, as well as varying the allocated power. The key underlying principle of the decision policy is carrying out an action that makes the DQN model obtain the maximum Q-value at each time slot. Following each action, the BS receives a penalty/reward,  $r_t$ , determined by the formulated objective function.

- **State of the RL model.** The state space consists of four parts: 1) the current phase shift of each reflecting element at the RIS; 2) the current 3D position of the RIS; 3) the current 2D position of each user; 4) the current power allocated from the BS to each user.
- **Action of the RL model.** The action space consists of three parts: 1) the variable quantity of the  $n$ -th reflecting element's phase shift; 2) the moving direction and distance of the RIS; 3) the variable quantity of the  $k$ -th user's transmit power.
- **Reward of the RL model.** The reward function is decided by the EE of the system. When action taken by the BS improves EE, the BS obtains a reward. Otherwise, when a reduction occurs in EE, the BS receives a penalty.

Fig. 18 characterizes the EE of the system in networks both with and without the assistance of an RIS. The EE is defined as the ratio between the system achievable sum mean opinion score (MOS) and the sum energy dissipation in Joule. It was shown that the EE of the system is enhanced by employing an RIS. The RIS-barycenter line indicates that the RIS is placed at the barycenter of all users. The RIS-random line indicates that the RIS is randomly deployed, while the RIS-optimal line indicates that the RIS is deployed at the optimal position derived from the proposed decaying double deep Q-network (D<sup>3</sup>QN) algorithm. The results of Fig. 18 confirm that there exists an optimal position for the RIS as far as the EE of the RIS-enhanced system is concerned. The performance of the RIS-enhanced system is improved by deploying the RIS at the optimal position compared to the random deployment strategy and the strategy of placing it at the barycenter.

#### E. Other ML Techniques for RIS-Enhanced Communication Systems

Besides DL and RL techniques, a range of supervised learning and unsupervised learning algorithms have been applied in the current generation wireless networks. Thus, these approaches can also be adopted for tackling challenges in the RIS-enhanced wireless systems.

1) *Supervised Learning Techniques for RIS-Enhanced Communication Systems:* As one of the key branches of ML, powerful supervised learning techniques, such as regression, decision tree and random forest, K-nearest neighbors (KNN), support vector machines (SVM), and Bayes classification, have been adopted in diverse scenarios for tackling challenges such as spectrum sensing [149], traffic/QoE prediction [150], channel/antenna selection [151], and networking association [152]. In the RIS-enhanced wireless networks, supervised learning algorithms can also be applied for solving the related problems with sufficient training data due to their advantages of low complexity and fast convergence speed.

2) *Unsupervised Learning Techniques for RIS-Enhanced Communication Systems:* In contrast to the supervised learning techniques, unsupervised learning methods do not rely on prior knowledge, which is not data-hungry. Thus, the

unsupervised learning algorithms [153] such as K-means clustering, expectation-maximization, principal component analysis (PCA), and independent component analysis (ICA) can be applied in the RIS-enhanced wireless networks for tackling challenges such as BS deployment, user clustering/association [154], channel/network state detection [155], data aggregation [156], and interference cancellation [157].

*3) Federated Learning Techniques for RIS-Enhanced Communication Systems:* Federated learning, which explores training statistical models directly on remote devices, has become a focal point in the area of large-scale ML and distributed optimization [158]. Since the federated learning algorithm is trained at the edge in distributed networks, the inaccessibility of private data is no longer a problem. Due to its privacy-preserving nature, federated learning algorithms can be applied for the deployment and design of multiple RISs, where each RIS can act as a distributed learner, trains its generated data and transfers its local model parameters instead of the raw training dataset to an aggregating unit. Thus, the deployment and design policy can be learned in a decentralized manner.

#### F. Discussions and Outlook

By exploiting ML learning capabilities, the aforementioned challenges encountered in RIS-enhanced wireless networks may be mitigated. This is due to the reason that RISs can learn by interacting with the environment and adapt the control/deployment policy based on the feedback of users to overcome the dynamic/uncertainty of the environment. It can be learned that the RL model can incorporate farsighted system evolution instead of only focusing on current states, which can reap long-term benefits for RIS-enhanced wireless networks. However, ML models will also pose some new challenges, including the layer design for the DL model, the state-action construction, and the reward function design for the RL model. In addition, simultaneously employing multiple RISs becomes more challenging due to the cooperation amongst RISs. Hence, intelligent deployment and design for multi-RIS enhanced wireless networks is highly desired. Finally, in the current ML models, either discrete or continuous state space is modeled for optimizing parameters in RIS-enhanced wireless networks, while joint discrete and continuous parameters exist in the networks. Hence, the joint discrete and continuous state space design in ML-enabled RIS-enhanced wireless networks is still challenging and constitutes an interesting topic.

## VI. INTEGRATING RISs WITH OTHER TECHNOLOGIES TOWARDS 6G

Current research contributions have proved that RIS-enhanced wireless networks are capable of obtaining tuned channel gains, improved QoS, enhanced coverage range, and reduced energy dissipation. These significant performance enhancements can be applied to diverse wireless communication networks. In this section, we identify the major issues and research opportunities on the path to 6G associated with the integration of RISs and other emerging technologies, such

as NOMA, PLS, SWIPT, UAV-enabled wireless networks, and autonomous driving networks.

#### A. NOMA and RIS

In an effort to improve the SE and user connectivity of RIS-enhanced wireless networks, power-domain NOMA technology is adopted, whose key idea is to superimpose the signals of two users at different powers for exploiting the spectrum more efficiently by opportunistically exploring the users' different channel conditions [66], [159]. Li *et al.* [160] considered a MISO-NOMA downlink communication network for minimizing the total transmit power by jointly designing the transmit precoding vectors and the reflecting coefficient vector. In [8], Yang *et al.* jointly optimized the phase shifts matrix of the RIS, as well as the power allocation from the BS to the users. Thus, the minimum decoding SINR of all users was maximized for optimizing the throughput of the system by considering user fairness. Ding and Poor [75] proposed a novel design of RIS assisted NOMA networks. It can be observed in [75] that, the directions of users' channel vectors can be aligned with the aid of the RIS, which emphasizes the importance of implementing NOMA technology. For an RIS-NOMA system, the core challenge is that the decoding order is dynamically changed due to the configuration of phase shifts of the RIS. Mu *et al.* [60] proposed an RIS-enhanced multiple-antenna NOMA transmission framework to maximize the throughput of the system by considering the NOMA SIC decoding order condition. The SCA technique and sequential rank-one constraint relaxation based algorithm were applied to obtain a locally optimal solution. Ni *et al.* [161] proposed a resource allocation framework in multi-cell RIS-NOMA networks, where the achievable sum rate was maximized by solving the joint optimization problem of user association, sub-channel assignment, power allocation, phase shifts design, and decoding order determination.

In contrast to the conventional MIMO-NOMA systems, RIS-NOMA technology can overcome the challenges of the dynamic environment such as random fluctuation of wireless channels, blocking, and user mobility in an energy-efficient manner. The NOMA system can obtain tuned channel gains, improved fair resource allocation, enhanced coverage range, and high EE with the aid of RISs [162]. However, NOMA also gives rise to new challenges when integrated with RISs. For multi-antenna NOMA transmission, the decoding order is not determined by the users' channel gains order, since additional decoding rate conditions need to be satisfied to guarantee successful SIC [60]. Additionally, both the active beamforming and passive phase shift design affect the decoding order among users and user clustering, which makes the decoding order design, user clustering, and joint beamforming design highly-coupled in RIS-NOMA networks.

#### B. PLS and RIS

It has been shown that RISs are capable of simultaneously enhancing the desired signal power at the intended user and mitigating the interference power at other unintended users [58]. Inspired by this result, several researchers explored the potential performance gain in the context of PLS by applying the RIS [119], [163]–[169]. Yu *et al.* [163]

considered an RIS-enhanced multiple-input single-output single eavesdropper (MISOSE) channel, where the eavesdropper is equipped with a single antenna. The secrecy rate was maximized by jointly optimizing the transmit beamforming and the RIS phase shift matrix by using an AO-based algorithm. It was demonstrated that the secrecy performance can be significantly improved by deploying the RIS. Cui *et al.* [164] focused on the scenario where the eavesdropper has a better direct channel condition than that of the legitimate receiver and they are also highly correlated in space, where the achievable secrecy rate is rather limited in conventional communications. However, it was shown that the direct signals and the reflected signals can be destructively combined at the eavesdropper with the aid of RISs, thus significantly improving the secrecy rate. The same problem was further investigated in [119], [166] by considering a multi-antenna eavesdropper or legitimate receiver. Chu *et al.* [165] minimized the transmit power while satisfying the secrecy rate requirement in the RIS-enhanced MISOSE system. Chen *et al.* [167] studied the minimum secrecy rate maximization problem in the RIS-enhanced multi-user multiple-input single-output multiple eavesdropper (MISOME) system by considering both the continuous and discrete RIS phase shifts. Injecting artificial noise (AN) is an effective technique to enhance the secrecy rate [170]. Motivated by this result, Guan *et al.* [168] examined the effectiveness of employing AN in an RIS-enhanced MISOME system. The achievable secrecy rate was maximized by jointly optimizing the transmit beamforming, the passive beamforming, and AN. The results verified the necessity of using AN, especially for systems with a large number of eavesdroppers. Yu *et al.* [169] considered an RIS-enhanced multi-user MISOME system under imperfect CSI with the aim of maximizing the sum rate, subject to the maximum information leakage constraint. An efficient AO-based algorithm was developed to optimize the transmit beamforming, the AN covariance matrix, and the RIS phase shifts. Numerical results showed that significant secrecy performance gains can be achieved by the RIS.

One critical issue of the RIS-enhanced PLS is that the joint design of transmit and passive beamforming requires the CSI of both AP-eavesdropper and RIS-eavesdropper links, which is quite challenging to obtain. This is because besides the nearly passive working mode of RISs, in practice, eavesdroppers usually stay almost silent to hide their positions and only detect signals in the air. Therefore, robust joint beamforming designs under the imperfect CSI of the eavesdropper are essential to guarantee secure transmission. Moreover, given the uncertainty of eavesdroppers, deploying the RIS may increase the probability of information leakage since the eavesdropper can receive not only the direct signals from the AP but also the reflected signals from the RIS. The situation may become even worse when there are multiple cooperative eavesdroppers. In this case, setting a protected zone to establish an eavesdropper-exclusion area with carefully deployed RISs would help to enhance the secrecy performance, which deserve further investigations.

### C. SWIPT and RIS

SWIPT is an attractive technique for future IoT networks. However, the low EE at the energy receivers is the main bottleneck in practical SWIPT systems. To overcome this limitation, deploying the RIS is a promising solution and the RIS-assisted SWIPT has been investigated in [19]–[22]. In [19], Wu and Zhang investigated an RIS-assisted SWIPT system, subject to individual SINR requirements of information receivers. The weighted sum power received by energy receivers was maximized by jointly optimizing the transmit and passive beamforming with the proposed AO-based algorithm. Moreover, Tang *et al.* [20] maximized the minimum received power among energy receivers. The results in [19] and [20] showed that deploying an RIS can improve the energy harvesting efficiency. Pan *et al.* [21] studied the weighted sum rate maximization problem in the RIS-assisted SWIPT MIMO system, subject to the energy harvesting requirement of each energy receiver. A block coordinate descent (BCD)-based algorithm was designed to find a Karush-Kuhn-Tucker (KKT) stationary point of the original optimization problem. Wu and Zhang [22] extended the RIS-assisted SWIPT system into a multi-RIS case, where the transmit power was minimized while satisfying the different QoS constraints at information users and energy users. It was shown that the RIS enlarges the wireless power transfer range and reduces the number of required energy beams.

Note that the above research contributions studied performance gain of deploying RISs for SWIPT mainly from the communication perspective and ignored the EM characteristic of RISs. As discussed in previous sections, there are substantial differences between the near-field region and the far-field region of RISs. Therefore, sophisticated EM-based wireless power transfer models are required for fully reaping the benefits of RISs, which need to be investigated in future work.

### D. UAV and RIS

RISs can be applied in UAV-enabled wireless networks, where UAVs are employed to complement and/or support the existing terrestrial cellular networks [171], [172]. An RIS enhances the UAV coverage and service quality by compensating for the power loss over long distances, as well as forming virtual LoS links between UAVs and mobile users via passively reflecting their received signals. Li *et al.* [24] jointly optimized the UAV trajectory and the RIS phase shifts in an iterative manner. It was shown in [24] that, the average achievable rates of the users were significantly improved with the aid of RISs. On the other hand, RISs can also be applied in UAV-aided wireless relay networks for enhancing performance. Zhang *et al.* [173] considered the effective placement of a single UAV, which was equipped with an RIS to assist the mmWave downlink transmission while considering user mobility. By jointly designing the UAV trajectory and the RIS reflection parameters, a virtual LoS connection between the BS and users was guaranteed. Thus, both the average data rate and the achievable downlink LoS probability were improved. Yang *et al.* [174] derived the analytical

expressions of outage probability, BER, and average capacity by approximating the PDF of the instantaneous SNR in RIS-assisted UAV relaying systems. Mu *et al.* [175] proposed a novel RIS-aided multi-UAV NOMA transmission framework, where an RIS was deployed to enhance the desired signal strength between UAVs with their served ground users while mitigating the inter-UAV interference. Liu *et al.* [176] integrated UAVs in RIS-enhanced wireless networks for enhancing the service quality of the UAV. With the aid of RISs, the energy consumption of the UAV was significantly reduced.

Due to the fact that UAVs are battery-powered, how to reduce their energy consumption is one of the key challenges. The limited flight-time of UAVs (usually under 30 minutes) hampers the wide commercial roll-out of UAV-aided networking. By deploying RISs, one can adjust the RIS phase shift instead of controlling the UAV movement for forming virtual LoS links between the UAV and the users. Therefore, the UAV can maintain hovering status only when the virtual LoS links can be formed even with the aid of the RIS. By invoking the aforementioned protocol, the total energy consumption of the UAV is minimized, which in turn, maximizes the UAV endurance. Additionally, by mounting a compact distributed laser charging (DLC) receiver or wireless power transmission (WPT) receiver antenna inside the UAVs, while a DLC/WPT transmitter is deployed on the ground or the building roof, the UAVs can be charged as long as they are flying within the coverage range of the DLC/WPT transmitter [131], [177]. However, the LoS connection between UAVs and the charging stations/vehicles have to be guaranteed, which is challenging in the urban scenario when the LoS link between UAVs and charging stations/vehicles are blocked by high-rise buildings with a high probability. RISs are capable of smartly reconfiguring the wireless propagation environment by forming virtual LoS links between UAVs and the charging stations/vehicles via passively reflecting their received signals. Thus, the quality of charging service is enhanced with the aid of the RIS.

The RIS-enhanced UAV communication scenario is naturally a highly dynamic one, which falls into the field of ML. When considering both the trajectory design of UAVs and the phase shift design of RISs, the former one can be formed as a continuous state space while the latter one is usually formed as a discrete one. Hence, how to simultaneously deal with both continuous and discrete state space is challenging in ML-empowered RIS-enhanced UAV networks.

#### E. Autonomous Driving/Connected Vehicles and RIS

RISs can also be deployed in vehicle-to-infrastructure (V2I) assisted autonomous driving systems, where V2I components are employed to complement the costly onboard units (OBUs). V2I networks enable autonomous vehicles (AVs) or connected vehicles (CVs) to receive reliable real-time traffic information from BSs, the information is collected by roadside base stations (RBSs) and transmitted from RBSs to BSs, which facilitates the interaction among AVs/CVs and road users, hence enhancing their safety and traffic efficiency [178]–[181]. Since AVs/CVs quality and reliability are non-negotiable, the AVs/CVs system must be real-time, while the transmission is supposed to be 100% reliable. However, the service quality

of current V2I communication systems cannot be guaranteed due to the complex channel terrain in the urban environment and the complexity of road conditions, such as bad weather. Makarfi *et al.* [182] and Wang *et al.* [183] proved that the performance of vehicular networks can be significantly improved with the aid of RISs. Since the RISs are made of EM material, which can be installed on key surfaces, such as building facades, highway polls, advertising panels, vehicle windows, and even pedestrians' clothes. With the massive deployment of RISs, a virtual LoS connection between the BSs and AVs, as well as between the RSUs and BSs will be guaranteed, which enhances the reliability of V2I communications.

In RIS-enhanced autonomous driving systems, the driving safety of AVs is the primary consideration. In terms of safety, collisions have to be avoided, while the traffic rules also need to obey. Additionally, in RIS-enhanced V2I-assisted autonomous driving systems, the wireless service quality for AVs has to be guaranteed at each timeslot. Hence, how to improve the reliability of RIS-enhanced autonomous driving systems is an open and challenging problem.

#### F. Discussions and Outlook

The studies of RISs have unveiled promising research opportunities, such as NOMA, PLS, SWIPT, UAV-enabled wireless networks, and autonomous driving networks. Recent research contributions have proved that RIS-enhanced wireless networks can achieve tuned channel gains, improved QoS, enhanced coverage range, and reduced energy dissipation. However, the network and beamforming designs are highly coupled due to dynamic control of the RIS phase shifts, which brings challenges to these new research directions.

## VII. CONCLUSIONS, CHALLENGES, AND POTENTIAL SOLUTIONS

#### A. Concluding Remarks

In this paper, recent research works on RIS-enhanced wireless networks proposed for applications to next-generation networks have been surveyed with an emphasis on the following aspects: operating principles of RISs, performance evaluation of multi-antenna assisted RIS systems, joint beamforming design and resource allocation for RISs, ML in RIS-enhanced wireless networks, and their integration with other key 6G technologies. We have highlighted the advantages and limitations of employing RISs for communication applications. Further research efforts are needed to bridge the complex physical models of the different RISs implementations with widely used communication models. We have considered the performance evaluation of multi-antenna assisted RIS systems by systematically surveying existing designs for RIS-enhanced wireless networks from the views of performance analysis, information theory, and optimization. In addition, we have discussed existing research contributions that applying ML tools for tackling the dynamic essence of the wireless environment such as random fluctuations of wireless channels and user mobility. Design guidelines for ML-empowered RIS-enhanced wireless networks have also been discussed. The benefits of

integrating RISs with NOMA, UAV-terrestrial networks, PLS, SWIPT, and AVs/CVs have been discussed. However, the research of RIS-enhanced wireless networks is still at a very early stage and there are ample opportunities for important contributions and advances in this field. Some of them are listed as follows.

### B. Challenges and Potential Solutions

1) *CSI Acquisition:* The acquisition of timely and accurate CSI plays a pivotal role in RIS-enhanced wireless systems, especially in MIMO-RIS and MISO-RIS networks. The majority of current research contributions assume perfect CSI available at the BS, RISs controllers, as well as the users. However, obtaining CSI in RIS-enhanced wireless networks is a non-trivial task, which requires a non-negligible training overhead. Additionally, in RIS-assisted NOMA networks, users in each cluster have to share the CSI with each other for implementing SIC. However, due to the passive characteristic of RISs, the CSI acquisition and exchanging are non-trivial. Potential solutions can be developed by employing DL methods for exploiting CSI structures beyond linear correlations.

2) *Pareto-Optimization for Satisfying Multiple Objectives:* In contrast to the conventional wireless networks, RIS-enhanced wireless networks are characterized by more rapidly fluctuating network topologies and more vulnerable communication links. Furthermore, RISs are more likely to be deployed in an environment with heterogeneous mobility profiles. Hence, the networks operate in a complex time-variant hybrid environment, where the classic mathematical models have limited accuracy. Additionally, the challenging optimization problems encountered in RIS-enhanced wireless networks usually have to satisfy multiple objectives (e.g., delay, throughput, BER, and power) in order to arrive at an attractive solution. To elaborate, by definition it is only possible to improve any of the metrics considered at the cost of degrading at least one of the others. The collection of Pareto-optimal points is referred to as the Pareto front. However, determining the entire Pareto-front of optimal solutions is still challenging. Potential solutions may be investigating near-real-time ML-aided Pareto-optimization for tackling the high-dynamic adaptation of RIS-enhanced wireless networks.

### REFERENCES

- [1] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, “The roadmap to 6G: AI empowered wireless networks,” *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
- [2] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.
- [3] Q. Wu and R. Zhang, “Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network,” *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.
- [4] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, and H. V. Poor, “Downlink and uplink intelligent reflecting surface aided networks: NOMA and OMA,” *IEEE Trans. Wireless Commun.*, early access, Feb. 2, 2021, doi: [10.1109/TWC.2021.3054841](https://doi.org/10.1109/TWC.2021.3054841).
- [5] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, “MIMO assisted networks relying on intelligent reflective surfaces,” 2019. [Online]. Available: arxiv:1910.00959.
- [6] Y.-C. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo, “Large intelligent surface/antennas (LISA): Making reflective radios smart,” *J. Commun. Inf. Netw.*, vol. 4, no. 2, pp. 40–50, 2019.
- [7] M. Di Renzo *et al.*, “Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.
- [8] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, “Intelligent reflecting surface meets OFDM: Protocol design and rate maximization,” *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.
- [9] N. S. Perović, M. Di Renzo, and M. F. Flanagan, “Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmWave environments,” in *Proc. IEEE Int. Commun. Conf. (ICC)*, 2020, pp. 1–7.
- [10] C. Huang *et al.*, “Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends,” *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.
- [11] M. Di Renzo *et al.*, “Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison,” *IEEE Open J. Commun. Soc.*, vol. 1, pp. 798–807, 2020.
- [12] E. Björnson, Ö. Özdogan, and E. G. Larsson, “Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?” *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.
- [13] S. Zhou, W. Xu, K. Wang, M. Di Renzo, and M.-S. Alouini, “Spectral and energy efficiency of IRS-assisted MISO communication with hardware impairments,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1366–1369, Sep. 2020.
- [14] Y. Cao and T. Lv, “Intelligent reflecting surface enhanced resilient design for MEC offloading over millimeter wave links,” 2019. [Online]. Available: arXiv:1912.06361.
- [15] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, “Latency minimization for intelligent reflecting surface aided mobile edge computing,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, May 2020.
- [16] Y. Cao and T. Lv, “Sum rate maximization for reconfigurable intelligent surface assisted device-to-device communications,” 2020. [Online]. Available: arXiv:2001.03344.
- [17] L. Yang, J. Yang, W. Xie, M. O. Hasna, T. Tsiftsis, and M. Di Renzo, “Secrecy performance analysis of RIS-aided wireless communication systems,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12296–12300, Oct. 2020.
- [18] C. Pan *et al.*, “Multicell MIMO communications relying on intelligent reflecting surfaces,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.
- [19] Q. Wu and R. Zhang, “Weighted sum power maximization for intelligent reflecting surface aided SWIPT,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 586–590, May 2020.
- [20] Y. Tang, G. Ma, H. Xie, J. Xu, and X. Han, “Joint transmit and reflective beamforming design for IRS-assisted multiuser MISO SWIPT systems,” in *Proc. IEEE Int. Commun. Conf. (ICC)*, 2020, pp. 1–6.
- [21] C. Pan *et al.*, “Intelligent reflecting surface enhanced MIMO broadcasting for simultaneous wireless information and power transfer,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.
- [22] Q. Wu and R. Zhang, “Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735–1748, Aug. 2020.
- [23] H. Wang *et al.*, “Performance of wireless optical communication with reconfigurable intelligent surfaces and random obstacles,” 2020. [Online]. Available: arxiv:2001.05715.
- [24] S. Li *et al.*, “Reconfigurable intelligent surface assisted UAV communication: Joint trajectory design and passive beamforming,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 716–720, May 2020.
- [25] D. Ma, M. Ding, and M. Hassan, “Enhancing cellular communications for UAVs via intelligent reflective surface,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2020, pp. 1–6.
- [26] A. U. Makarfi, K. M. Rabie, O. Kaiwartya, O. S. Badarneh, X. Li, and R. Kharel, “Reconfigurable intelligent surface enabled IoT networks in generalized fading channels,” in *Proc. IEEE Int. Commun. Conf. (ICC)*, 2020, pp. 1–6.
- [27] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, “Intelligent reflecting surface enhanced indoor robot path planning: A radio map-based approach,” *IEEE Trans. Wireless Commun.*, early access, Mar. 4, 2021, doi: [10.1109/TWC.2021.3062089](https://doi.org/10.1109/TWC.2021.3062089).

- [28] C. Liaskos, S. Nie, A. Tsoliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.
- [29] H. Gacanin and M. Di Renzo, "Wireless 2.0: Toward an intelligent radio environment empowered by reconfigurable meta-surfaces and artificial intelligence," *IEEE Veh. Technol. Mag.*, vol. 15, no. 4, pp. 74–82, Dec. 2020.
- [30] M. Di Renzo *et al.*, "Smart radio environments empowered by AI reconfigurable meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun.*, vol. 2019, no. 1, pp. 1–20, May 2019.
- [31] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.
- [32] S. Gong *et al.*, "Towards smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2283–2314, 4th Quart., 2020.
- [33] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, early access, Jan. 18, 2021, doi: [10.1109/TCOMM.2021.3051897](https://doi.org/10.1109/TCOMM.2021.3051897).
- [34] D. R. Smith, O. Yurduseven, L. P. Mancera, P. Bowen, and N. B. Kundtz, "Analysis of a waveguide-FED metasurface antenna," *Phys. Rev. Appl.*, vol. 8, no. 5, 2017, Art. no. 054048.
- [35] B. O. Zhu *et al.*, "Dynamic control of electromagnetic wave propagation with the equivalent principle inspired tunable metasurface," *Sci. Rep.*, vol. 4, no. 1, pp. 1–7, 2014.
- [36] J. Y. Dai, J. Zhao, Q. Cheng, and T. J. Cui, "Independent control of harmonic amplitudes and phases via a time-domain digital coding metasurface," *Light Sci. Appl.*, vol. 7, no. 1, p. 90, 2018.
- [37] S. R. Rengarajan and Y. Rahmat-Samii, "The field equivalence principle: Illustration of the establishment of the non-intuitive null fields," *IEEE Antennas Propag. Mag.*, vol. 42, no. 4, pp. 122–128, Aug. 2000.
- [38] V. S. Asadchy, M. Albooyeh, S. N. Tsvetkova, A. Díaz-Rubio, Y. Ra'di, and S. Tretyakov, "Perfect control of reflection and refraction using spatially dispersive metasurfaces," *Phys. Rev. B, Condens. Matter*, vol. 94, no. 7, 2016, Art. no. 075142.
- [39] B. H. Fong, J. S. Colburn, J. J. Ottusch, J. L. Visher, and D. F. Sievenpiper, "Scalar and tensor holographic artificial impedance surfaces," *IEEE Trans. Antennas Propag.*, vol. 58, no. 10, pp. 3212–3221, Oct. 2010.
- [40] N. M. Estakhri and A. Alù, "Wave-front transformation with gradient metasurfaces," *Phys. Rev. X*, vol. 6, no. 4, 2016, Art. no. 041008.
- [41] B. O. Zhu, J. Zhao, and Y. Feng, "Active impedance metasurface with full 360 reflection phase tuning," *Sci. Rep.*, vol. 3, p. 3059, Oct. 2013.
- [42] N. K. Emani, A. V. Kildishev, V. M. Shalaev, and A. Boltasseva, "Graphene: A dynamic platform for electrical control of plasmonic resonance," *Nanophotonics*, vol. 4, no. 1, pp. 214–223, 2015.
- [43] G. Gradoni and M. Di Renzo, "End-to-end mutual-coupling-aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances," *IEEE Wireless Commun. Lett.*, early access, Jan. 11, 2021, doi: [10.1109/LWC.2021.3050826](https://doi.org/10.1109/LWC.2021.3050826).
- [44] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, Sep. 2020.
- [45] V. Arun and H. Balakrishnan, "RFocus: Beamforming using thousands of passive antennas," in *Proc. 17th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2020, pp. 1047–1061.
- [46] A. Welkie, L. Shangguan, J. Gummesson, W. Hu, and K. Jamieson, "Programmable radio environments for smart spaces," in *Proc. 16th ACM Workshop Hot Topics Netw.*, 2017, pp. 36–42.
- [47] M. Dunna, C. Zhang, D. Sievenpiper, and D. Bharadia, "ScatterMIMO: Enabling virtual MIMO with smart surfaces," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, 2020, pp. 1–14.
- [48] R. J. Bell, K. R. Armstrong, C. S. Nichols, and R. W. Bradley, "Generalized laws of refraction and reflection," *J. Opt. Soc. America A*, vol. 59, no. 2, pp. 187–189, 1969.
- [49] J. Huang and J. A. Encinar, *Reflectarray Antennas*. Hoboken, NJ, USA: Wiley, 2008.
- [50] F. H. Danufane, M. Di Renzo, J. de Rosny, and S. Tretyakov, "On the path-loss of reconfigurable intelligent surfaces: An approach based on green's theorem applied to vector fields," 2020. [Online]. Available: arxiv:2007.13158.
- [51] R. C. Johnson, H. A. Ecker, and J. S. Hollis, "Determination of far-field antenna patterns from near-field measurements," *Proc. IEEE*, vol. 61, no. 12, pp. 1668–1694, Dec. 1973.
- [52] H.-T. Chen, A. J. Taylor, and N. Yu, "A review of metasurfaces: Physics and applications," *Rep. Progr. Phys.*, vol. 79, no. 7, 2016, Art. no. 076401.
- [53] J. Huang and R. J. Pogorzelski, "A Ka-band microstrip reflectarray with elements having variable rotation angles," *IEEE Trans. Antennas Propag.*, vol. 46, no. 5, pp. 650–656, May 1998.
- [54] J. Xu and Y. Liu, "A novel physics-based channel model for reconfigurable intelligent surface-assisted multi-user communication systems," 2020. [Online]. Available: arxiv:2008.00619.
- [55] X. Qian, M. Di Renzo, J. Liu, A. Kammoun, and M.S. Alouini, "Beamforming through reconfigurable intelligent surfaces in single-user MIMO systems: SNR distribution and scaling laws in the presence of channel fading and phase noise," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 77–81, Jan. 2021.
- [56] Z. Zhang, Y. Cui, F. Yang, and L. Ding, "Analysis and optimization of outage probability in multi-intelligent reflecting surface-assisted systems," 2019. [Online]. Available: arxiv:1909.02193.
- [57] A. A. Boulogeorgos and A. Alexiou, "Ergodic capacity analysis of reconfigurable intelligent surface assisted wireless systems," in *Proc. IEEE 3rd 5G World Forum (5GWF)*, 2020, pp. 395–400.
- [58] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
- [59] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
- [60] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884–6898, Oct. 2020.
- [61] R. Karasik, O. Simeone, M. Di Renzo, and S. S. Shitz, "Beyond max-SNR: Joint encoding for reconfigurable intelligent surfaces," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2020, pp. 2965–2970.
- [62] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Capacity and optimal resource allocation for IRS-assisted multi-user communication systems," *IEEE Trans. Commun.*, early access, Feb. 26, 2021, doi: [10.1109/TCOMM.2021.3062651](https://doi.org/10.1109/TCOMM.2021.3062651).
- [63] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1057–1071, Apr. 2021.
- [64] W. Tang *et al.*, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, Jan. 2021.
- [65] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.
- [66] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.
- [67] W. Yi, Y. Liu, and A. Nallanathan, "Modeling and analysis of D2D millimeter-wave networks with Poisson cluster processes," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5574–5588, Dec. 2017.
- [68] T. Hou, Y. Liu, Z. Song, X. Sun, and Y. Chen, "Exploiting NOMA for UAV communications in large-scale cellular networks," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6897–6911, Oct. 2019.
- [69] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996–1019, 3rd Quart., 2013.
- [70] M. A. Kishk and M. Alouini, "Exploiting randomly-located blockages for large-scale deployment of intelligent surfaces," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1043–1056, Apr. 2021.
- [71] M. Di Renzo and J. Song, "Reflection probability in wireless networks with metasurface-coated environmental objects: An approach based on random spatial processes," *EURASIP J. Wireless Commun.*, vol. 2019, no. 1, pp. 1–15, Apr. 2019.
- [72] C. Zhang, W. Yi, and Y. Liu, "Reconfigurable intelligent surfaces aided multi-cell NOMA networks: A stochastic geometry model," 2020. [Online]. Available: arxiv:2008.08457.

- [73] K. S. Ali, M. Haenggi, H. ElSawy, A. Chaaban, and M.-S. Alouini, "Downlink non-orthogonal multiple access (NOMA) in Poisson networks," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1613–1628, Feb. 2019.
- [74] Z. Ding, R. Schober, and H. V. Poor, "On the impact of phase shifting designs on IRS-NOMA," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1596–1600, Oct. 2020.
- [75] Z. Ding and H. V. Poor, "A simple design of IRS-NOMA transmission," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1119–1123, May 2020.
- [76] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, and G. K. Karagiannidis, "Non-orthogonal multiple access (NOMA) with multiple intelligent reflecting surfaces," 2020. [Online]. Available: arxiv:2011.00211.
- [77] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, "Reconfigurable intelligent surface aided NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2575–2588, Nov. 2020.
- [78] J. Lyu and R. Zhang, "Spatial throughput characterization for intelligent reflecting surface aided multiuser system," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 834–838, Jun. 2020.
- [79] Z. Tang, T. Hou, Y. Liu, J. Zhang, and L. Hanzo, "Physical layer security of intelligent reflective surface aided NOMA networks," 2020. [Online]. Available: arxiv:2011.03417.
- [80] P. Xu, G. Chen, Z. Yang, and M. Di Renzo, "Reconfigurable intelligent surfaces-assisted communications with discrete phase shifts: How many quantization levels are required to achieve full diversity?" *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 358–362, Feb. 2021.
- [81] L. Dai *et al.*, "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results," *IEEE Access*, vol. 8, pp. 45913–45923, 2020.
- [82] T. Hou, Y. Liu, Z. Song, X. Sun, and Y. Chen, "MIMO-NOMA networks relying on reconfigurable intelligent surface: A signal cancellation-based design," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 6932–6944, Nov. 2020.
- [83] K. Ntontin, J. Song, and M. Di Renzo, "Multi-antenna relaying and reconfigurable intelligent surfaces: End-to-end SNR and achievable rate," 2019. [Online]. Available: arxiv:1908.07967.
- [84] H. Zhang, B. Di, L. Song, and Z. Han, "Reconfigurable intelligent surfaces assisted communications with limited phase shifts: How many phase shifts are enough?" *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4498–4502, Apr. 2020.
- [85] J. Yuan, Y. C. Liang, J. Joung, G. Feng, and E. G. Larsson, "Intelligent reflecting surface-assisted cognitive radio system," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 675–687, Jan. 2021.
- [86] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, Nov. 2020.
- [87] W. Shi, X. Zhou, L. Jia, Y. Wu, F. Shu, and J. Wang, "Enhanced secure wireless information and power transfer via intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 25, no. 4, pp. 1084–1088, Apr. 2021.
- [88] B. Lyu, D. T. Hoang, S. Gong, D. Niyato, and D. I. Kim, "IRS-based wireless jamming attacks: When jammers can attack without power," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1663–1667, Oct. 2020.
- [89] S. Zhang and R. Zhang, "Intelligent reflecting surface aided multiple access: Capacity region and deployment strategy," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2020, pp. 1–5.
- [90] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels. I. Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.
- [91] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.
- [92] H. Han, J. Zhao, D. Niyato, M. Di Renzo, and Q. Pham, "Intelligent reflecting surface aided network: Power control for physical-layer broadcasting," in *Proc. IEEE Int. Commun. Conf. (ICC)*, 2020, pp. 1–7.
- [93] M. Fu, Y. Zhou, and Y. Shi, "Reconfigurable intelligent surface empowered downlink non-orthogonal multiple access," *IEEE Trans. Commun.*, early access, Mar. 17, 2021, doi: 10.1109/TCOMM.2021.3066587.
- [94] J. Zhu, Y. Huang, J. Wang, K. Navaie, and Z. Ding, "Power efficient IRS-assisted NOMA," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 900–913, Feb. 2021.
- [95] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753–757, Apr. 2020.
- [96] G. Zhou, C. Pan, H. Ren, K. Wang, M. Di Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, Oct. 2020.
- [97] T. Lipp and S. Boyd, "Variations and extension of the convex-concave procedure," *Optim. Eng.*, vol. 17, no. 2, pp. 263–287, 2016.
- [98] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 5092–5106, Aug. 2020.
- [99] A. Zappone, M. Di Renzo, F. Shams, X. Qian, and M. Debbah, "Overhead-aware design of reconfigurable intelligent surfaces in smart radio environments," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 126–141, Jan. 2021.
- [100] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, 2019, pp. 735–740.
- [101] X. Yu, D. Xu, and R. Schober, "Optimal beamforming for MISO communications via intelligent reflecting surfaces," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2020, pp. 1–5.
- [102] B. Ning, Z. Chen, W. Chen, and J. Fang, "Beamforming optimization for intelligent reflecting surface assisted MIMO: A sum-path-gain maximization approach," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1105–1109, Jul. 2020.
- [103] K. Ying, Z. Gao, S. Lyu, Y. Wu, H. Wang, and M. Alouini, "GMD-based hybrid beamforming for large reconfigurable intelligent surface assisted millimeter-wave massive MIMO," *IEEE Access*, vol. 8, pp. 19530–19539, 2020.
- [104] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.
- [105] C. You, B. Zheng, and R. Zhang, "Intelligent reflecting surface with discrete phase shifts: Channel estimation and passive beamforming," in *Proc. IEEE Int. Commun. Conf. (ICC)*, 2020, pp. 1–6.
- [106] C. Huang, A. Zappone, M. Debbah, and C. Yuen, "Achievable rate maximization by passive intelligent mirrors," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2018, pp. 3714–3718.
- [107] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate optimization for intelligent reflecting surface enhanced wireless networks," 2019. [Online]. Available: arxiv:1905.07920.
- [108] M. Jung, W. Saad, M. Debbah, and C. S. Hong, "On the optimality of reconfigurable intelligent surfaces (RISs): Passive beamforming, modulation, and resource allocation," *IEEE Trans. Wireless Commun.*, early access, Feb. 18, 2021, doi: 10.1109/TWC.2021.3058366.
- [109] M. M. Zhao, Q. Wu, M. J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 2–17, Jan. 2021.
- [110] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 7748–7764, Dec. 2020.
- [111] G. Yang, X. Xu, and Y.-C. Liang, "Intelligent reflecting surface assisted non-orthogonal multiple access," 2019. [Online]. Available: arxiv:1907.03133.
- [112] Y. Yang, S. Zhang, and R. Zhang, "IRS-enhanced OFDMA: Joint resource allocation and passive beamforming optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 760–764, Jun. 2020.
- [113] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7170–7183, Nov. 2020.
- [114] X. Li, J. Fang, F. Gao, and H. Li, "Joint active and passive beamforming for intelligent reflecting surface-assisted massive MIMO systems," 2019. [Online]. Available: arxiv:1912.00728.
- [115] H. Xie, J. Xu, and Y.-F. Liu, "Max-min fairness in IRS-aided multi-cell MISO systems with joint transmit and reflective beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1379–1393, Feb. 2021.
- [116] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Joint deployment and multiple access design for intelligent reflecting surface assisted networks," *IEEE Trans. Wireless Commun.*, early access, May 23, 2020, doi: 10.1109/TWC.2021.3075885.

- [117] X. Liu, M. Chen, Y. Liu, Y. Chen, S. Cui, and L. Hanzo, "Artificial intelligence aided next-generation networks relying on UAVs," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 120–127, Feb. 2021.
- [118] C.-X. Wang, M. Di Renzo, S. Stanczak, S. Wang, and E. G. Larsson, "Artificial intelligence enabled wireless networking for 5G and beyond: Recent advances and future challenges," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 16–23, Feb. 2020.
- [119] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, Sep. 2019.
- [120] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian, "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 60–69, Sep. 2019.
- [121] Z. Qin, H. Ye, G. Y. Li, and B. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 93–99, Apr. 2019.
- [122] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7331–7376, Oct. 2019.
- [123] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.
- [124] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems," 2019. [Online]. Available: arxiv:1912.03619.
- [125] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.
- [126] S. Liu, Z. Gao, J. Zhang, M. Di Renzo, and M. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmWave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, Aug. 2020.
- [127] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-Wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, Sep. 2020.
- [128] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2019, pp. 1–5.
- [129] J. Gao, C. Zhong, X. Chen, H. Lin, and Z. Zhang, "Unsupervised learning for passive beamforming," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1052–1056, May 2020.
- [130] S. Khan and S. Y. Shin, "Deep-learning-aided detection for reconfigurable intelligent surfaces," 2019. [Online]. Available: arxiv:1910.09136.
- [131] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7957–7969, Aug. 2019.
- [132] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double  $Q$ -learning," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 2094–2100.
- [133] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, "Dueling network architectures for deep reinforcement learning," 2015. [Online]. Available: arxiv:1511.06581.
- [134] M. Fortunato *et al.*, "Noisy networks for exploration," 2017. [Online]. Available: arxiv:1706.10295.
- [135] W. Dabney, M. Rowland, M. G. Bellemare, and R. Munos, "Distributional reinforcement learning with quantile regression," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2892–2901.
- [136] V. Mnih *et al.*, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937.
- [137] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017. [Online]. Available: arxiv:1707.06347.
- [138] P. Hämäläinen, A. Babadi, X. Ma, and J. Lehtinen, "PPO-CMA: Proximal policy optimization with covariance matrix adaptation," 2018. [Online]. Available: arxiv:1810.02541.
- [139] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," 2015. [Online]. Available: arxiv:1509.02971.
- [140] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor–critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018. [Online]. Available: arxiv:1801.01290.
- [141] N. C. Luong *et al.*, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 4th Quart., 2019.
- [142] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, Aug. 2020.
- [143] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, May 2020.
- [144] A. Taha, Y. Zhang, F. B. Mismar, and A. Alkhateeb, "Deep reinforcement learning for intelligent reflecting surfaces: Towards standalone operation," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2020, pp. 1–5.
- [145] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021, doi: 10.1109/TWC.2020.3024860.
- [146] Q. Zhang, W. Saad, and M. Bennis, "Millimeter wave communications with an intelligent reflector: Performance optimization and distributional reinforcement learning," 2020. [Online]. Available: arxiv:2002.10572.
- [147] J. Ye, S. Guo, and M.-S. Alouini, "Joint reflecting and precoding designs for SER minimization in reconfigurable intelligent surfaces assisted MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5561–5574, Aug. 2020.
- [148] M. Jung, W. Saad, and G. Kong, "Performance analysis of large intelligent surfaces (LISs): Uplink spectral efficiency and pilot training," 2019. [Online]. Available: arxiv:1904.00453.
- [149] K. Umebayashi, M. Kobayashi, and M. López-Benítez, "Efficient time domain deterministic-stochastic model of spectrum usage," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1518–1527, Mar. 2018.
- [150] Z. Feng, X. Li, Q. Zhang, and W. Li, "Proactive radio resource optimization with margin prediction: A data mining approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9050–9060, Oct. 2017.
- [151] K. G. M. Thilina, E. Hossain, and D. I. Kim, "DCCC-MAC: A dynamic common-control-channel-based MAC protocol for cellular cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3597–3613, May 2016.
- [152] P. Abouzar, K. Shafiee, D. G. Michelson, and V. C. Leung, "Action-based scheduling technique for 802.15.4/ZigBee wireless body area networks," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2011, pp. 2188–2192.
- [153] J. Wang, C. Jiang, H. Zhang, Y. Ren, K.-C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1472–1514, 3rd Quart., 2020.
- [154] X. Liu, Y. Liu, and Y. Chen, "Reinforcement learning in multiple-UAV networks: Deployment and movement design," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8036–8049, Aug. 2019.
- [155] A. Assra, J. Yang, and B. Champagne, "An EM approach for cooperative spectrum sensing in multiantenna CR networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1229–1243, Mar. 2016.
- [156] A. Morell, A. Correa, M. Barceló, and J. L. Vicario, "Data aggregation and principal component analysis in WSNs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3908–3919, Jun. 2016.
- [157] J. Li, H. Zhang, and M. Fan, "Digital self-interference cancellation based on independent component analysis for co-time co-frequency full-duplex communication systems," *IEEE Access*, vol. 5, pp. 10222–10231, 2017.
- [158] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.
- [159] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
- [160] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA intelligent reflecting surface-aided downlink communication networks," 2019. [Online]. Available: arxiv:1909.06972.

- [161] W. Ni, X. Liu, Y. Liu, H. Tian, and Y. Chen, "Resource allocation for multi-cell IRS-aided NOMA networks," *IEEE Trans. Wireless Commun.*, early access, Feb. 11, 2021, doi: 10.1109/TWC.2021.3057232.
- [162] A. S. D. Sena *et al.*, "What role do intelligent reflecting surfaces play in multi-antenna non-orthogonal multiple access?" *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 24–31, Oct. 2020.
- [163] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.
- [164] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.
- [165] Z. Chu, W. Hao, P. Xiao, and J. Shi, "Intelligent reflecting surface aided multi-antenna secure transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 108–112, Jan. 2020.
- [166] L. Dong and H. Wang, "Secure MIMO transmission via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 787–790, Nov. 2020.
- [167] J. Chen, Y. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82599–82612, 2019.
- [168] X. Guan, Q. Wu, and R. Zhang, "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 778–782, Jun. 2020.
- [169] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.
- [170] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, Jun. 2008.
- [171] Q. Wang, Z. Chen, H. Li, and S. Li, "Joint power and trajectory design for physical-layer secrecy in the UAV-aided mobile relaying system," *IEEE Access*, vol. 6, pp. 62849–62855, 2018.
- [172] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: The vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.
- [173] Q. Zhang, W. Saad, and M. Bennis, "Reflections in the sky: Millimeter wave communication with UAV-carried intelligent reflectors," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.
- [174] L. Yang, F. Meng, J. Zhang, M. O. Hasna, and M. Di Renzo, "On the performance of RIS-assisted dual-hop UAV communication systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12296–12300, Sep. 2020.
- [175] X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent reflecting surface enhanced multi-UAV NOMA networks," 2021. [Online]. Available: arxiv:2101.09145.
- [176] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, early access, Dec. 2, 2020, doi: 10.1109/JSAC.2020.3041401.
- [177] Q. Liu *et al.*, "Charging unplugged: Will distributed laser charging for mobile wireless power transfer work?" *IEEE Veh. Technol. Mag.*, vol. 11, no. 4, pp. 36–45, Dec. 2016.
- [178] L. Yao, J. Wang, X. Wang, A. Chen, and Y. Wang, "V2X routing in a VANET based on the hidden Markov model," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 889–899, Mar. 2018.
- [179] R. P. D. Vivacqua, M. Bertozi, P. Cerri, F. N. Martins, and R. F. Vassallo, "Self-localization based on visual lane marking maps: An accurate low-cost approach for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 582–597, Feb. 2018.
- [180] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Enhancing the fuel-economy of V2I-assisted autonomous driving: A reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8329–8342, Aug. 2020.
- [181] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Information-theoretic model predictive control: Theory and applications to autonomous driving," *IEEE Trans. Robot.*, vol. 34, no. 6, pp. 1603–1622, Dec. 2018.
- [182] A. Makarf *et al.*, "Reconfigurable intelligent surfaces-enabled vehicular networks: A physical layer security perspective," 2020. [Online]. Available: arxiv:2004.11288.
- [183] J. Wang, W. H. Zhang, X. Bao, T. Song, and C. Pan, "Outage analysis for intelligent reflecting surface assisted vehicular communication networks," 2020. [Online]. Available: arxiv:2005.00996.



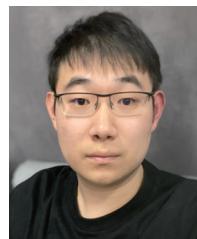
**Yuanwei Liu** (Senior Member, IEEE) received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications in 2011 and 2014, respectively, and the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2016.

He was with the Department of Informatics, King's College London from 2016 to 2017, where he was a Postdoctoral Research Fellow. He has been a Lecturer (Assistant Professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London since 2017. His research interests include non-orthogonal multiple access, 5G/6G networks, machine learning, and stochastic geometry. He received IEEE ComSoc Outstanding Young Researcher Award for EMEA in 2020. He received the 2020 Early Achievement Award of the IEEE ComSoc Signal Processing and Computing for Communications (SPCC) Technical Committee. He received the Exemplary Reviewer Certificate of IEEE WIRELESS COMMUNICATIONS LETTERS in 2015, IEEE TRANSACTIONS ON COMMUNICATIONS in 2016 and 2017, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2017 and 2018. He has served as the Publicity Co-Chair for VTC 2019-Fall. He is the leading contributor for "Best Readings for Non-Orthogonal Multiple Access" and the primary contributor for "Best Readings for Reconfigurable Intelligent Surfaces." He serves as the Chair of Special Interest Group (SIG) in SPCC Technical Committee on the topic of signal processing Techniques for next generation multiple access, the vice-chair of SIG Wireless Communications Technical Committee on the topic of Reconfigurable Intelligent Surfaces for Smart Radio Environments, and the Tutorials and Invited Presentations Officer for Reconfigurable Intelligent Surfaces Emerging Technology Initiative. He currently an Editor on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON COMMUNICATIONS, and IEEE COMMUNICATIONS LETTERS. He serves as the leading Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on Next Generation Multiple Access, a Guest Editor for IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING special issue on Signal Processing Advances for Non-Orthogonal Multiple Access in Next Generation Wireless Networks. He has served as a TPC Member for many IEEE conferences, such as GLOBECOM and ICC. See for more info <http://www.eecs.qmul.ac.uk/~yuanwei>.



**Xiao Liu** (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Communication Systems Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London.

His research interests include unmanned aerial vehicle aided networks, machine learning, non-orthogonal multiple access techniques, and reconfigurable intelligent surface. He received the Exemplary Reviewer Certificate of IEEE TRANSACTIONS ON COMMUNICATIONS in 2019 and 2020.



**Xidong Mu** (Graduate Student Member, IEEE) received the B.S. degree in information engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015, where he is currently pursuing the Ph.D. degree.

From September 2020, he is a visiting student (online enrolment) with the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. His research interests include non-orthogonal multiple access, IRSs/RISs aided communications, UAV communications, and optimization theory. He received the Exemplary Reviewer Certificate of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2020.



**Tianwei Hou** received the M.S. and Ph.D. degrees in information and communication engineering from Beijing Jiaotong University.

He has been an Associate Professor with the School of Electronic and Information Engineering, Beijing Jiaotong University since 2021. His research interests include UAV communications, non-orthogonal multiple access, reconfigurable intelligent surface and Internet of Things. He received the Exemplary Reviewer Certificate of IEEE TRANSACTIONS ON COMMUNICATIONS in 2019 and the IEEE COMMUNICATIONS LETTERS in 2018 and 2019.

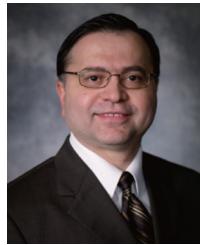


**Marco Di Renzo** (Fellow, IEEE) received the Laurea (*cum laude*) and Ph.D. degrees in electrical engineering from the University of L'Aquila, Italy, in 2003 and 2007, respectively, and the Habilitation à Diriger des Recherches (Doctor of Science) degree from University Paris-Sud, France, in 2013. Since 2010, he has been with the French National Center for Scientific Research (CNRS), where he is a CNRS Research Director (CNRS Professor) of the Laboratory of Signals and Systems, Paris-Saclay University–CNRS and

CentraleSupélec, Paris, France. In Paris-Saclay University, he serves as the Coordinator of the communications and networks research area with the Laboratory of Excellence DigiCosme and as a Member of the Admission and Evaluation Committee of the Ph.D. School on Information and Communication Technologies. He has received several individual distinctions and research awards, which include the IEEE Communications Society Best Young Researcher Award for Europe, Middle East and Africa, the Royal Academy of Engineering Distinguished Visiting Fellowship, the IEEE Jack Neubauer Memorial Best System Paper Award, the IEEE Communications Society Young Professional in Academia Award, the SEE-IEEE Alain Glavieux Award, and a 2019 IEEE ICC Best Paper Award. In 2019, he was a recipient of a Nokia Foundation Visiting Professorship for conducting research on metamaterial-assisted wireless communications at Aalto University, Finland. He currently serves as the Editor-in-Chief of IEEE COMMUNICATIONS LETTERS, and is a Distinguished Speaker of the IEEE Vehicular Technology Society. From 2017 to 2020, he was a Distinguished Lecturer of the IEEE Vehicular Technology Society and IEEE Communications Society. He served as an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and an Associate Editor-in-Chief of IEEE COMMUNICATIONS LETTERS. He currently serves as the Founding Chair of the Special Interest Group on Reconfigurable Intelligent Surfaces of the Wireless Technical Committee of the IEEE Communications Society. He is the Founding Lead Editor of the IEEE Communications Society Best Readings in Reconfigurable Intelligent Surfaces. He is a Highly Cited Researcher (Clarivate Analytics, 2019), a World's Top 2% Scientist from Stanford University (2020), and a Fellow of the IET (2020).



**Jiaqi Xu** (Student Member, IEEE) received the B.S. degree from Peking University, China, in 2019. He is currently pursuing the Ph.D. degree with the Queen Mary University of London, U.K. His research interests include modeling, design, and performance analysis of wireless communication systems.



**Naofal Al-Dhahir** (Fellow, IEEE) received the Ph.D. degree from Stanford University, where he was a Principal Member of Technical Staff with GE Research and AT&T Shannon Laboratory from 1994 to 2003. He is an Erik Jonsson Distinguished Professor and an ECE Department Associate Head with UT-Dallas. He is co-inventor of 43 issued patents, coauthored about 450 papers and co-recipient of four IEEE best paper awards. He is received 2019 IEEE SPCE Technical Recognition Award, and served as the Editor-in-Chief of IEEE TRANSACTIONS ON COMMUNICATIONS from January 2016 to December 2019.