Reconfigurable Intelligent Surface for B5G & 6G

未来移动通信论坛
FuTURE MOBILE COMMUNICATION FORUM

Abstract

In the emerging study of 6G technology, Reconfigurable Intelligent Surface (RIS) is a new type of artificial electromagnetic surface that can intelligently and dynamically control the electromagnetic characteristics of each unit of the electromagnetic surface through digital programming. It will realize accurate and efficient control of electromagnetic waves and change the electromagnetic wave propagation environment, bringing a new 6G airspace transmission paradigm to the future.

This white paper introduces RIS from multiple perspectives such as application scenarios, requirements, key technologies and algorithms, current research status, field verification, and proposes suggestions on industry maturity, development direction, and standardization research direction.

Contents

Chapter I Overview

1.1. Concept

About every ten years, there is a new generation of cellular mobile communication systems, in which new key technologies are introduced to improve service quality. The 5th-generation mobile communication (5G) aim to provide mobile communication infrastructures in a higher standard. It supports multiple scenarios, such as Enhanced Mobile Broadband (eMBB), Massive Machine Type Communications (mMTC), and Ultra-reliable and Low-latency Communications (uRLLC). In the meantime, mobile communication network development is changed toward software-defined development. That is, software is used to configure and optimize the network in real time. However, there are many randomnesses and uncertainties in the wireless environment, bringing lots of uncontrollable factors to mobile communication networks. As the research on 6th-generation mobile communication (6G) commences, 6G is considered to provide full-coverage, full-spectrum, all-application, and strong-security services to meet people's various increasing communication requirements. 6G will provide greater capacity, ultra-low latency, high reliability, high security, and full-space coverage. To explore and break through the restriction that there are many uncontrollable factors in the wireless environment and reshape the wireless transmission environment is the new direction for 6G development.

In the last two years, the reprogrammable metasurface technology has drawn great attention in the field of mobile communication. The reprogrammable metasurface technology was first proposed and verified in lab in 2014 by Cui Tiejun, Academician of CAE in Southeast University [1]. Figure 1-1 shows its basic structure. The reprogrammable metasurface adopts a two-dimensional thin artificial electromagnetic surface structure with reprogrammable electromagnetic features, and it can be used for various frequency bands from microwaves to visible light [2]. The reprogrammable metasurface is composed of delicately designed and regularly arranged electromagnetic units. Such units are usually made of metal, medium, and adjustable elements. The adjustable elements in the electromagnetic units allow you to modify the electromagnetic parameter settings related to reflected electromagnetic waves through reprogramming. Such parameters include the phase and amplitude. This technology connects the physical electromagnetic world of the metasuface and the digital world of information technology [3], and is particularly attractive to mobile communication applications.

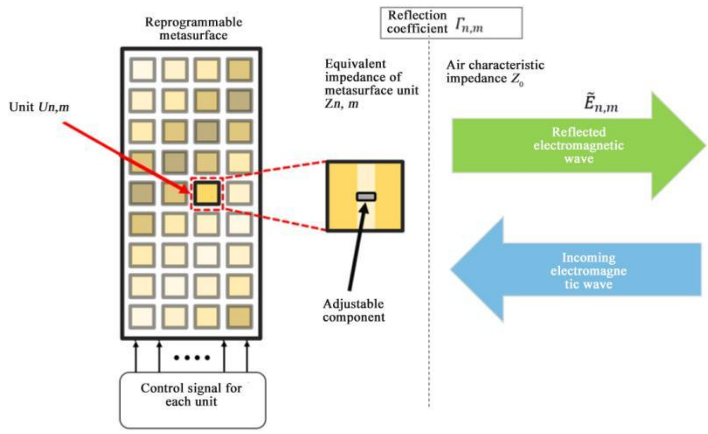


Figure 1-1 Reprogrammable Metasurface

The Reconfigurable Intelligent Surface (RIS) is composed of and enabled by the reprogrammable metasurface, and can be used to adjust the electromagnetic signals of mobile communication in real time. The primary application for RIS is the RIS-based wireless relay. The RIS allows people to adjust the wireless channel environment to significantly improve the transmission performance between communication devices. Wireless signals from the transmitting terminal to the receiving terminal may be attenuated and scattered to a certain degree due to the absorption of objects and natural signal diffusion in the space. This increases the computational complexity and reduces the performance for signal recovery on the receiving terminal. In traditional mobile communication systems, people cannot control the wireless environment. The only thing that can be done is to create channel characteristic models through massive channel sounding and delicately design algorithms for the transceivers. However, with RIS in the communication system, flexible control of the electromagnetic transmission can proactively improve the wireless transmission environment. RIS-based wireless relays can implement redirection and beamforming for electromagnetic signals in the wireless channels to utilize wireless signal energy more efficiently and improve the system performance.

1.2. Potential Application Scenarios and Requirements

Currently, the potential application for RIS is the relay of wireless transmission. It adjusts the phase and/or amplitude of electromagnetic waves to form desired reflected beams, enhancing transmission.

When the RIS is operating as a relay, it improves the wireless signal transmission through reflection/transparent transmission. The adjustments made by the electromagnetic components in the RIS can change the transmission direction of the reflected or transmitted electromagnetic waves on the RIS to form the desired beam pattern. Specifically, the application scenarios for RIS to be adopted as a relay are as follows:

1) Coverage Holes Filling

In wireless cellular networks, there can be regional weak coverage areas or coverage holes. They may occupy small areas, but will hamper user experience. Besides, the mmWave frequency band used for 5G is vulnerable to the blockage of obstacles, causing deteriorated signal quality. Moving vehicles, pedestrians, growing vegetation, and leaves can be potential obstacles.

RIS can resolve all of the above coverage issues. Deploy the RIS at proper locations (such as the surface of a wall) can artificially build a line of sight (LOS) link, improving the weak coverage and wireless signal transmission robustness. See Figure 1-2.

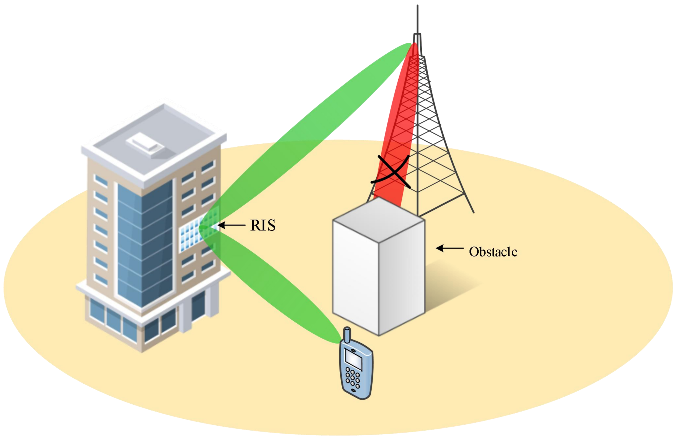


Figure 1-2 Coverage Hole Filling by RIS

2) Regional Traffic Increase

For hotspot areas, RIS can reduce the multi-path effect for wireless channels and increase the number of channel subspace, to increase the traffic and expand capacity. Unlike coverage hole filling, when RIS is used for regional traffic increase, the base station must be able to implement channel estimation for the cascading channels between the RIS and the UE. Therefore, it is more complicated to design the transmission scheme.

3) Indoor Coverage Enhancement

For the scenarios of outdoor base stations covering indoor areas, the transparent RIS material can be used to design a transparent glass film that does not block light and reduces loss of wireless signals when they penetrate glass, to improve indoor network coverage quality. For common weak coverage scenarios such as the stairway corner and the deep sunken area of stadium stand, deploy the RIS at proper locations such as the walls of the stairway corner or stadium. A LOS link from the base station to the weak coverage area can be planned based on the requirement to improve the coverage of weak coverage and wireless signal transmission robustness.

1.3. Current Research at Home and Abroad

Massive low-cost electromagnetic intelligent control units of the RIS can control the reflections of wireless signals, in order to rearrange the wireless transmission environment. Therefore, this technology is an emerging technology that is being widely discussed. It has great potential to improve the transmission rate, coverage area, and energy efficiency of mobile systems. Through adjustments of the reflection phase shift of the RIS electromagnetic unit for wireless channels, the signals transmitted through RIS reflections and signals transmitted through other paths can be superposed in a co-phase way, to enhance the receive signal quality. Compared with traditional relays, RIS-based wireless relays can implement full-duplex transmissions without introducing interference. In addition, the RIS features lightweight, low cost, and low power consumption, and has great potential.

In [4], a single-cell wireless communication system with the RIS-assisted relay is studied, in which a RIS-assisted multi-antenna access point communicates with multiple single-antenna UEs.

Joint optimization is implemented for the transmitted beams from the active antenna arrays of the access point and the reflected beams of the RIS, to minimize the total transmit power of the access point while maintaining a specified signal-to-noise ratio (SNR) of receive signals on the UE. Significantly different from communications with passive signal scattering, RIS-based wireless relays are mainly used to enhance the current communication link performance, but not to transmit information of the relay itself through reflection. In the RIS-assisted communication, the signals transmitted in the direct path and the signals that are reflected carry the same information, and therefore the coherent signals can be superposed at the receiver to enhance the transmission and maximize the receiver power. In [5], the RIS is used for the downlink of a multi-antenna base station serving multiple UEs, and a power optimization model is proposed for the RIS-assisted communication system. This model utilizes the number of RIS-based reflection units and their phase control capabilities. Under maximum power and minimum service quality, it maximizes the energy efficiency to optimize phase shift distribution of the RIS electromagnetic units and downlink transmit power.

In a mobile communication system with RIS-assisted relay, the channel status information (CSI) is critical for the passive beamforming of RIS. In [6], the performance of an RIS-assisted large-scale antenna system in different propagation scenarios is evaluated by formulating a tight upper bound of the ergodic spectral efficiency. It is demonstrated that the ergodic spectral efficiency is related to the reflected phase shift distribution of RIS. In addition, considering a hardware non-ideal factor, an optimal phase shift design based on the upper bound of the ergodic spectral efficiency and statistical channel state information is proposed to maximize the ergodic spectral efficiency under the condition that the phase quantization bits of RIS are limited. In [7], the beamforming optimization is studied for the RIS-assisted mobile communication system under discrete phase shift. Assume that the electromagnetic units can assist the communication between the multi-antenna access point and the single-antenna UE with limited phase shift status. A solution that is closest to the optimal and least complex is proposed based on the alternative optimization technology. That is, joint optimization is implemented for the continuously transmitted beams of the access point and the discretely reflected beams from the RIS, to minimize the transmit power for the access point while maintaining a certain SNR on the UE. The simulation shows that RIS with discrete phase shifts enables the same performance as that of the continuous phase shift when there are massive reflective electromagnetic units. In [4], [5], [6], and [7], the reflective phase shift design is implemented for the RIS electromagnetic units under the assumption that the perfect CSI is known. This assumption helps us understand the upper bound of the system performance.

There is also research on how the RIS can improve the security of mobile communication. In [8], the adaptive adjustment of the reflective phase shift of the RIS can enhance the desired signals and suppress the undesired signals. Joint design of the beamformings for the transmitted signals of the access point and the passively reflected signals on the RIS can maximize the encryption of legitimate communication links. In [9], how the RIS enhances physical-layer encryption of the communication is explored. In a Multi-Input Single-Output (MISO) broadcast system, the base station transmits independent data flows to multiple legitimate receivers, and encryption is conducted against multiple eavesdroppers. Joint optimization is implemented for the beamformer of the base station and the reflected phase shift distribution of the RIS. Under actual constraints, the minimum encryption is maximized, and the path tracing algorithm based on alternative optimization reduces the computational complex.

In the preceding publications, theories have been used to study the RIS-assisted relay's mobile communication system, and multiple optimization algorithms have been proposed to improve system performance. In fact, there have been a few early studies on the system realizations of the RIS-based wireless relay.

In [10], an RIS-based wireless relay named RFocus moves the beamforming function from wireless terminals to the wireless environment. In typical indoor scenarios, every reflective electromagnetic unit of RFocus is configured by software controllers to maximize receive signal power of the receiver. In theoretical analysis and actual measurement, RFocus can improve signal intensity by 10.5 times and channel capacity by 2 times in average. In [11], there are experiments proving good robustness of the RFocus in case of electromagnetic units' failures. The relative performance improvement will not drop to 0 but will drop by 50% even though one third of these units fail. In [10] and [11], the feasibility of the RIS applied into actual communications systems is practically verified.

In [12], a large array composed of 36 low-cost antenna units is deployed in an indoor household environment to adjust the wireless environment, with a channel decomposition algorithm designed to quickly estimate the wireless channel environment, and the phase distribution of the large array is configured in real time to align phases of multiple sub-channels. This system realizes flexible reprogrammable wireless channels. This experiment indicates that reconfiguring wireless environment can improve the system throughput by 24% in average. In addition, the Shannon capacity is improved by 51.4% compared with that of the baseline single-antenna links, and by 12.23% to 18.95% compared with that of the baseline antenna links. In [13], an RIS-based scattering MIMO system is proposed to improve the spatial multiplexing gains by enhancing the scattering effect with the low-cost RIS. It pairs with active access points to create virtual passive access points. In the experiment, the configuration is optimized to enable virtual access points to provide signals with the same power of real active access points and improve the coverage range of each single access point. On the other hand, algorithms of low computational complexity are designed to optimize the RIS for the scattering MIMO for each UE. The RIS-based scattering MIMO system has reduced interference, and the power requirement for distributed MIMO systems is reduced. In the experiment, the RIS increases the average throughput by 2 times in an existing commercial MIMO-Wi-Fi network after deployment in the system.

In [14], the concept of smart space is proposed. In the space, the wireless environment is reprogrammable to have required link quality in the wireless space. Low-cost devices are embedded into the wall of a building to improve the wireless link quality by passively or actively reflecting wireless signals. The experiment has proved that it is feasible to use passive elements to change the wireless channels. The 2×2 MIMO channel matrix condition number is increased by 1.5 dB, and the signal intensity is improved by 26 dB. The RIS, a new technology that can flexibly control the incoming electromagnetic waves, can be deployed on the surface of the scatterer of a large space to intelligently control the wireless environment, enhance the coverage of 5G-advanced networks, and implement transformation of the communication paradigm. The RIS reflection mode is an important application direction for the RIS, and many institutions in China have carried out research on the RIS in the form of relays.

Southeast University has built an measurement system for the RIS free space path loss (FSPL) to verify the theoretical model of the RIS FSPL. The system can measure the FSPL of multiple types of RISs to verify the theoretical formula for the RIS FSPL.

In addition, a series of live network tests have been carried out, producing lots of inspiring results. China Mobile, together with the Cui Tiejun academician team of Southeast University and Hangzhou Qiantang Information Co., Ltd., has completed the technical verification of the RIS on live networks in Nanjing [15]. The preliminary test result shows that the RIS can flexibly adjust the signal beams in the wireless environment based on the UE distribution, to significantly improve the signal intensity, network capacity, and user rate in weak coverage areas of the live network. In the outdoor test scenarios, the cell edge coverage is improved by 3 to 4 dB in average, and the cell edge user throughput is improved by about over 10 times. In the test scenarios that outdoor base stations cover indoor areas, the indoor coverage is improved by about 10 dB, and the user throughput is improved by about 2 times.

ZTE, together with China Unicom, has completed the technical verification test of the RIS reflective panel in 5G intermediate frequency (IF) network outfield in Shanghai [16]. The test result shows that at the non-line-of-sight (NLOS) cell edge of the 5G IF base stations, the reference signal received power (RSRP) of 5G UEs can reach 10 dB, and the performance of 5G cell edge users can be improved by over 40%. Therefore, the RIS reflection technology will be a scientific and feasible innovative approach for the in-depth coverage of 5G IF base station networks.

ZTE, together with China Telecom, has completed the far-distance technical verification test of the RIS reflective panel in 5G high frequency network outfield in Shanghai [17]. The test result shows that the reference signal receive intensity of 5G UEs is improved by 12.5 dB, and the 5G high frequency weak coverage UE performance in improved by 296% for NLOS coverage holes or week coverage over 150 meters away from 5G high frequency (26 GHz frequency band) base stations. The RIS reflection technology will be a scientific and feasible innovative approach for the in-depth coverage of 5G high frequency base station networks.

In addition, ZTE, together with China Mobile Beijing Branch, has published the RIS cascading technology prototype verification results for the commercial networks in 2.6 GHz. Two-level cascaded RIS panels are used to reflect 5G signals in controlled circumstance, to actively explore the feasibility of RIS in 5G networks [18].

The early test results preliminarily show the feasibility of the RIS, but there are four challenges for its standardization and actual engineering application. Challenge one: The basic theory is not complete. The reflection and transparent transmission characteristics of the RIS are to be determined, the channel transmission model is incomplete, and there is no modeling for actual transmission environment. Challenge two: The key technology needs to be created. There are things to be studied and standardized, such as co-channel and adjacent-frequency interference characteristics, interference coordination among operators, and beamforming and channel estimation algorithms. Challenge three: The component maturity and reliability are not sufficient. The current RIS in the industry is only a prototype, with insufficient adjustable angle and component adjustment speed, and there are massive units for quickly locating and identifying unit failures. Challenge four: The deployment scenarios are limited. The RIS has a large size and a large surface, and active and wired control will reduce its application scenarios. Further optimized engineering design can improve its deployment flexibility.

The intelligent reflector is of relatively high technological and industrial maturity, and three-phase development can be adopted for the intelligent reflector. Phase one: To realize the passive static reflection surface that can be used for quick deployment and satisfies the needs for expanding network coverage and filing coverage holes; Phase two: To realize the semi-static controllable reflector. Adjustment and control on component units implements beam selection to expand the coverage of the metasurface beams and improve the cell capacity and rate; Phase three: To realize the dynamic intelligent reflector. Coding algorithms dynamically track UE location and match the channel environment, to intelligently adjust and control the transmissions of electromagnetic waves.

References

[1]. C T J, Q M Q, W X, et al. Coding metamaterials, digital metamaterials and programmable metamaterials[J]. Light Science & Applications, 2014(3): e218.

[2]. Zhang Lei, Liu Shuo, and Cui Tiejun, Theory and application of coding metamaterials [J], Chinese Optics, 2017,10(1): 1-12.

[3]. C T J, L S, Z L. Information metamaterials and metasurfaces[J]. Journal of Materials Chemistry C, 2017(5): 3644-3668.

[4]. W Q, Z R. Itelligent reflecting surface enhanced wireless network via joint active and passive beamforming[J]. IEEE Transactions on Wireless Communications, 2019,18(11): 5394-5409.

[5]. H C, Z A, A G C, et al. Reconfigurable intelligent surfaces for energy efficiency in wireless communication[J]. IEEE Transactions on Wireless Communications, 2019,18(8): 4157-4170.

[6]. H Y, T W K, J S, et al. Large intelligent surface assed wireless communication exploiting statistical CSI[J]. IEEE Transactions on Vehicular Technology, 2019,68(8): 8238-8242.

[7]. W Q, Z R. Beamforming optimization for intelligent reflecting surface with discrete phase shifts[C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, U.K., 2019: 7830-7833.

[8]. 8C M, Z G, Z R. Secure wireless communication via intelligent reflecting surface[J]. IEEE Wireless communications letters, 2019,8(5): 1410-1414.

[9]. 9C J, L Y C, P Y, et al. Intelligent reflecting surface: A programmable wireless environment for physical layer security[J]. IEEE Access, 2019(7): 82599-82612.

[10]. V A, H B. RFocus: Practical beamforming for small devices[J]. arXiv preprint arXiv: 1905.05130, 2019.

[11]. V A, H B. RFocus: Beamforming using thousands of passive antennas[C]//NSDI ’20. Santa Clara, CA, USA, 2020.

[12]. L Z Q, X Y X, S L F. Towards programming the radio environment with large arrays of inexpensive antennas[C]//NSDI ’19. Boston, MA, USA, 2019.

[13]. M D, Z C, D S, et al. Scatter MIMO: Enabling virtual MIMO with smart surfaces[C]//MobiCom ’20. London, U.K., 2020.

[14]. A W, S L F, J G, et al. Programmable radio environments for smart spaces[C]//HotNets-XVI. Palo Alto, CA, USA, 2017.

[15]. Interdisciplinary innovation. China Mobile, together with the Cui Tiejun academician team, has completed the first technical experiment of the RIS [E].https://www.c114.com.cn/news/118/a1168348.html. 2021.

[16]. China Mobile, together with China Telecom, has completed the first technical verification of the RIS under the 5G IF network outfield

[17]. [E].https://www.c114.com.cn/news/127/a1167167.html. 2021.

[18]. ZTE, together with China Telecom, has completed the first technical verification of the RIS under the 5G high-frequency outfield [E].https://www.c114.com.cn/news/127/a1167281.html. 2021.

[19]. Beijing 5G pilot city innovation summit, discussing 5G+ smart city construction hot topics [E].http://www.xinhuanet.com/tech/20210927/5ab4b56cffcd4ea0a3be906868b024ec/c.html. 2021.

Chapter II Key Technologies and Algorithms of RIS

2.1. RIS Structure Design and Adjustment

The RIS is a two-dimensional electromagnetic meta-material based on theories such as the traditional periodic structure electromagnetic theory, Huygens-Fresnel principle, and generalized reflection/refraction principles. With such theories and principles, a specially designed thin PCB structure can be used to have full control of the amplitude, phase, polarization, spread, and momentum of electromagnetic waves. The electromagnetic metasurface can have all the characteristics of electromagnetic meta-materials. In addition, it features multiple functions, simple structure, easy-to-integrate, flexible manufacturing, and low cost, and therefore is widely applied in the electromagnetic wave control field.

With the same metasurface, the adjustable material or component can be controlled to implement multiple functions, greatly expanding the application scenarios of the metasurface and reducing the application cost of the metasurface. In order to enable the metasurface to be applicable to more scenarios (such as applications in information transmission), high-speed control circuits can be introduced to the adjustable metasurface for quickly switching functions of the metasurface. In the meantime, artificial atoms or artificial units with powerful functions can be introduced to the metasurface to implement specific functions, enabling evolution from adjustable metasurface to intelligent metasurface. Thus, features that can be realized for metasurface are determined by: ① Coding strategy (unit control strategy); ② Unit structural characteristics (size and shape); ③ Characteristics of adjustable materials or components; ④ Unit layout in the space; ⑤ Control circuit characteristics.

The metasurface can be used to construct many new functions that are primarily done by the coding strategy, such as full adjustment and control of the amplitude, phase, polarization, frequency spectrum, and momentum of electromagnetic waves. The unit structure design of metasurface can be relatively fixed, to cover typical scenarios with only several types of unit structures of metasurface. Peripheral control circuits control adjustable materials or adjustable components to implement the coding strategy. However, introducing active components and attached control circuits on the metasurface will cause certain impact on the design, usage and original characteristics of the metasurface. Therefore, joint optimization and design should be implemented on the peripheral control circuit, and active components or adjustable materials during new metasurface design. The following sections will analyze the basic performance of the adjustable components and adjustable metasurface, in order to analyze their application scenarios in wireless communications systems.

The excellent performance of the passive metasurface is determined by its unit structure and layout (periodic or non-periodic structure). However, once the metasurface is made, its functions are fixed. The non-adjustable characteristics limit its application in the wireless communications system. Introducing switching diodes or other adjustable materials to the passive metasurface units and using the different operating statuses of the adjustable materials or components to control the metasurface unit resistance can help control the electromagnetic wave characteristics, thus developing adjustable metasurface and intelligent metasurface. Therefore, the operating performance of these adjustable materials or components will have vital impact on functions that can be realized for the metasurface.

Adjustable materials or components used on the metasurface have certain characteristics. For example, such substances can have drastic or rapid physical property changes under external environmental stimulation. When there are changes in the external electronic field, magnetic field, temperature, humidity, pressure, or light intensity, the properties of such materials will change accordingly, thus causing changes to the functions of the metasurface and realizing different modulation mechanisms. Examples: The nematic liquid crystals and graphene in the adjustable electrical structure, the ferrite magnetic bar or ferrite magnetic plate in the adjustable magnetic structure, the Si, GaAs, or semiconductor optoelectronic materials in the adjustable optical structure, and phase shift material, phase shift materials (PCM) such as VO2 in the adjustable thermal structure can be used. Since the current wireless communications system is based on the electrical control mechanism, the adjustable electrical structure and the corresponding metasurface should be preferentially considered.

Adjustable electrical materials or adjustable electrical components have excellent performance and can be flexibly controlled, and therefore are widely used in metasurface unit and metasurface design. Future wireless communications systems use the metasurface to realize more new functions, such as to flexibly control electromagnetic environment, change channel quality, suppress inter-UE interference, converging transmission energy, and simplifying system architecture. In the meantime, the future wireless communications system will possibly work on different frequency bands and multiple frequency bands, or on ultra-large bandwidth. Therefore, select the proper metasurface to satisfy different needs.

2.2. Channel Measurement and Modeling

The channel characteristics of the RIS will determine the performance limit and optimization solution of the RIS communications system. Therefore, it is important to measure the channels of the RIS and build an easy-to-use and sufficiently accurate channel model for the RIS.

The RIS wireless communication system has different path losses in different scenario applications. Common expressions for the FSPL of the RIS-assisted path are not simple enough.

In the RIS-assisted far-field beamforming scenarios, the FSPL of the reflection-assisted path is proportional to (d1*d2)^2, where d1 and d2 are the distance between the base station and the RIS and the that between the base station and the UE, respectively. In the meantime, the beamforming gain is proportional to the square of the number of electromagnetic units. In the RIS-assisted near-field signal broadcast scenario, the overall FSPL of the reflection-assisted path is proportional to (d1+d2)^2. That is, the FSPL is proportional to the square of the sum of the two transmission paths, and there is no beamforming gain in this scenario. The RIS FPSL model needs to be further simplified and optimized in the future.

Currently, large-scale attenuation characteristics (path loss and shadow attenuation), fast attenuation characteristics (multi-path effect, delay expansion and angle expansion, Doppler spectrum, and relativity), spatial consistency, and so on are considered in wireless channel measurement and modeling. The commonly used channel models include the deterministic model, random model, or mixed channel model. The deterministic models include the geometric optical model and ray tracing model. The scenario-map-based model and the point cloud model are common simplified ray tracing models. In recent years, the random channel model that can be classified into GBCM and CBCM has been widely used in academia and industry. GBCM and the digital-map-based hybrid channel model are the channel models that are mainly adopted in the current 3GPP and ITU standard research. The following must be further considered for the RIS channel measurement and modeling based on the current wireless channel models:

(1) Physical model abstraction of RIS

To implement model abstraction for different RIS physical implementation solutions, the following needs to be considered:

① RIS unit model

Polarization model (single/dual polarization, polarization leakage/twist, and anisotropy)

Amplitude/phase adjustment and control model

Insertion loss model

Non-ideal factors

② RIS panel model

BS-RIS far-field plane wave feeding excitation model

BS-RIS near-field spherical wave feeding excitation model

RIS-UE near-field model

RIS codebook model

(2) RIS channel cascading

Selection and optimization of the RIS channel cascading link

For the merge of absolute-latency-based RIS cascading links, considering the balance between simplicity and usability and between complications and accuracy is the key for creating an effective channel model.

In recent preliminary researches, people are attempting to explain the path loss model of the RIS. In [1-2], the physical and electromagnetic characteristics of the RIS are considered, and FSPL models of RIS-assisted wireless communications are created for different scenarios. The measurement in the lab has further proved the proposed path loss mode. The measurement results are well matched with the modeling results. In [3], the general scalar diffraction theory and Huygens-Fresnel principle are used to propose the closed expression for calculating the RIS receive power and to determine the conditions for the RIS to serve as the anomalous specular reflection. In [4], a physical and electromagnetic compatible communications model is introduced for the RIS-assisted wireless system. The model features end-to-end, electromagnetic compatibility, mutual-coupling sensing, and unit amplitude and phase coupling. This model is compatible with the traditional communications theory frame.

For the multi-path attenuation (small-scale attenuation), the typical method for modeling estimation is to use the well-known distribution (such as Rayleigh attenuation and Rician attenuation). In [5], a controllable smart electromagnetic surface reflection model based on ray tracing is proposed in addition to the map-based hybrid channel model, to create a channel modeling method for the smart electromagnetic surface to be deployed in complex scenarios. In this method, multiple virtual logic base stations are assumed to exist on the electromagnetic surface, in order to simplify the model algorithm and reduce calculations while maintaining the high accuracy.

2.3. Channel Estimation and Feedback

The RIS, one of the candidate key technologies for 6G, can be applied in the electromagnetic active wireless environment of physical surfaces. The electromagnetic characteristics of the arrays on the metasurface are altered through intelligent methods to form an electric field with controllable polarization, amplitude, and phase, where the energy is concentrated in a designated three-dimensional space for transmission and reception. This increases the energy efficiency and reduces interference, to implement electromagnetic environment sensing, communications, and control in a completely new way. The RIS can control the scattering, reflection, and refraction of radio waves, to neutralize the negative effect of multi-path fading. Without complicated coding and decoding processes and RF processing, directional reflection can be implemented for the incoming electromagnetic waves.

In traditional communications systems, the channel conditions are completely determined by the environment, and optimization can only be carried out on the Tx/Rx nodes. However, the RIS enables controllable channel conditions, and therefore joint optimization can be implemented on the Tx/Rx nodes and channels, to flexibly control the transmission of electromagnetic waves. The RIS is of positive significance for solving NLOS-related problems, increasing coverage range, expanding transmission freedom, reducing electromagnetic pollution, and enabling ultra-large-scale terminal access, environment sensing, and positioning, and it provides certain support for the unification of future communications. It is necessary to obtain accurate CSI to assist in realizing the preceding functions. This section analyzes the related information of scenario requirements, algorithm design, and engineering implementation of the RIS channel estimation and feedback.

The RIS can operate on different frequency bands. The channel characteristics vary under different frequencies. Generally, the low-frequency bands have strong penetration but severe multi-path effect, the mmWave channels feature sparsity and have a significantly shorter transmission distance than that of the low-frequency bands, and terahertz channels feature abundant scattering paths. For terahertz channels, reflections are no longer ample, and the energy is prone to molecular absorption. In channel estimation algorithm designing, the channel characteristics under different frequency bands, such as sparsity, need to be considered and fully utilized, and the channel estimation method needs to be designed according to the specific frequency bands.

The RIS is greatly different from the traditional MIMO, relays, and other technologies, especially in channel presentation and estimation. The antenna spacing generally equals a half wavelength for the traditional MIMO, which is usually used for far-field situations. However, the antenna spacing can be less than a half wavelength for the RIS. When there are massive RIS elements and the distance between the base station and the RIS is small, the distances from the base station to different metasurface arrays do not need to be the same. In this situation, the near-field model of the channel needs to be analyzed. Traditional MIMOs acting as signal transmitters or receivers do not affect the electromagnetic wave propagation environment, while the RIS may change the electromagnetic field of the propagation environment, bringing a nonlinear impact on the electromagnetic wave propagation. In relay transmission, the segmental channel CSI must be obtained, while the RIS imposes different requirements for the channel CSI in different transmission scenarios, and the cascading channel CSI obtained is sufficient to support communication in some transmission scenarios.

The RIS channel estimation and feedback also has many challenges. The RIS features passive reflection, and does not need massive RF links to be configured, but embraces a large number of arrays, making it difficult to obtain segmental channels. Compared with MIMOs and relays configured with strong signal processing capabilities, general RISes are only equipped with simple onboard signal processing units. For effective channel estimation, new algorithms and protocols must be designed to simplify the RIS channel estimation to the greatest extent and avoid complex onboard signal processing operations. Reasonable channel models are essential for the analysis and research of RIS communication. The current dual-polarization backscattering channel model, spatial scattering channel model, large-scale path loss analysis, and other studies are still in need, while the near-/far-field channels of the RIS have different propagation properties. The introduction of the RIS affects the electromagnetic field, posing new challenges to the characterization and simplification of RIS channels.

For typical application scenarios of the RIS, the BS and the RIS remain fixed in their places. Channels between these two are high-dimensional, varying slowly and quasistatic, while those between UEs with the BS and RIS are time-varying and low-dimensional because users are moving. Channels in these two cases are estimated separately. Although the mobile channels for UEs still need to be frequently estimated, those with large dimensions, between the BS and the RIS, no longer need to be estimated frequently, so the average pilot overhead is reduced.

To estimate the channels between the BS and the RIS under the limitation that the RIS cannot receive or transmit signals, round-trip pilot Rx/Tx methods were mentioned in some references. In such cases, the BS must support the duplex, transmit downlink pilots to the RIS, and receive uplink pilots reflected. BS-RIS channels can be estimated based on the received and transmitted pilots. After BS-RIS channels are estimated, the only thing that needs to do is to transmit pilots through UEs within each hourly scale, so as to estimate the channels from the BS to UEs and those from the RIS to UEs. This helps to complete the RIS-UE channel estimation with a relatively small pilot overhead.

In high-frequency scenarios, transmission signals attenuate more severely, the path signals after multiple reflections are quite weak and can generally be ignored, and only a small number of paths can reach the receiver, so the high-frequency channels are generally considered to be sparse. Cascading channels are estimated by building a combined sparse matrix and designing matrix completion issues considering that the RIS is programmable and that the channels are rank-deficient. In this solution, the RIS is fully passive, and the channel estimation is based on the combination of bilinear sparse matrix decomposition and matrix completion, so that cascading channels can be estimated accurately. However, this method featuring high computational complexity can only obtain cascading channels (but not decompose segmental channels), so it does not apply to the designs and scenarios where the segmental channel CSI must be known.

For high-frequency communication scenarios such as millimeter wave and terahertz, the pilot overhead is reduced based on the sparsity of multi-UE channels in the angle domain. On the one hand, as all UEs communicate with the BS through the same RIS, all non-zero elements of the multi-UE angle-domain cascading channel are in the same row. On the other hand, since all UEs share some scatterers between the RIS and UEs, some non-zero elements of the multi-UE angle-domain cascading channel are in the same column. This can be called structured sparsity. Based on this nature, a support set detection algorithm based on dual-structured sparsity was introduced in [6]. The multi-UE joint estimation is employed to get the common row where the non-zero elements of the angle-domain cascading channels are located and the common column where the non-zero elements of the angle-domain cascading channel are located. It is possible to estimate each UE's non-common columns where the non-zero elements of the UE's angle-domain cascading channels are located. Because the multi-UE joint estimation employed can suppress noise to a certain extent, the estimation accuracy is improved under the same low pilot overhead.

In order to make full use of information about collected data or to solve the channel estimation issues due to unknown channel models, a potential idea is proposed to use new AI methods for channel estimation. Among them, the data-driven channel estimation has less stringent requirements for model accuracy and does not even need to know the transmission model. In recent years, the rapidly developing AI provides new processing paradigms for traditional wireless communication, bringing new solutions to the channel estimation of the RIS wireless communication. This type of method creates databases through offline training, and, during the online training stage, just needs to obtain signals received by pilots to output the channel estimation results, with high robustness and anti-noise capability. However, existing AI-based solutions generally use data information directly for channel estimation, but have seldom dived into data + model dual-driven strategies that may simplify the training process and achieve better training results.

The channel estimation and beam matching collaboration solutions may be designed to adapt to more practical application scenarios. The codebook-based channel estimation solutions may be used for the RIS channel estimation. The characteristics of the RIS segmental transmission challenge the traditional codebook-based solutions. It is urgent to find out how to design the codebooks of BS-RIS and RIS-UE channels and match them well. In addition, codebook-based channel estimation and transmission solutions face more issues in algorithm and protocol design, such as random access design, beam selection method, etc.

In RIS-assisted communication systems, the CSI acquisition is the prerequisite for joint precoding between the BS and the RIS. However, since the RIS is all composed of passive and near-passive units, it cannot actively receive/transmit or process pilot signals, posing a serious challenge to the channel estimation. In addition, the massive RIS units lead to a huge pilot overhead required for channel estimation. A low-overhead channel estimation based on compressed sensing was introduced in [6] according to a hybrid active/passive RIS structure, and the estimation accuracy was improved using the Denoising Convolutional Neural Network (DnCNN). Specifically, the hybrid RIS structure means that only a few RIS units are active and can receive/transmit or process signals and most RIS units are still passive, so as to ensure low costs and low power consumption. In the channel estimation stage, a few RIS units can receive pilot signals, and thanks to the channel sparsity in the angle time delay domain, the compressed sensing algorithm may be employed to recover high-dimensional channels from a small number of received pilots. In order to further improve the estimation accuracy, [6] denoised the channels estimated by the compressed sensing algorithm in the DnCNN, thus realizing the accurate channel estimation with a low pilot overhead. However, the introduction of active units increased the RIS cost and power consumption. In order to achieve the channel estimation under passive RISes, most traditional algorithms directly estimate cascading channels from the BS to UEs via the RIS, so that the RIS channel issues can be converted into channel estimation in a large-scale MIMO system. The AI-based channel estimation in traditional MIMO systems can also be transplanted to the RIS channel estimation.

However, it may be difficult to model accurately RIS channels for different application scenarios, and only measured channel data can be collected. In such cases, the traditional channel estimation cannot work as expected. To this end, we can generate deep learning technologies to estimate channels. For example, based on the measured channel data, we can first use generative adversarial networks (GANs) to learn the channel distribution (including the channels from the BS to the RIS and those from the RIS to UEs), and then in the channel estimation stage, by optimizing the GAN inputs, output the estimate of the current channel to maximize the correlation between the received pilot signals and the channel estimation. More efforts are needed to explore channel estimation based on generation models in future studies, because it is essential for performance comparison and unified evaluation of different algorithms. The performance indicators currently considered include channel estimation accuracy, BS/RIS/terminal computational complexity, time delay, RS, signaling overhead, etc.

2.3.1. Analysis on difficulties of channel estimation

To design a RIS-assisted communication system, obtaining CSI in the RIS wireless communication system is an essential issue to be tackled. In wireless communication systems collaborated by traditional active devices (such as repeaters), the CSI can be estimated based on the training sequence sent by active devices. However, in the RIS collaborative-based communication systems, the passive device RIS composed of a large number of passive reflection units does not have the capability of active Tx/Rx and signal processing. Therefore, the CSI containing massive unknown parameters is to be estimated in the BS or UEs when they are in the active state, which without doubt brings tremendous difficulties and challenges to channel estimation.

Most of the previous efforts assumed that all individual channels between the RIS with the BS and UEs had ideal CSI, and such CSI was known to the BS and RIS. The RIS-based channel estimation and feedback have the following challenges:

- The RIS can only passively reflect signals and has no signal transmission/processing capability, RF links, or sensing and signal amplification capability. In fact, it is difficult to estimate its channels with the BS and UEs.
- The RIS is usually composed of a large number of reflecting elements. The traditional "one-off" channel estimation, i.e. computing the cascading channels of all RIS reflecting elements at one time, requires a long pilot length, and such length increases as the reflecting elements augment, resulting in excessive pilot overheads. In addition, due to massive RIS units, high-dimensional channel matrix, complex channel estimation, and high feedback overhead,
- Channel interoperability may not necessarily be established, making it difficult to estimate the channel uplink and downlink.

- The system performance is sensitive to the accuracy of channel estimation, i.e. the BS and RIS beamforming are also affected.

2.3.2. RIS-based channel estimation and feedback algorithm

1) Bistatic reflection-based channel estimation and feedback

First of all, the entire channel estimation time is divided into several substages [7]. In the first substage, all reflection units in the RIS are closed, and the BS only needs to estimate the direct channel from the BS to UEs. Therefore, the wireless communication system based on RIS collaboration can be simplified into a traditional wireless communication system without RIS. The direct channel can be estimated through classic solutions, such as the least square (LS) or Minimum Mean Square Error (MMSE). In the following sub-stages, each individual reflection unit of the RIS is opened in turn and the remaining are kept closed. The BS estimates the cascading channels from the BS to reflection units and those from reflection units to UEs. Finally, the channel estimation of the entire system is completed based on the estimation results of all substages through the LS/MMSE.

2) MMSE-based unbiased channel estimation and feedback

All reflection units of the RIS remain active throughout the training stage. First of all, the entire cascading channel estimation time is also divided into several substages. In each substage, the optimal phase shift matrix of the RIS is a discrete Fourier transform (DFT) matrix [8]. Therefore, the cascading channel estimation is based on all pilot signals received in all substages.

However, both of the above methods result in significant pilot overhead. Because the unknown channel coefficient in the wireless communication system based on RIS coordination is much larger than that of the conventional wireless communication system, the huge pilot overhead definitely impairs the system performance. Therefore, in the wireless communication system based on RIS collaboration, the channel estimation is quite necessary to reduce the pilot overhead.

3) Multi-time scale-based channel estimation and feedback

After being deployed, the RIS remains static relative to the BS. It can be considered that the RIS-BS channel is quasistatic, with a long coherence time. Due to user mobility, UE-BS and UE-RIS channels are dynamic and fast-changing, with a short coherence time. Therefore, the channel estimation with multi-dimensional time scales can be set for different channels.

BS-RIS channels are quasistatic, but not highly time-varying, with a relatively long channel estimation period. UE-BS and UE-RIS channels are dynamic, with a shorter estimation period. This solution can effectively reduce the pilot overhead, and for these two types of time scales, different settings may be configured according to different RIS deployment scenarios to adapt more flexibly to complicated and ever-changing environments.

In [9], the authors introduced a two-timescale channel estimation framework. The key idea of the framework was based on the high-dimensional and quasistatic nature of BS-RIS channels and the mobile and low-dimensional nature of RIS-UE channels. In order to estimate the quasistatic transmitter-RIS channel, they proposed a dual-link pilot transmission solution, i.e. the transmitter sends downlink pilots and receives uplink pilots reflected from the RIS. In addition, they also introduced a BS-RIS channel recovery algorithm based on coordinate transformation. The numerical results showed that the proposed two-timescale channel estimation framework could achieve accurate channel estimation with lower pilot overhead.

4) Compressed sensing-based channel estimation and feedback

The BS and RIS are generally installed at the same level. Under HF communication conditions, with limited scatterers around the BS and RIS, the angle-domain channel appears to be sparse to some degree. In order to reduce the pilot overhead of the RIS channel estimation, the DFT dictionary helps to convert the space-domain channel to the angle domain and model the channel estimation issues into a sparse signal recovery issue, so that the compressed sensing algorithm can be employed to reduce the pilot overhead. The compressed sensing effectively reduces the pilot overhead, whether it is single-structured sparse for segmental channels or dual-structured sparse for cascading channels.

Reference [10] proposed a method of channel estimation and joint beamforming design for RIS-assisted mmWave systems. The inherent sparsity of the mmWave channel was considered to reduce the training overhead. First of all, the sparse representation of cascading channels was given, and then a channel estimation method based on compressed sensing was proposed to work out the joint beamforming design according to the estimated channels.

5) AI-based channel estimation and feedback

Compared with traditional channel estimation algorithms, AI-based channel estimation can obtain highly accurate CSI with just a small number of dynamic RIS units. Moreover, AI does not need any RIS geometric knowledge, nor does it take into account the channel sparsity, so the solution is more universal. By creating databases through offline training, and obtaining a small number of pilot signals in the online stage, the channel estimation can output results with the anti-noise capability superior to those of traditional channel estimation algorithms. In order to achieve these gains, AI needs to collect sufficient data sets. The current AI-based channel estimation and feedback mainly rely on strong data drivers to obtain reliable gains, while the studies on dual-driven (data-driven and model-driven) models that can simplify the training process are quite limited.

6) Block training-based channel estimation and feedback

The channel estimation for block training is also named the channel estimation for layered training. It divides RIS elements into multiple groups and estimates the cascading channels of all elements in each group. This solution greatly reduces the pilot overhead because the number of pilot symbols required is only related to the number of groups rather than all RIS elements. However, this may undermine the beamforming accuracy as the reflection coefficients are the same for all RIS elements in each group, resulting in lower freedom of the RIS beamforming.

In order to compensate for the beamforming accuracy loss, [11] proposed to break down each group of RIS elements further into smaller subgroups. Reflection coefficient vectors of all RIS units over a given pilot symbol period are decomposed into Kronecker products of two vectors, called base reflection vector and intragroup reflection vector respectively. Intragroup reflection vectors are common to all groups; while base reflection vectors are designed to effectively estimate the valid channels of each group. This method takes into account the beamforming accuracy and the complexity of channel estimation as well.

7) Location-aided channel estimation and feedback

Generally, once deployed, the RIS remains unchanged, so the RIS location information can be applied in the RIS-BS channel estimation. The relatively fixed location between the RIS and the BS helps to obtain information such as the angle of arrival (AOA) through low-complexity channel estimation algorithms. In the cascading channel estimation, the RIS location information is regarded as the known information for channel estimation, to obtain additional gains.

8) Matrix theory-based channel estimation and feedback

First, the sparse matrix is employed to export the channel matrices between the BS and the RIS and between the RIS and UEs based on the received signals. Second, the on/off state matrix containing all RIS reflection units helps to clear ambiguity matrices decomposed from the above matrices Third, the whole RIS reflection unit undergoes the channel estimation finally by recovering missing items based on the matrix characteristics.

2.3.3. Future Research Direction of RIS Channel Estimation and Feedback

1) Channel estimation of multi-cell communication system

The current RIS-based channel estimation is mostly based on the channel estimation of single and multiple UEs in single-cell scenarios, while there are few studies on the single- and multi-UE channel estimation in multi-cell scenarios. Considering the inter-cell interference exists among multiple cells, it is a big challenge to overcome such interference. Existing models seldom consider the multi-RIS cases. The RIS has the advantage that it may have the wireless environment under control after large-scale deployment. The future RIS deployment will be definitely diversified and multi-dimensional, so the channel estimation in multi-RIS scenarios will be a top topic for future research.

2) Channel estimation in highly dynamic scenarios

RISes are usually deployed in fixed locations such as buildings and street lights. For some special applications, RISes must be mounted on dynamic locations, such as unmanned aerial vehicles (UAVs), high-speed railways (HSRs), etc. In these dynamic scenarios, RIS-BS channels become dynamic, while RIS-US channels are quasistatic. As the RIS-BS channel matrix is high-dimensional and dynamic, the channel estimation becomes more difficult, and the challenge gets bigger. It should be explored that whether the existing sparsity-based compressed sensing algorithms and block estimation algorithms are applicable to dynamic channels. Future researches will also focus on solving the RIS-based channel estimation in dynamic scenarios.

2.4. Beamforming Design

The RIS can adjust the beamforming to transmit signals toward a specific direction, so as to increase the power of the required signals and weaken the interference simultaneously. Analysis results from numerous references showed that by increasing the number of RIS reflecting elements, the interruption and capacity of the communication system could be significantly improved. The current massive beamforming design efforts are based on the instantaneous CSI of perfect/imperfect cascading channels (i.e., channels between the BS - each RIS reflecting element - the BS). In order to obtain these instantaneous CSI, a series of estimation methods have been proposed in the academic community. However, in the case of an extremely large number of RIS reflecting elements, the instantaneous CSI estimation overhead of cascading channels is tremendous, seriously reducing the frequency spectrum efficiency of the system. Considering that the statistical CSI changes much more slowly than the instantaneous CSI, the RIS beamforming design based on statistical CSI is beneficial to improving the frequency spectrum efficiency of the system.

It can be seen from the components of the RIS system that the RF processing units and intelligent controller units at the BS and receiver are crucial for intelligent communication and transmission. In particular, the deployment of large-scale antenna arrays and the use of cheap RF units will greatly reduce the system overhead. However, it is impractical that the hardware system is always working in a perfect and stable state due to the environmental noise, I/Q imbalance, phase noise (PN), amplifier nonlinearity, working point drifting of relevant core components (e.g. low-precision A/D converters), etc. in actual physical communication scenarios. Although some measures such as transmitter signal correction and receiver signal compensation can relieve the impact of hardware impairments on the system, the impact of residual hardware impairments or hardware limitations on the system cannot be ignored, otherwise, the beamforming algorithm in the ideal hardware state would seriously distort the training pilots and the signals expected to receive.

The RIS beamforming design with hardware limitations is an essential research direction in this field because it can achieve communication transmission stability and performance optimization from a practical perspective. Affected by actual factors such as transmitter PN and channel estimation errors, reflector PN, receiver distorting noise, etc., traditional methods for beamforming optimization and design of RIS systems are not applicable because these actual factors will also undermine the performance of traditional algorithms. It is quite necessary to consider the impact of hardware impairment in advance from the perspective of system modeling, but taking these factors into consideration will exacerbate the complexity and difficulty in optimizing the original parameter coupling. Therefore, the key to improving the RIS system performance is to design a joint beamforming and phase optimization algorithm that is less complex, reliable, and analytic.

The RIS configured with active electromagnetic units has certain signal detection capabilities. Due to the limited digital processing capabilities of the RIS, conventional pilot training methods cannot be directly employed to estimate RIS channels. In actual application systems, generally, the transmitting terminal can only obtain some imperfect CSI. To solve this issue, robust active Tx/Rx beams and RIS passive beams must be jointly designed to improve the system performance under imperfect CSI conditions.

As the demands of wireless communication users increase, signal transmission requires higher bandwidth, and the frequency band required for wireless communication needs to be increased to obtain sufficient bandwidth resources. However, when the frequency band increases, the corresponding path loss increases, and then a larger antenna array is needed to provide higher beamforming gains. Since each antenna in the array has to be connected to a phase shifter, the power consumption and cost of the phase shifter will get unaffordable as the array size increases. Considering that the RIS features low cost, low power consumption, and intelligent phase control, it is possible to use the RIS auxiliary base stations for joint beamforming.

Because most of the existing RISes can only achieve low-precision phase shift, there are non-convex limitations on joint beamforming between the BS and the RIS, and a highly complex exhaustive search algorithm is required to obtain the optimal solution. For example, in the case of 1-bit quantification, with 64 antennas, the search space would be up to 264, which is not available in practice. As the RIS phase shift is low-bit quantitative, the phase shift of each unit is quantified in a fixed set, that is, the quantification of low-bit quantitative phase shift for each unit can be regarded as a classification issue. Given the excellent performance of machine learning algorithms in solving non-convex classification issues, we can use the deep learning methods and input the channel matrix to predict the classification results of the phase shift for each unit of the RIS. After offline training, the performance close to the best practices can eventually be achieved with low complexity.

Compared with large-scale MIMOs, RISes usually have more units and larger channel dimensions, and thus face greater difficulty in obtaining channel information. Especially in mobile scenarios, users' movement causes channels to change rapidly as time varies and requires repeated beam training to obtain accurate channel information, which will result in huge pilot overheads and seriously limit the actual deployment of the RIS. For this issue, beam training assisted by beam tracking can be employed to update the channel information quickly at a low overhead according to the past channel information and the time-varying rules of channels.

Time-varying channels are affected by many environmental factors, such as users' movement speed, movement direction, and whether the link is blocked. These factors are usually difficult to model, so unsupervised reinforcement learning can be selected, and dynamic environmental characteristics can be adaptively captured using historical beam training information, to significantly reduce pilot overheads. Specifically, the RIS beam tracking process is first modeled as a Markov Decision Process (MDP). Assuming that the optimal beam at the previous moment has been obtained, the action space can be defined as the distance between the optimal beam at this moment and the optimal beam at the previous moment within the beam domain. The reward is defined as the combined rate reachable by the optimal beam predicted based on the current action. Based on Q-learning strategies, online learning is selected specifically for environmental information, to intelligently change the beam selection strategy, and greatly reduce pilot overheads under time-varying channels of the RIS.

The RIS consists of numerous controllable elements. Each of them can independently change the phase of the incoming wave and reflect it back to the environment. However, with the introduction of the RIS containing numerous controllable components, the communication system becomes more complex. The major challenge is the absence of closed-form solutions to the preceded vector at the BS and the precoded matrix at the RIS in the RIS-assisted MIMO system. Therefore, many convex optimization methods are employed to iteratively optimize the precoded vector and matrix, and fixed-design signal detection methods are used at clients. On this basis, the entire system is optimized by separately optimizing individual modules, but this does not guarantee the global optimization of the entire system. To solve this issue, the RIS-assisted communication system can be jointly optimized based on the concept of joint optimization of end-to-end communication systems.

The RIS passively reflects signals to assist the communication between the BS and UEs. The RIS reflection coefficient is regulated by the RIS controller from the BS. Therefore, the precoding at the BS side is called the active beamforming, while the reflected beamforming at the RIS side is named the passive beamforming. The key to improving the RIS system performance is the design of a combined active and passive beamforming. Usually, such joint optimization is non-convex and involves highly coupled variables, bringing some challenges to the solutions.

RIS parameters are always challenged by the following factors:

- Due to tremendous RIS units, numerous configured parameters, and high computational complexity, plus the fact that the RIS itself has no computing power, the computing loads of the BS or the edge server are increased.
- Current optimization algorithms are mostly based on perfect instantaneous CSI. Obtaining these CSI leads to an extremely large overhead for estimating cascading channels. Therefore, it is necessary to consider some optimization algorithms with some CSI or statistical CSI.
- As the reflector phase is coupled to channels, if both parameters are optimized simultaneously, there are great difficulties in achieving the joint optimization design of active and passive beamforming under certain optimization objectives and constraints.

Therefore, different algorithms are proposed to solve issues for different target functions and application scenarios. This section summarizes several optimization algorithms for the parameter design of the RIS.

2.4.1. Common Parameter Optimization Algorithms

1) SDR

At present, the RIS modeling at the communication layer is usually designed as an N-dimensional diagonal matrix, represented by $\Theta = diag(\beta_1 e^{j\theta_1},...,\beta_N e^{j\theta_N},)$. N indicates the number of reflection units in the RIS. The nth diagonal element in the matrix represents the amplitude and phase shift of the nth reflection unit in the RIS. Most of the current references consider the ideal phase shift, i.e., $\beta_n = 1, \theta_n \in [0,2\pi]$, which means each RIS unit has a unit-modulus constraint (UMC) that is non-convex. The SDR is a typical algorithm for UMC processing and is widely used in precoding research of the RIS.

The general process based on the SDR algorithm is to turn optimization issues into a convex semidefinite programming (SDP) issue to be solved by convex optimization solvers, such as CVX. Generally, however, the solution to the issue is not that to Rank 1, so it should be restored by Gaussian randomization [12]. Currently, studies have shown that this algorithm can obtain approximately the optimal solution and ensure high-quality solutions within the polynomial time. However, due to tremendous units in the RIS, this method is too complex to obtain desirable solutions in practice.

2) MM

MM represents "Majorize-Minimization" or "Minorize-Maximization", depending on whether the required optimization is minimization or maximization. The Majorize-Minimization is to obtain the minimum value of the upper bound functions of each target function at each iteration. The Minorize-Maximization is to get the maximum value of the lower bound functions of each target function at each iteration.

The MM algorithm is an iterative optimization method that uses the function convexity to obtain the maximum or minimum value. When it is difficult to optimize target function f(θ), the algorithm does not directly solve the optimal solution of the target function. Instead, it seeks a series of easy-to-optimize substitutes, substitute function g(θ), for target function f(θ) and then solves substitute function g(θ). The optimal solution of substitute function g(θ) is close to that of f(θ). After each iteration, a new substitute function for the next iteration is created based on the solution. The new substitute function is then optimized to find the solution for the next iteration. The solution increasingly closer to the optimal solution of the target function is obtained after several iterations.

The MM algorithm can reach a desirable compromise between the computing amount of each iteration and the total number of iterations. The algorithm has the characteristic of monotonous improvement. For example, in Majorize-Minimization, it will monotonously reduce the value of the true target after each interaction, which means that the target value is convergent.

3) Quantitation

Under the assumption of limited phase shift, quantitation relaxes each phase shift variable to a continuous one. Each of the continuous variables resulting from relaxation will be quantified to the discrete value closest to it. However, quantitation may degrade system performance and on non-convex unit-modulus constraint will survive the continuous relaxation.

4) Deep learning

The latest studies optimized RIS parameters with the assistance of deep learning. [13] developed a DRL-based algorithm to obtain a joint design by observing the predefined rewards and interacting with the environment through trial and error under continuous states and actions, instead of using complex mathematical formulas and numerical optimization techniques.

If the direct transmission between a base station and a user is completely blocked, the join design produced in research on the transmission beamforming of the base station and RIS reflectivity matrix can be used as the output of a DRL neural network to maximize the utilization of the DRL multi-user downlink MISO system and its speed. Specifically, the policy-based deep deterministic policy gradient (DDPG) is used to tackle the CW formation matrix and phase shift.

2.4.2. Application Scenarios

1) Single-user transmission

Issues in RIS single-user transmission are typically non-convex and hard to find the optimal solution. To represent the theoretical performance gain brought by RIS, [12] assumed that the CSI of all channels in the AP was known and addressed the unit-modulus constraint with SDR. Nevertheless, this approach can only generate an approximate solution, which has no guarantee of being optimal. Prior RIS research has mainly hypothesized an ideal phase-shift model in which a unified reflection amplitude is used without considering the phase shift of individual elements. However, when the reflection amplitude depends on phase shift, such a reflection design is generally no longer optimal and can lead to performance degradation. [14] considered a practical phase-shift model in which the amplitude and phase shift cannot be separately adjusted, compounding the problem. For this reason, alternative optimization (AO) and punishment-based optimization techniques were employed to efficiently find the second-best solution. In 2020, Yu et al. proposed a branch-and-bound (BnB) algorithm to address the major issues in beamformer optimization[15]. This first globally optimal algorithm developed by a RIS-assisted MISO system in the literature ensured the acquisition of the optimal solution in single-user circumstances and, though with high complexity, could address discrete phase shift. In addition, approximate optimal solutions were obtained with a manifold optimization-based algorithm taking the proposed BnB algorithm as the baseline, especially in large-scale wireless systems. For RIS-assisted point-to-point MISO communication systems, [16] put forward two new algorithms to optimize AP beamforming and RIS phase shift, with the unit-modulus constraint addressed respectively by fixed-point iteration and the multi-optimization technique. Different from the results of [12], both proposed algorithms ensure the partial optimal solution of the beamformer at AP and RIS phase shifts and are superior to the most advanced SDR approaches in spectral efficiency and computational complexity.

2) Multi-user transmission

Compared with single-user transmission, multi-user systems significantly improve the SNR performance of all users in a network by jointly optimizing both base station transmission beamforming and RIS reflection beamforming. To solve joint optimization problems, AO is generally used to divide a joint optimization problem into two sub-problems, i.e., active base station transmission beamforming and passive RIS beamforming, and then obtain their optimal solutions respectively. This approach is advantageous in that, for a given vector of passive RIS beamforming, it converts the problem of designing active base station transmission beamforming to a conventional optimization problem, for which numerous optimization tools are readily available to solve its solution. Notwithstanding, for a given vector of base station beamforming, the passive RIS beamforming remains a knotty problem to be tackled. The main challenges include the unit-modulus constraint and the discrete nature of the inherent feasible set of the RIS reflection unit. [17] researched a multi-user MISO system comprising of several antenna base stations (BSs), each communicating with a single-antenna UE. The RIS was used to assist wireless transmission and inhibit inter-cell interference. Efficient algorithms based on alternative optimization can be utilized to maximize users' minimum weighted SNR by jointly optimizing the coordinated transmission beamforming of BSs and the RIS reflection beamforming.

To raise the system summation rate as far as possible, Huang et al employed the combination of MM and alternative optimization in the context of a multi-user MIMO system. They managed to increase the system summation rate by over 40% without increasing any energy consumption [18]. In addition, Wang et al proposed in 2020 a self-defined block-coordinate accelerated proximal gradient (APG) algorithm [19] to jointly optimize the transmission beamformer on base stations and continuous or discrete phase shifts on the RIS, thus maximizing the summation rate.

To minimize the total AP transmission power, [8] put forward a punishment-based optimization algorithm and a two-phase optimization algorithm for obtaining the second-best solution in the context of multi-user transmission. The two algorithms offer different tradeoffs between complexity and performance. The punishment-based approach delivers better performance with higher complexity. Under the constraint of user signal to interference plus noise ratio (SINR), [12] generalized the single-user case to the multi-user setting, which more commonly found, to minimize the total AP transmission power. Similar to the single-user case, the algorithm used alternative optimization, but the difference was that for AP transmission beamforming, the principle of minimum mean square error (MMSE) was applied to address the multi-user interference, instead of MRT being used in the single-user case where there was no interference. Moreover, inspired by the maximization of combined channel gain in single-user situations, researchers divided the design of joint beamforming into two beamforming sub-issues and utilized the two-phase algorithm, which was less complex than the alternative optimization algorithm, to optimize the phase shift and transmission beamforming respectively.

References

[1]. W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "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.

[2]. W. Tang, X. Chen, M. Z. Chen, J. Y. Dai, Y. Han, M. Di Renzo, S. Jin, Q. Cheng, and T. J. Cui, "Path loss modeling and measurements for reconfigurable intelligent surfaces in the millimeter-wave frequency band," arXiv: 2101.08607, Jan. 2021.

[3]. M. Di Renzo, F. H. Danufane, X. Xi, J. de Rosny, and S. A. Tretyakov, "Analytical Modeling of the Path-loss for Reconfigurable Intelligent Surfaces - Anomalous Mirror or Scatterer?" IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), May 2020.

[4]. 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., vol. 10, no. 5, pp. 938-942, May 2021.

[5]. Dou Jianwu, Chen Yijian, Zhang Nan, et al. On the Channel Modeling of Intelligent Controllable Electro-Magnetic-Surface [J]. Chinese Journal of Radio Science, 2021, 36(3): 368-377.

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

[7]. D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (IEEE ICASSP'19), Brighton, UK, May 2019, pp. 4659–4663.

[8]. Q. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M. S. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," IEEE Open J. Commun. Soc., vol. 1, pp. 661–680, May 2020.

[9]. Hu C , Dai L . Two-Timescale Channel Estimation for Reconfigurable Intelligent Surface

Aided Wireless Communications[J]. 2019.

[10]. Wang P , Fang J , H Duan, et al. Compressed Channel Estimation and Joint Beamforming for Intelligent Reflecting Surface-Assisted Millimeter Wave Systems[J]. 2019.

[11]. You C , Zheng B , Zhang R . Channel Estimation and Passive Beamforming for Intelligent Reflecting Surface: Discrete Phase Shift and Progressive Refinement[J]. 2019.

[12]. Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in Proc. IEEE GLOBECOM, Dec. 2018.

[13]. Huang C , Member, IEEE, et al. Reconfigurable Intelligent Surface Assisted Multiuser MISO Systems Exploiting Deep Reinforcement Learning[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(8):1839-1850.

[14]. S. Abeywickrama, R. Zhang, Q. Wu and C. Yuen, "Intelligent Reflecting Surface: Practical Phase Shift Model and Beamforming Optimization," in IEEE Transactions on Communications, vol. 68, no. 9, pp. 5849-5863, Sept. 2020.

[15]. X. Yu, D. Xu and R. Schober, "Optimal Beamforming for MISO Communications via Intelligent Reflecting Surfaces," 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2020, pp. 1-5,

[16]. X. Yu, D. Xu and R. Schober, "MISO Wireless Communication Systems via Intelligent Reflecting Surfaces : (Invited Paper)," 2019 IEEE/CIC International Conference on Communications in China (ICCC), 2019, pp. 735-740.

[17]. H. Xie, J. Xu and Y. Liu, "Max-Min Fairness in IRS-Aided Multi-Cell MISO Systems via Joint Transmit and Reflective Beamforming," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), 2020, pp. 1-6.

[18]. C. Huang, A. Zappone, M. Debbah and C. Yuen, "Achievable Rate Maximization by Passive Intelligent Mirrors," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 3714-3718.

[19]. Li Q , Cui X , Wu S X , et al. Sum Rate Maximization for Multiuser MISO Downlink with Intelligent Reflecting Surface[J]. 2019.

Chapter III **RIS Realization and Prototype Validation**

3.1. Experimental Validation 1

To explore the actual performance of RIS, its application scenarios, as well as possible problems in its indoor and outdoor deployment, China Mobile completed in Nanjing the testing validation of the feasibility of the RIS new technology in a 5G live network environment in June 2021, in collaboration with the team of Cui Tiejun, an academician of CAE, at Southeast University and the Hangzhou Qiantang Information Co., Ltd. The RIS new technology validated will enable adjustable electromagnetic unit components and flexible control of the beamforming direction.

1. Testing Environment

To ensure that the test result would truly reflect the improving effect of RIS at cell edges and in weak coverage areas, the working frequency was locked for all UEs used in the validation. Further, the validation was conducted in several scenarios possible for actual application, including under-tower shadow zone, indoor coverage from outdoors, and outdoor traversing.

1) Under-tower shadow zone

An "under-tower shadow zone", i.e., a weak coverage zone, often exists under a base station tower, due to the limitations of the downtilt angle of the base station antenna and antenna direction. A RIS was installed in a location with good reception inside the test cell and the RIS panel parameters were adjusted so that signals from the base station would be reflected to the under-tower shadow zone. The CDF distribution differences in indicators, such as RSRP, SINR, and throughput of users in the zone, were compared to examine the overall improving effect of RIS on signal coverage in the under-tower shadow zone and validate whether the RIS could satisfy the high-speed services for users in the zone. Figure 3-1 shows the site of the test case. The selected cell base station mainly covered urban roadways. The RIS, installed at the diagonally opposite side of the road in front of a building, received signals from the base station (46 meters high) on the 13th floor of the building and reflected the signals to the weak signal coverage area at the back of the building. The straight-line distance between the base station and the RIS was about 120 meters and the transmission in between traveled on a line-of-sight (LOS) path.

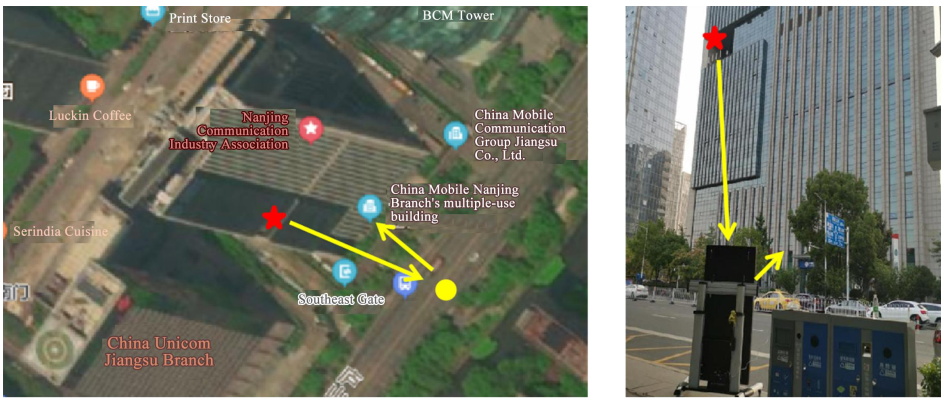


Figure 3-1 Image of the Site for the Under-Tower Shadow Zone Test

2) Indoor coverage from outdoors

During propagation, electromagnetic signals will be significantly attenuated by factors such as reinforced concrete building walls, glass curtain walls, and aluminum alloy building materials. Therefore, in most cases, indoor environments are typical weak coverage settings. Capable of tuning the reflection beam shape, the RIS can enhance indoor signal coverage by converging beams for greater penetration. In this test case of indoor weak coverage, the overall effect of RIS in the "indoor coverage from outdoors" scenario was examined by comparing the transmission performance difference before and after RIS installation at multiple points. The indoor scenario of the test case was set up in an office building with glass curtain walls. Stable transmission tests were conducted at multiple points inside the building. The RIS reflection panel was installed on the opposite of the road under the building. The straight-line distance between the base station and the RIS was about 65 meters and the transmission in between traveled on a line-of-sight (LOS) path. See Figure 3-2 for the image of the site for the test case.

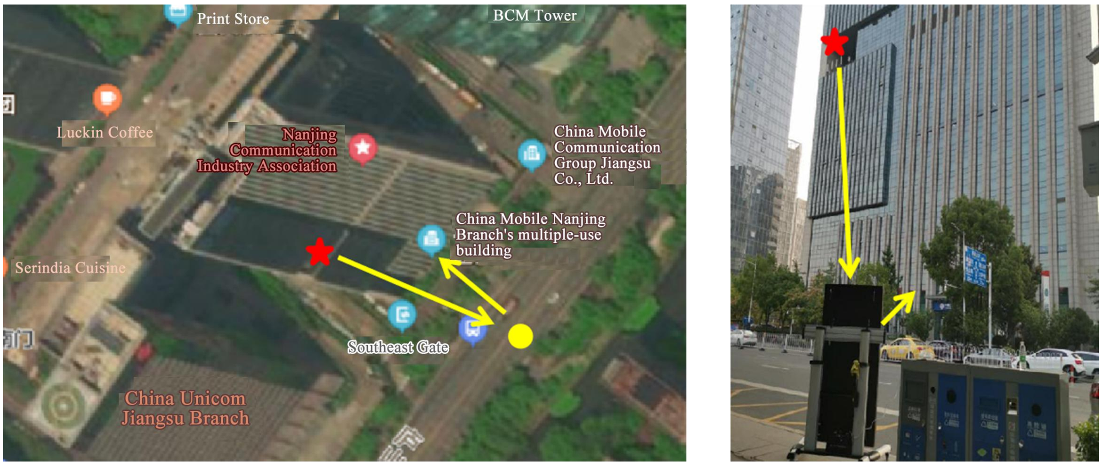


Figure 3-2 Image of the Site for the "Indoor Coverage from Outdoors" Test

3) Outdoor traversing

RIS can flexibly control the reflection beams and the beam shape and thus can be used for filling coverage holes, improving coverage at cell edges, and other scenarios. Given that, the outdoor traversing test was conducted in an urban area. The improving effect of the installed RIS on base station coverage was investigated by comparing the differences in indicators of UE traversing in the cell under overlapped coverage before and after the installation. A pole-mounted site (10 meters high) in an urban area with a dense population was selected for the test case. Occluded by high buildings in the test cell, the signals of the pole-mounted site only covered a limited range. (The road perpendicular to the one where the pole-mounted site stood was a weak coverage area.) The RIS was deployed at a crossroad 70 meters away from the pole-mounted site to receive signals from the pole-mounted site and reflect the signals to the weak coverage road in the test cell. UE traversing was conducted on the weak coverage road at a constant speed before and after the RIS deployment respectively. The transmission path between the RIS and the UE and that between the UE and the base station were good LOS paths. See Figure 3-3 for the image of the site for the test case.

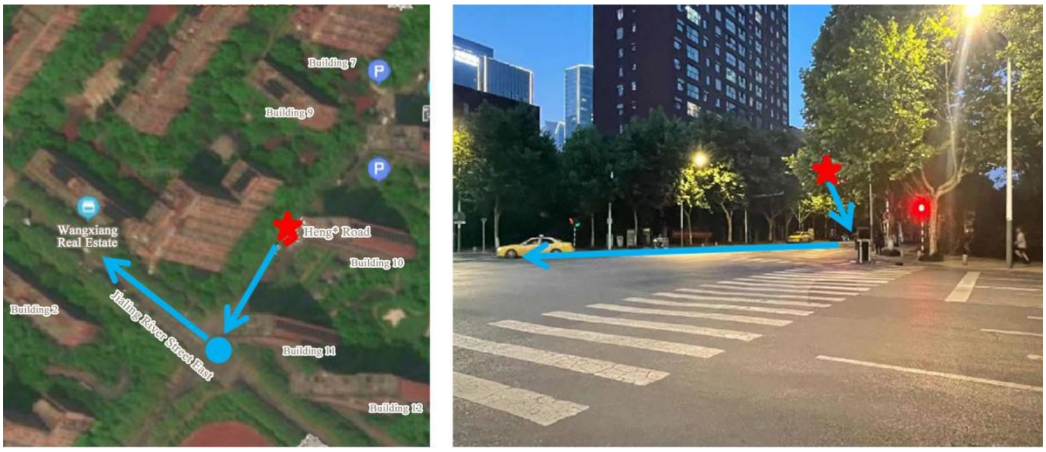


Figure 3-3 Image of the Site for the Outdoor Traversing Test

2. Base Station and Cell Configuration

Two outdoor base stations were involved in the above test cases. One was on the 13th floor of a building with glass curtain walls and the other was a pole-mounted site in an urban area with a dense population. See the table below for information on the base stations and cells.

Table 3-1 Base Station Configuration Information

Test Case	Type	Transmission Power	RRU Type	Base Station Model	Downtilt Angle of Installation	Direction Angle of Installation
Under-tower shadow zone	Outdoor base station	327 W	64 channels	Huawei	9°/10°	60°
Indoor coverage from outdoors						
Outdoor traversing	Outdoor base station	327 W	64 channels	Huawei	6°/ 3°	200°

Table 3-2 Cell Configuration Information

Test Case	Sector Number	Downlink Frequency Point	Downlink Bandwidth	Physical Cell Identification	Cell Duplex Mode	Timeslot Ratio
Under-tower shadow zone	1	504990	100	301	TDD	8:2
Indoor coverage from outdoors						
Outdoor traversing	2	504990	100	13	TDD	8:2

3. RIS Panel Configuration

The field performance test was jointly conducted by China Mobile, Southeast University, and Hangzhou Qiantang Information Co., Ltd. using a 160 cm*80 cm panel.

Figure 3-4 RIS Prototype for Field Test

4. Validation Conclusions

1) Under-tower shadow zone

Respectively before and after RIS deployment, a UE was carried to traverse the under-tower shadow zone at a constant speed along the same route. This resulted in the RSRP, SINR, and throughput dot figures before and after the RIS deployment, as shown in Figure 3-5.

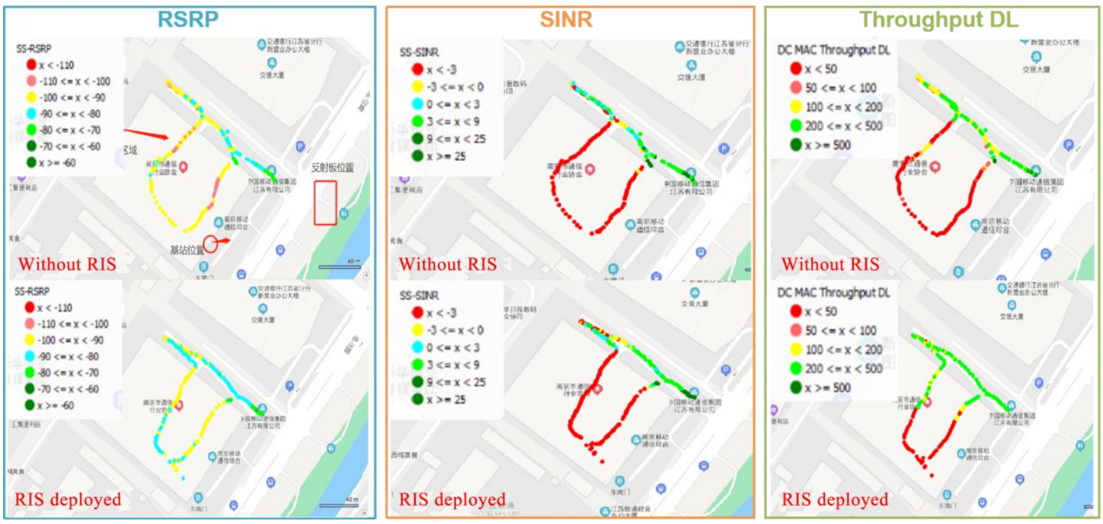


Figure 3-5 RSRP, SINR, Throughput Dot Figures Before and After RIS Deployment in Under-Tower Shadow Zone Test Case

With the RIS deployed, the RSRP and throughput were found to significantly improve and there were relatively fewer weak, low-quality overage areas. No significant changes were observed in SINR. According to analysis, this may be because the RIS also magnified the interference signals of adjacent cells while reflecting the base station signals. Moreover, relatively substantial performance improvement was observed at the back of the office building as well, although the UE-RIS transmission path was NLOS. Reasons for this phenomenon may be that a setting full of refraction and diffraction formed by nearby obstacles made the base station downlink signals reflected by the RIS receivable by the UE within the area.

The test results indicated that before RIS deployment, both the RSRP of cell edge users and the average RSRP of users were relatively lower at -102.18 dBm and -94.93 dBm, respectively. After RIS deployment, RSRP was increased to some extent. The figure was increased by 4.03 dB for the RSRP of cell edge users and 3.8 dB for the average RSRP of users. The RIS deployment raised the average user throughput from 91.50 Mbps to 109.00 Mbps, up about 19%. However, it made no significant difference to SINR.

2) Indoor coverage from outdoors

The indoor scenario was set up on the 2nd and 4th floors of an office building with glass curtain walls. For the office area on the 2nd floor, eight points were selected, covering three typical scenarios: conference room, office, and studio. The 4th floor was a supermarket, where stable transmission tests were conducted at four locations. See Figure 3-6 for a diagram of the test point locations.

Figure 3-6 Schematic Plan for Fixed-Point Transmission Test in the Office Building

It was found during the test process that the UE could not receive downlink signals transmitted from the base station when it was behind the indoor reinforced concrete walls. This suggested that the RIS version used in the test had lower penetration capability to cover indoor environments from outdoors such that it was not able to further penetrate an interior wall after penetrating a glass curtain wall. The following data were obtained by averaging the data recorded after one minute of stable signal transmission at each point that can receive downlink signals.

Preliminary conclusions were drawn by analysis of the fixed-point transmission data before and after RIS deployment. Specifically, after RIS deployment, RSRP, SINR, and downlink transmission rate were increased at most points. Despite the loss after signals penetrating the glass curtain wall, RSRP increases of 3~17 dB were stilled found, averaging at 10 dB across all points, and the transmission rate increases of 5~137 Mbps were also observed, averaging at 78.19 Mbps across all points. The increment varied significantly across the points probably due to signal fluctuations and limited range of RIS coverage. Although the gain was substantially improved, the signals failed to continue to penetrate an interior wall (15 dB of interior wall penetration loss or above) as the indoor coverage had poor overall quality and the basic level was lower in the test setting (about -100 dB after RIS deployment). In addition, the gain was not significantly improved in the studio on the 2nd floor. Since the studio was on the rightmost side of the building and the RIS reflection beams were limited in width, it was likely that the test area was beyond the coverage of RIS reflection beams.

3) Outdoor traversing

The outdoor traversing test was conducted by carrying a UE to traverse the test cell at a constant speed along the same route, respectively before and after RIS deployment. This resulted in the RSRP, SINR, and throughput dot figures shown in Figure 3-7.

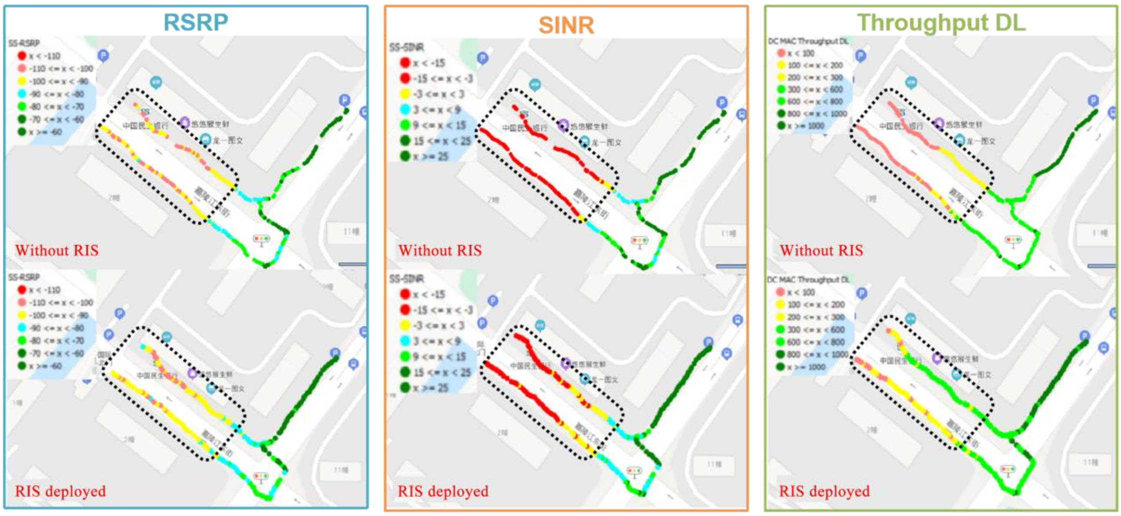


Figure 3-7 RSRP, SINR, Throughput Dot Figures in the Outdoor Traversing Test

The dot figures show that RIS deployment did not lead to significant changes to performance indicators in areas where RSRP was greater than -80 dBm, whereas in areas where RSRP was smaller than -90 dBm, RSRP, SINR, and throughput after RIS deployment were significantly better than those before RIS deployment.

Further, the test data suggest that RIS deployment had a significant influence on cell edge users, with RSRP, SINR, and throughput increasing by about 3.3 dB, 1.45 dB, and 79 Mbps, respectively. In contrast, the average RSRP for users was improved by 1.25 dB, meaning no significant influence on the average gain for users. In addition, to examine the improving effect of RIS on the cell coverage, a remote base station test was conducted in the scenario. Specifically, respectively before and after RIS deployment, a tester carrying a UE kept moving until the UE could not receive signals. A comparison between the two locations of communication interruption showed that the longest reception distance before RIS deployment was 150 meters and was increased by about 60 meters to 210 meters after. Results of the test case demonstrated that RIS was of substantial practical value in improving performance for cell edge users and extending cell coverage.

3.2. Experimental Validation 2

From 2018 to 2021, NTT DOCOMO, a Japanese company, completed several validation experiments on RIS prototypes using different prototype systems, confirming potential application scenarios of intelligent metamaterial surfaces such as filling outdoor coverage holes, interference control, and outdoor-to-indoor coverage enhancement.

Figure 3-8 Onboard Millimeter-Wave Metamaterial Reflector (left); Test Scenario and Results (right) [1]

In 2018, DOCOMO worked with Metawave Corp. to conduct an outdoor field validation experiment on the coverage enhancement capability of metamaterial reflector in a 28 GHz millimeter-wave system [1]. The metamaterial reflector used in the experiment, manufactured by Metawave Corp., was deployed on the top of the test vehicle, as shown in Figure 3-8. The metamaterial reflector was 80 cm × 80 cm × 5 cm and weighed around 700 g. With plane wave incidence taken into account, its reflection beam width was 18 degrees. The base station was deployed on the roof of a building in the experimental area. The roads under the building were blind zones, to which the metamaterial reflector was used to reflect 28 GHz millimeter-wave signals in the experiment. Results showed that using a metamaterial reflector to fill coverage holes could improve the SNR by over 15 dB and increase the rates by over 500 Mbps. Thus, the experiment verified the capability of metamaterial surfaces to fill coverage holes.

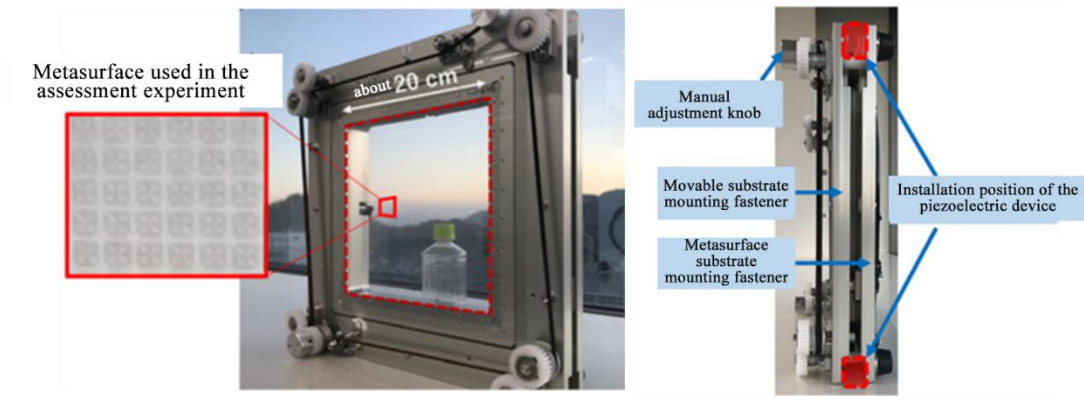


Figure 3-9 Appearance (Left) and Structure (Right) of a Millimeter-Wave Transparent Dynamic Metasurface[2]

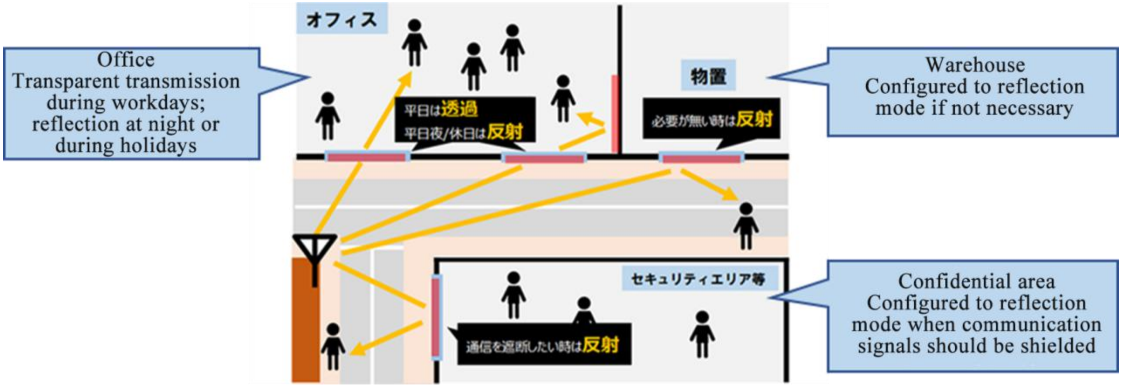
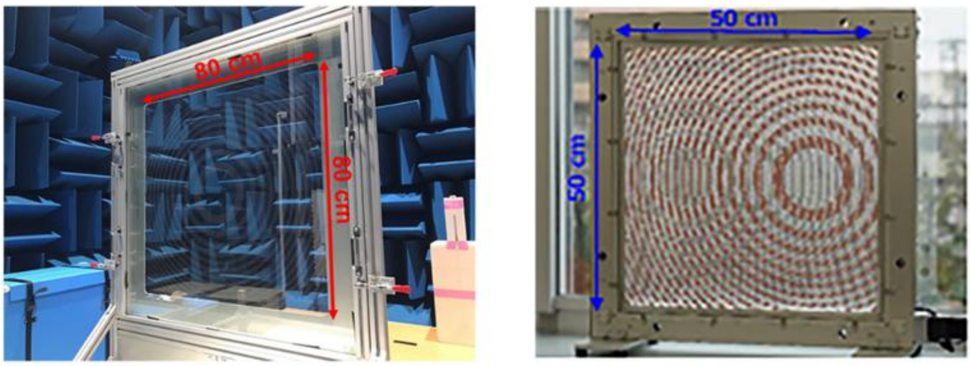


Figure 3-10 Imagined Application Scenarios of Transparent Dynamic Metasurface[2]

DOCOMO cooperated with AGC in 2020 to develop a transparent dynamic metasurface for millimeter waves and conducted an experiment in DOCOMO R&D Center to verify the dynamic tuning capability and transparency effect of the metasurface[2]. The metamaterial surface used in the experiment, designed by DOCOMO and manufactured by AGC, comprised of two substrates, one of which contained transparent electromagnetic material. As shown in Figure 3-9, the metamaterial surface could switch between three modes, i.e., transparent transmission, semi-reflection, and reflection, through adjustments to the gap between the two substrates. The observed results showed that when the metasurface was in transparent transmission mode, 28 GHz millimeter waves were only attenuated by about 1 dB. However, this figure was more than 10 dB when the metasurface was in reflection mode. Figure 3-10 gives an example of the application and deployment of the transparent metamaterial surface. Transparent dynamic metasurfaces can be flexibly configured to transparent transmission or reflection mode to meet the communication or interference shielding needs across various scenarios, such as office, warehouse, and confidential areas. The experiment proved the capability of metamaterial surfaces being applied in interference control scenarios.

Meanwhile, metamaterial surfaces with transparent design can be deployed at a large scale on windows, building surfaces, and other positions to realize the required functions since they do not influence the appearance of the surrounding environment.



Meta-surface lens (before transparency processing)

Meta-surface lens with focus control function (before transparency processing)

Figure 3-11 Fixed-Focus Metasurface Lens (Left) and Metasurface Lens with Focus Control Function (Right) [3]

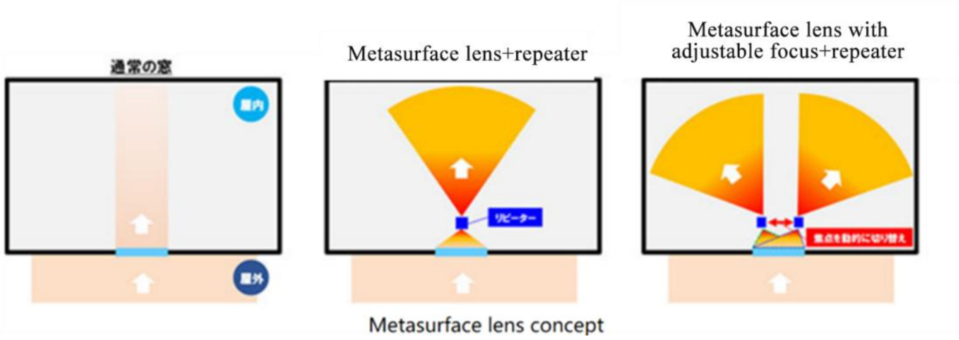


Figure 3-12 Common Glass, Fixed Metasurface Lens, and Dynamic Metasurface Lens (from Left to Right) [3]

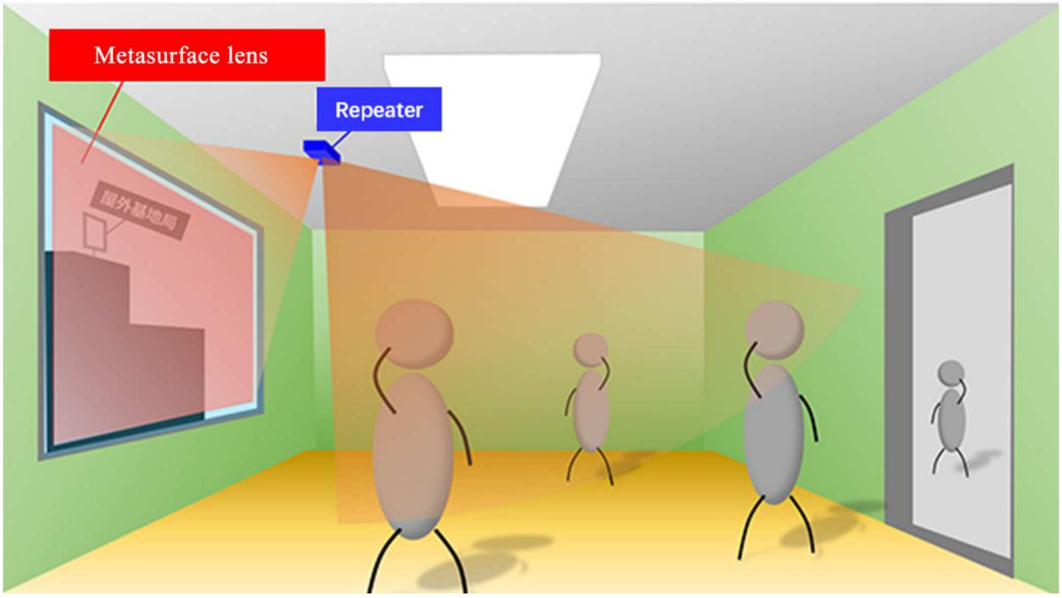


Figure 3-13 Metasurface Lens Combined with a Repeater installed on the Focus to Enhance Indoor Coverage [3]

In 2021, DOCOMO continued its partnership with AGC and developed metasurface lenses for millimeter waves, shown in Figure 3-11, validating the focusing and dynamic focus control capabilities of metasurfaces[3]. The metasurface lenses were also manufactured with transparency material and capable of focusing signals and dynamically changing the focus. The experimental result showed that, compared with common glass, metasurface lenses improved the SNR at the focus by around 24 dB. The dynamic metasurface lens designed and manufactured by DOCOMO and AGC were capable of focus control and could switch between 1-2 focus, as shown in Figure 3-12. When the metasurface lens operated in single-focus mode, the gain at the focus was about 11 dB. When it operated in dual-focus mode, the gains at the focuses were 6 dB respectively. Millimeter-wave outdoor-to-indoor coverage enhancement is a typical application scenario of the metamaterial lens. The metamaterial lens can be used in combination with a repeater or other relay devices to improve indoor millimeter-wave coverage. As shown in Figure 3-13, a repeater can be deployed at the focus of the metamaterial lens so that the communication link between the repeater and the outdoor millimeter-wave base station will be efficient and stable. At the same time, after being magnified and forwarded by the repeater, the coverage of the outdoor millimeter-wave base station can be expanded to the indoor areas, enhancing the outdoor-to-indoor coverage.

3.3. Experimental Validation 3

In June 2021, China Unicom and ZTE validated the RIS technology in a 5G field environment in Shanghai. In the 5G live network environment, the gNB, RIS, and UE were deployed as shown in Figure 3-14. The C-band gNG was deployed on the roof of a prototype building, whereas the UE was deployed in the channel. The communication path between the gNB and the UE was an NLOS path. Due to limitations of the on-site environment, the RIS was deployed about 60 meters away from the gNB and about 52 meters away from the UE, to ensure that a LOS transmission path existed between the RIS and the gNB and to shorten the distance between the RIS and the UE. Nevertheless, conditions remained to be met for the LOS path transmission between the RIS and the UE. Before the RIS was deployed, the rank fluctuated between 1-2 and the throughput was about 250 Mbps because there was no LOS transmission path between the gNB and the UE, and the RSRP and the SINR were relatively lower around the UE in the channel. With the RIS deployed, although the RSRP and the SINR around the UE were not significantly increased, significant improvements were observed for the wireless channel rank and throughput. The rank fluctuated between 3-4 and the throughput rose to 350 Mbps. The test results suggest that performance was improved by over 40% at the edge of the NLOS coverage cell of the 5G IF base station.

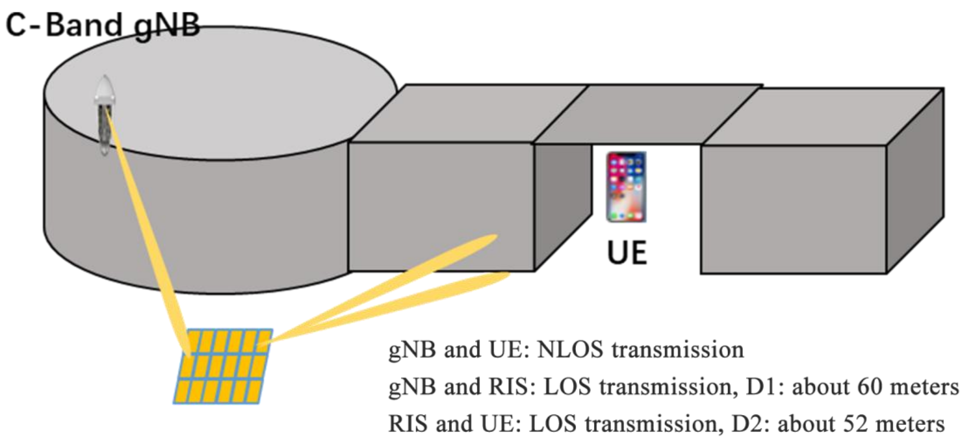


Figure 3-14 Deployment of the RIS Field Test Conducted by China Unicom and ZTE

Figure 3-15 Image of the RIS Field Test Conducted by China Unicom and ZTE

The RIS field test confirmed that the technology can reconstruct the wireless transmission environment between a 5G IF base station and commercial endpoints and boost the in-depth coverage capability of the 5G network at extremely low energy consumption cost. It thus will provide a revolutionary wireless technology to enable the new service of 5G-Advanced ultra-high experience.

References

[1]. D. Kitayama, D. Kurita, K. Miyachi, Y. Kishiyama, S. Itoh and T. Tachizawa, "5G Radio Access Experiments on Coverage Expansion Using Metasurface Reflector at 28 GHz," 2019 IEEE Asia-Pacific Microwave Conference (APMC), 2019, pp. 435-437.

[2]. NTT DOCOMO, "DOCOMO Conducts World's First Successful Trial of Transparent Dynamic Metasurface," https://www.nttdocomo.co.jp/english/info/media_center/pr/2020/0117_00.html.

[3]. NTT DOCOMO, "DOCOMO and AGC Use Metasurface Lens to Enhance Radio Signal Reception Indoors," https://www.nttdocomo.co.jp/english/info/media_center/pr/2021/0126_00.html.

Chapter IV **Technical Challenges and Standardization**

4.1. Technical Challenges

So far, the industry has planned the RIS test validation and trial application in some scenarios to promote the key technological validation and comprehensive performance assessment of RIS. However, it's not uncommon to find that severe challenges remain in the hardware realization, engineering deployment, theoretical and solution design, control solutions, network architecture, and networking of RIS.

For hardware realization, RIS material and devices are not quite mature and are relatively costly. The performance of adjustable components cannot meet control needs and their structural design remains to be optimized. With the limitations of the control rates of adjustable components, the RIS has not been able to be dynamically controlled at a high frequency band. The RIS has a certain degree of reflection loss, thus having difficulties in achieving ultra-long distance coverage. Currently, the RIS more commonly used on low-bit adjustable elements is faced with the problem of increased large-angel beamforming grating lobes, which will degrade other aspects of user communication performance. As a result, in multi-user communication situations, networking may produce interference to other cells and even the networks of other operators. Further, a metamaterial panel comprises hundreds of cycle units. Some faulty tuning components on the metasurface will cause metamaterial units to fail to deliver functions as expected and it will be difficult to troubleshoot and repair the faulty units. The increased number of faulty metamaterial units will lead to magnetic-wave tuning performance degradation of the entire metamaterial. This includes problems such as lower beam gain, beam pointing deviation, increase in minor lobes, etc.

In terms of engineering deployment, the larger size of the RIS panel makes it necessary to communicate with the property management company and owners and entails great wind resistance. The power feeding requirement of RIS will limit its deployment and may be faced with weak-current interference.

Regarding the theoretical and solution design, the transmission solution design of RIS systems lacks strong theoretical support as reliable, complete transmission theoretical basis, channel models, and system models are yet to be developed. The existing OTA transmission solutions are highly complex and costly and have limited feasibility. Therefore, an complete, trustworthy evaluation system taking non-ideal factors into account should be established for actual system performance. Further, channel estimation, joint beamforming, and other realizable basic OTA transmission solutions are to be designed based on the compromise between performance and complexity. In addition, further research is required as to whether the RIS can support sub-band dispatching solutions and whether the phase tuning of high and low-frequency arrays is faster enough.

The control method of RIS has an important influence on the design of network architecture, power consumption, and deployment. A compromise between power consumption and network complexity should be considered for truly passive/semi-passive and dynamic control. In terms of networking, further assessment and investigation are warranted as to whether the RIS can obtain performance gain in multi-bandwidth, multi-system communication mode and how the transmission solution should be designed.

Introducing RIS into wireless networks will bring network coexistence challenges. In actual networks, the wireless signals transmitted to the RIS panel include both "target signals" to be optimized and tuned by the RIS and "non-target signals". Both types of signals will be tuned by the RIS. The RIS will tune the unintended "non-target signals" while reinforcing "target signals" by tuning the amplitude, phase, and polarization of electromagnetic waves. Without control, the RIS will perform unintended abnormal tuning on "non-target signals" from other networks, leading to a serious network coexistence problem. Therefore, it is necessary to carry out in-depth research on the network coexistence problem caused by large-scale RIS deployment and provide effective solutions. Moreover, the presence of the network coexistence problem also suggests that large-scale RIS deployment must be controlled by the network so as to constrain its arbitrary, unintended abnormal tuning behavior on "non-target signals" in the wireless environment and prevent severe network performance degradation.

4.2. Standardization Research Directions

Research on the RIS technology is currently at an initial stage. Research regarding the RIS structure and the effect of its application on communication systems are not sufficiently comprehensive. Therefore, standardization research has not officially begun. Based on prior standardization experience, standardization research on RIS serving as a reflector to assist communication can be conducted in the following aspects:

1) Physical channel design: The exiting functions of RIS mean that it does not need to perform autonomous downlink or uplink data transmission, but may be required to receive and send some control information. In addition, the design of UCI and DCI can be simpler as the RIS itself is not involved in the dispatching of physical resources. Therefore, new DCI and UCI formats should probably be defined.

2) Reference signal design: RIS-related reference signals are mainly used for channel measurement. The design of reference signals should consider the exact type of channel information to be provided by the RIS. In addition, given its low cost and low complexity characteristics, RIS may not be capable of demodulating complex signals. Therefore, reference information demodulation should not be excessively complex and must be compatible with the capabilities of RIS.

3) Channel measurement and channel feedback solutions: Channel measurement and feedback solutions are to be designed in a targeted manner by considering the technical requirements on RIS. Further, since the base station and the RIS are located at relatively fixed places, the large-scale parameters of channels will be kept constant for a long time. On the contrary, the RIS's position relative to users may change in real time, depending on user mobility. Therefore, the channel between the base station and that between RIS and users may require different channel measurement and feedback solutions.

4) Beamforming design of RIS devices: For the optimal quality of reception signals, RIS-assisted communication systems require the optimization of both the digital beamforming of the base station and the analog beamforming of the RIS device. In actual settings, the existence of multiple users and multiple cells compound the problem of BS-RIS joint beamforming. The beamforming generation solution for RIS devices and beam instruction signaling should be further improved based on the channel measurement information fed back.

5) Standardization of RIS device indicators: The electromagnetic signal control function of RIS devices can be realized by various technical solutions. Different solutions will have different device characteristics, such as reflection/transparent transmission signals, pure passive RIS/active and passive hybrid RIS, continuous/discrete phase tuning, phase tuning response time, working bandwidth, and working frequency point. In view of the time-frequency structure of communication systems and the signal costs for a base station to control RIS devices, communication systems must define the macro-indicators of RIS nodes and contain necessary parameter reporting interfaces. Efforts should also be made to guide the healthy, well-organized, and compliance-oriented development of RIS technology research institutes and RIS device manufacturers.

6) Evolution of network deployment solutions: Different from the IAB nodes in the LTE/NR system and the smart repeaters introduced by NR, simple RIS devices can be flexibly deployed at scale within cell coverage areas. This has enabled the network architecture to evolve from the

original base station-endpoint star topology network to base station-RIS-endpoint tree/net topology network. The latter necessitates the design of new data routing/RIS device dispatching solutions to maximize the assisting role of RIS devices in communication.

Acknowledgment

Responsible editors: Yi Zhiling, Wang Anna, and Miao Long

Contributing units and experts

S/N	Contributing Unit	Contributor
1	China Mobile Communication Technology Co., Ltd.	Li Ya, Gu Qi, and Wu Dan
2	Hangzhou Qiantang Information Co., Ltd.	Yang Linjun
3	ZTE Corporation	Zhao Yajun and Jian Mengnan
4	China Telecom Corporation Limited	Li Nanxi, Zhu Jianchi, and Li Pengxiang
5	Beijing Xiaomi Mobile Software Co., Ltd.	Duan Gaoming and Chi Liangang
6	Lenovo Research	Bao Tingnan
7	vivo Mobile Communication Co., Ltd.	Jiang Dajie and Yang Kun
8	NTT DoCoMo	Wang Xin and Hou Xiaolin
9	China United Network Communications Group Co., Ltd.	Liu Qiuyan

Acronyms

Acronym	Full Name
RIS	Reconfigurable Intelligent Surface
MIMO	Multi-Input Multi-Output
BFSK	Binary Frequency Shift Keying
QPSK	Quadrature Phase Shift Keying
PSK	Phase Shift Keying
ASK	Amplitude Shift Keying
QAM	Quadrature Amplitude Modulation
BS	Base Station
UE	User equipment
CSI	Channel State Information
DnCNN	Denoising Convolutional Neural Network
GAN	Generative Adversarial Networks
LS	Least Sqaure
MMSE	Minimum Mean Square Error
DFT	Discrete Fourier Transform
MISO	Multi-Input Single-Output
AOA	Angle-of-Arrival
SDP	Semidefinite Program
DRL	Deep Reinforcement learning
DDPG	Deep Deterministic Policy Gradient
SINR	Signal to Interference plus Noise Ratio
AP	Access Point
RSRP	Reference Signal Receiving Power