

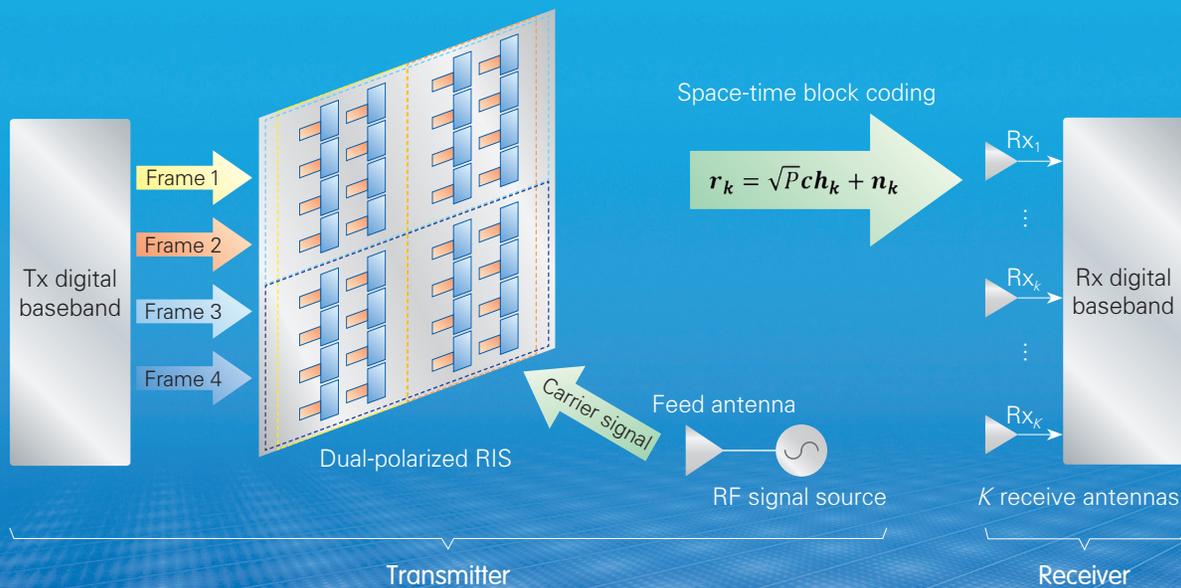


# ZTE COMMUNICATIONS

中兴通讯技术(英文版)

March 2022, Vol. 20 No. 1

## Special Topic: Reconfigurable Intelligent Surface (RIS)



# The 9th Editorial Board of ZTE Communications

**Chairman** GAO Wen, Peking University (China)

**Vice Chairmen** XU Ziyang, ZTE Corporation (China) | XU Chengzhong, University of Macau (China)

## Members (Surname in Alphabetical Order)

AI Bo	Beijing Jiaotong University (China)
CAO Jiannong	Hong Kong Polytechnic University (China)
CHEN Chang Wen	The State University of New York at Buffalo (USA)
CHEN Yan	Northwestern University (USA)
CHI Nan	Fudan University (China)
CUI Shuguang	UC Davis (USA) and The Chinese University of Hong Kong, Shenzhen (China)
GAO Wen	Peking University (China)
GAO Yang	Nanjing University (China)
GE Xiaohu	Huazhong University of Science and Technology (China)
HE Yejun	Shenzhen University (China)
HWANG Jenq-Neng	University of Washington (USA)
Victor C. M. LEUNG	The University of British Columbia (Canada)
LI Xiangyang	University of Science and Technology of China (China)
LI Zixue	ZTE Corporation (China)
LIAO Yong	Chongqing University (China)
LIN Xiaodong	ZTE Corporation (China)
LIU Chi	Beijing Institute of Technology (China)
LIU Jian	ZTE Corporation (China)
LIU Yue	Beijing Institute of Technology (China)
MA Jianhua	Hosei University (Japan)
MA Zheng	Southwest Jiaotong University (China)
PAN Yi	Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (China)
PENG Mugen	Beijing University of Posts and Telecommunications (China)
REN Fuji	Tokushima University (Japan)
REN Kui	Zhejiang University (China)
SHENG Min	Xidian University (China)
SU Zhou	Xi'an Jiaotong University (China)
SUN Huifang	Mitsubishi Electric Research Laboratories (USA)
SUN Zhili	University of Surrey (UK)
TAO Meixia	Shanghai Jiao Tong University (China)
WANG Chengxiang	Southeast University (China)
WANG Haiming	Southeast University (China)
WANG Xiang	ZTE Corporation (China)
WANG Xiaodong	Columbia University (USA)
WANG Xiyu	ZTE Corporation (China)
WANG Yongjin	Nanjing University of Posts and Telecommunications (China)
XU Chengzhong	University of Macau (China)
XU Ziyang	ZTE Corporation (China)
YANG Kun	University of Essex (UK)
YUAN Jinhong	University of New South Wales (Australia)
ZENG Wenjun	EIT Institute for Advanced Study (China)
ZHANG Honggang	Zhejiang University (China)
ZHANG Jianhua	Beijing University of Posts and Telecommunications (China)
ZHANG Yueping	Nanyang Technological University (Singapore)
ZHOU Wanlei	City University of Macau (China)
ZHUANG Weihua	University of Waterloo (Canada)

## Special Topic ►

### Reconfigurable Intelligent Surface (RIS)

01 Editorial ..... YUAN Yifei, JIN Shi, Marco Di RENZO

03 Recent Progress in Research and Development of Reconfigurable Intelligent Surface .....  
..... YUAN Yifei, GU Qi, WANG Anna, WU Dan, LI Ya

The authors aim to provide a comprehensive overview of the progress in research and development of RIS over the last 2 - 3 years, especially when RIS is used as relays in next-generation mobile networks. Major areas of research in academia are outlined, including fundamental performance, channel estimation, joint optimization with antenna precoding at base stations, propagation channel modeling and meta-material devices of RIS elements. Development in industry is surveyed from the aspects of performance potentials and issues, realistic joint optimization algorithms, control mechanisms, field trials and related activities in SDOs. Views on how to carry out the engineering-aspect study on RIS for 6G systems are also presented.

14 Some Observations and Thoughts about Reconfigurable Intelligent Surface Application for 5G Evolution and 6G ..... HOU Xiaolin, LI Xiang, WANG Xin, CHEN Lan, SUYAMA Satoshi

The typical applications of RIS for 5G evolution and 6G are discussed, including RIS-aided smart channels and RIS-aided mega MIMO. Then, several observations from RIS trials and system-level simulations are presented, especially those on the deployment strategy and the potential performance gain of RIS for coverage enhancement. The near-field effect and a two-step dynamic RIS beamforming method are also discussed. Finally, challenges and opportunities of the RIS technology for 5G evolution and 6G are summarized. The authors also suggest a step-by-step commercialization strategy as a conclusion.

21 Recent Developments of Transmissive Reconfigurable Intelligent Surfaces: A Review .....  
..... TANG Junwen, XU Shenheng, YANG Fan, LI Maokun

This paper reviews some of the latest developments of the transmissive RIS. The approaches for transmissive RIS designs are classified and described briefly. Numerous designs with different phase resolutions, such as 1-bit, 2-bit or continuous 360° phase shifts, are presented, with detailed discussions on their operating mechanisms and transmission performances. The design solutions for various transmissive RIS elements are summarized and compared.

28 IRS-Enabled Spectrum Sharing: Interference Modeling, Channel Estimation and Robust Passive Beamforming ..... GUAN Xinrong, WU Qingqing

IRS-enabled spectrum sharing is investigated, from the perspectives of interference modeling, efficient channel estimation and robust passive beamforming design. Specifically, the authors first characterize the interference in a spectrum sharing system consisting of a single PU pair and a single SU pair, and extends it to the large-scale network by leveraging the PPP. Then, they propose an efficient channel estimation framework based on decoupling the cascaded IRS channels. Moreover, the tradeoff between spectrum efficiency and energy efficiency is derived from the view of channel estimation accuracy. Finally, they discuss the robust passive beamforming design in presence of imperfect channel estimation and nonideal/discrete phase shifts.

Submission of a manuscript implies that the submitted work has not been published before (except as part of a thesis or lecture note or report or in the form of an abstract); that it is not under consideration for publication elsewhere; that its publication has been approved by all co-authors as well as by the authorities at the institute where the work has been carried out; that, if and when the manuscript is accepted for publication, the authors hand over the transferable copyrights of the accepted manuscript to *ZTE Communications*; and that the manuscript or parts thereof will not be published elsewhere in any language without the consent of the copyright holder. Copyrights include, without spatial or timely limitation, the mechanical, electronic and visual reproduction and distribution; electronic storage and retrieval; and all other forms of electronic publication or any other types of publication including all subsidiary rights.

Responsibility for content rests on authors of signed articles and not on the editorial board of *ZTE Communications* or its sponsors.

All rights reserved.

**36 Resource Allocation for Two-Tier RIS-Assisted Heterogeneous NOMA Networks** .....  
..... XU Yongjun, YANG Zhaohui, HUANG Chongwen, YUEN Chau, GUI Guan

The authors design an RA scheme for a RIS-assisted HetNet with non-orthogonal multiple access to improve spectrum efficiency and transmission rates. In particular, they jointly optimize the transmit power of the small-cell base station and the phase-shift matrix of the RIS to maximize the sum rates of all small-cell users. An efficient suboptimal RA scheme is proposed based on the alternating iteration approach, and successive convex approximation and logarithmic transformation approach. Simulation results verify the effectiveness of the proposed scheme in terms of data rates.

**48 Markovian Cascaded Channel Estimation for RIS Aided Massive MIMO Using 1-Bit ADCs and Oversampling** .....  
..... SHAO Zhichao, YAN Wenjing, YUAN Xiaojun

The authors consider RIS aided MIMO system, where the base station employs a large antenna array with low-cost and low-power 1-bit ADCs. Bussgang decomposition is applied to deal with the coarse quantization, and a Markov chain is developed to characterize the banded structure of the oversampling filter. An approximate message-passing based algorithm is proposed for the estimation of the cascaded channels.

**57 RIS: Spatial-Wideband Effect Analysis and Off-Grid Channel Estimation** .....  
..... JIAN Mengnan, ZHANG Nan, CHEN Yijian

The authors analyze the impact of the spatial-wideband effect on the RIS channel to account for the propagation delay across RIS elements and estimate sparse channel parameters such as angle and gain through a super-resolution CS algorithm. The simulation results explore the influence of the spatial-wideband effect on the RIS channel and verify the effectiveness of the proposed algorithm.

**63 Dual-Polarized RIS-Based STBC Transmission with Polarization Coupling Analysis** .....  
..... ZHOU Mingyong, CHEN Xiangyu, TANG Wankai, KE Jun Chen, JIN Shi, CHENG Qiang, CUI Tie Jun

The authors propose a dual-polarized RIS-based transmitter, which realizes 4-transmit STBC transmission by properly partitioning RIS's unit cells and utilizing the degree of freedom of polarization. The proposed scheme is evaluated through a prototype system that utilizes a fabricated dual-polarized phase-adjustable RIS. In particular, the polarization coupling phenomenon in each unit cell of the employed dual-polarized RIS is modeled and analyzed, and an initial research effort is made on characterizing the polarization coupling property in the dual-polarized RIS.

**Research Paper** ▶

**76 Metric Learning for Semantic-Based Clothes Retrieval** ..... YANG Bo, GUO Caili, LI Zheng

A novel method called SCR is proposed, the semantic similarities between clothes are measured, and a new adaptive loss based on these similarities is designed. The margin in the proposed adaptive loss can vary with different semantic similarities between the anchor and negative samples. The authors use Recall@K and nDCG as evaluation metrics to conduct experiments on the DeepFashion dataset and have achieved better performance.

**Roundup** ▶

13 New Member (HE Yejun) of *ZTE Communications* Editorial Board

20 New Member (WANG Chengxiang) of *ZTE Communications* Editorial Board

Serial parameters: CN 34-1294/TN\*2003\*q\*16\*82\*en\*P\*¥ 30.00\*2200\*10\* 2022-03

**Statement**

This magazine is a free publication for you. If you do not want to receive it in the future, you can send the "TD unsubscribe" mail to magazine@zte.com.cn. We will not send you this magazine again after receiving your email. Thank you for your support.



# Editorial: Special Topic on Reconfigurable Intelligent Surface (RIS)

## Guest Editors >>>



**YUAN Yifei** received his Bachelor & Master degrees from Tsinghua University, China, and Ph.D. degree from Carnegie Mellon University, USA. He was with Alcatel-Lucent from 2000 to 2008, working on 3G/4G key technologies. From 2008 to 2020, he was with ZTE Corporation as a technical director and the Chief Engineer, responsible for standard research on LTE-Advanced and 5G. Since January 2020, he has been with China Mobile Research Institute as the Chief Expert, responsible for advanced research of 6G air interface. His research interests include MIMO, channel coding, non-orthogonal multiple access (NOMA), Internet of Things (IoT), and resource scheduling. He has extensive publications, including seven books on LTE-Advanced relay, LTE-Advanced key technologies & system performance, narrow-band (NB) IoT, 5G new radio (NR) channel coding, 5G ultra-dense networks (UDN), 5G non-orthogonal multiple access, and 5G random access enhancements, respectively. He has over 50 granted US patents. He is the rapporteur of NOMA study item in 3GPP. He is the recipient of the Best Paper Award by IEEE Communications Society Asia-Pacific Board for a paper on NOMA in *IEEE Communications Magazine*. He serves as the editor of *IEEE Communications Letters* and *China Communications*, and the Editor in Chief of *Hans Wireless Communications*.

advanced research of 6G air interface. His research interests include MIMO, channel coding, non-orthogonal multiple access (NOMA), Internet of Things (IoT), and resource scheduling. He has extensive publications, including seven books on LTE-Advanced relay, LTE-Advanced key technologies & system performance, narrow-band (NB) IoT, 5G new radio (NR) channel coding, 5G ultra-dense networks (UDN), 5G non-orthogonal multiple access, and 5G random access enhancements, respectively. He has over 50 granted US patents. He is the rapporteur of NOMA study item in 3GPP. He is the recipient of the Best Paper Award by IEEE Communications Society Asia-Pacific Board for a paper on NOMA in *IEEE Communications Magazine*. He serves as the editor of *IEEE Communications Letters* and *China Communications*, and the Editor in Chief of *Hans Wireless Communications*.



**JIN Shi** received the Ph.D. degree in information and communications engineering from Southeast University, China in 2007. From June 2007 to October 2009, he was a research fellow with the Adastral Park Research Campus, University College London, U.K. He is currently with the faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include space time wireless communications, random matrix theory, and information theory. He serves as the associate editor of *IEEE Transactions on Communications*, *IEEE Transactions on Wireless Commu-*

*IEEE Transactions on Communications*, *IEEE Transactions on Wireless Commu-*

*nications*, *IEEE Communications Letters*, and *IET Communications*. Dr. JIN and his co-authors were awarded the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory and the 2010 Young Author Best Paper Award by the IEEE Signal Processing Society.



**Marco Di RENZO** received the Laurea (cum laude) and Ph.D. degrees in electrical engineering from the University of L'Aquila, Italy in 2003 and 2007, respectively, and the Habilitation à Diriger des Recherches (Doctor of Science) degree from University Paris-Sud (now Paris-Saclay University), France in 2013. Since 2010, he has been with the French National Center for Scientific Research (CNRS), where he is a CNRS Research Director (Professor) with the Laboratory of Signals and Systems (L2S) of Paris-Saclay University—CNRS and CentraleSupélec, France. In Paris-Saclay University, he serves as the coordinator of the Communications and Networks Research Area of the Laboratory of Excellence DigiCosme, and as a member of the Admission and Evaluation Committee of the Ph.D. School on Information and Communication Technologies. He is the Editor-in-Chief of *IEEE Communications Letters* and a distinguished speaker of the IEEE Vehicular Technology Society. In 2017 - 2020, he was a distinguished lecturer of the IEEE Vehicular Technology Society and IEEE Communications Society. He has received several research distinctions, which include the SEE-IEEE Alain Glavieux Award, the IEEE Jack Neubauer Memorial Best Systems Paper Award, the Royal Academy of Engineering Distinguished Visiting Fellowship, the Nokia Foundation Visiting Professorship, the Fulbright Fellowship, and the 2021 EURASIP Journal on Wireless Communications and Networking Best Paper Award. He is a Fellow of the IEEE, a Fellow of the UK Institution of Engineering and Technology (IET), a Fellow of the Asia-Pacific Artificial Intelligence Association (AAlA), an ordinary member of the European Academy of Sciences and Arts (EASA), and an ordinary member of the Academy of Europe (AE). Also, he is a highly cited researcher.

Paris-Saclay University—CNRS and CentraleSupélec, France. In Paris-Saclay University, he serves as the coordinator of the Communications and Networks Research Area of the Laboratory of Excellence DigiCosme, and as a member of the Admission and Evaluation Committee of the Ph.D. School on Information and Communication Technologies. He is the Editor-in-Chief of *IEEE Communications Letters* and a distinguished speaker of the IEEE Vehicular Technology Society. In 2017 - 2020, he was a distinguished lecturer of the IEEE Vehicular Technology Society and IEEE Communications Society. He has received several research distinctions, which include the SEE-IEEE Alain Glavieux Award, the IEEE Jack Neubauer Memorial Best Systems Paper Award, the Royal Academy of Engineering Distinguished Visiting Fellowship, the Nokia Foundation Visiting Professorship, the Fulbright Fellowship, and the 2021 EURASIP Journal on Wireless Communications and Networking Best Paper Award. He is a Fellow of the IEEE, a Fellow of the UK Institution of Engineering and Technology (IET), a Fellow of the Asia-Pacific Artificial Intelligence Association (AAlA), an ordinary member of the European Academy of Sciences and Arts (EASA), and an ordinary member of the Academy of Europe (AE). Also, he is a highly cited researcher.

Over the past two years, reconfigurable intelligent surface (RIS), as a promising emerging technology for Beyond 5G (B5G) and 6G mobile communications systems, has attracted enormous interest from both academia and industry worldwide. In IMT-2030 (6G) Promotion Group of China, the RIS task force was created in June 2020. In IEEE, an emerging technology initiative (ETI) of RIS was formed in July 2020. In ETSI, an industry specifications group (ISG) was kicked off in September 2021. In 2022, RIS Technology Alliance (RISTA) will be established. Initial field trials of RIS have been conducted by several companies and universities.

A RIS panel is made up with many surface-laid elements, where each element can be controlled to change the phase/am-

plitude of incident electromagnetic waves. Benefiting from the rapid development of meta-material device manufacturing, RIS panels with hundreds of programable meta-surface elements now become feasible, so that RIS can significantly improve, rather than adapt to, radio environments. By properly cooperating with base station antennas, RIS can extend the network coverage and increase the system throughput. RIS is primarily a passive device requiring little or a very small amount of power. Hence its operation is more energy-efficient than deploying traditional small cell base stations or relays which usually contain active radio devices such as power amplifiers. However, RIS is a relatively new technology that still faces a lot of challenges and issues for wireless and device researchers/engineers to solve.

In this special topic, we assembled eight papers with a reasonable divide between theoretical research and practical engineering: five of them are coming from academia and three from the industry. They span over the general trends of RIS re-

DOI: 10.12142/ZTECOM.202201001

Citation: Y. F. Yuan, S. Jin, and M. Di Renzo, "Editorial: special topic on reconfigurable intelligent surface (RIS)," *ZTE Communications*, vol. 20, no. 1, pp. 1–2, Mar. 2022. doi: 10.12142/ZTECOM.202201001.

search and development, RIS devices, signal processing and control aspects of RIS systems, and prototype RIS transmitter and receiver system. In terms of research fields, certain balance is also considered to ensure diversity. Apart from the two general papers, four papers are in the field of wireless communications, e.g., signal processing and physical layer controls. Two papers are primarily about hardware and electromagnetic physics, e.g., circuit device and test systems.

The two papers “Recent Progress in Research and Development of Reconfigurable Intelligent Surface” and “Some Observations and Thoughts about Reconfigurable Intelligent Surface Application for 5G Evolution and 6G” come from wireless operators, China Mobile and NTT DOCOMO, respectively. They provide big pictures of recent R&D of RIS at both academia and the industry. The former paper has a more comprehensive coverage, including the fundamental performance bounds, channel estimation, joint optimization algorithms, propagation channel modeling, realistic issues, control mechanisms, field trials and related activities in standardization development organizations (SDOs) of RIS. The latter paper emphasizes a little bit on demo trial and system-level computer simulations carried out by NTT DOCOMO for 28 GHz millimeter wave band. For large-scale RIS panels, near field effect is analyzed and a two-step beamforming method is proposed.

The paper “Recent Developments of Transmissive Reconfigurable Intelligent Surfaces: A Review” is contributed by professors from Tsinghua University. It reviews design approaches for RIS devices: Rx-Tx structure and frequency selective surface (FSS) structure; representative designs with different phase resolutions are illustrated, such as single-polarized 1-bit, dual-polarized 1-bit, 2-bit and continuous 360° phase shifts. The operating mechanisms, schematics and transmission performances of those RIS elements are discussed in details.

Four papers in this special topic are devoted in signal processing and control aspects of RIS, reflecting the importance of control algorithms in fulfilling the performance potential of RIS. The paper “IRS-Enabled Spectrum Sharing: Interference Modeling, Channel Estimation and Robust Passive Beamforming” from Macau University pays special attention to the interference modeling in multi-user and multi-RIS panel scenarios. By decoupling the cascaded RIS channels, a novel channel estimation framework is proposed that includes off-line estimation and on-line estimation, with certain tradeoff consideration between spectral efficiency and energy efficiency of the systems. Also, in this paper, RIS beamforming is optimized in a more realistic setting, e.g., imperfect channel estimation and non-ideal or finite bits for phase shifts. The paper “Resource Allocation for Two-Tier RIS-Assisted Heterogeneous NOMA Networks” comes from Zhejiang University and considers heterogeneous networks that consist of macro base stations, small cell base stations and RIS units. In such setting, cross-tier interference becomes a limiting factor to the overall system performance. The problem is formulated as the joint optimization

of transmit power of small cell base stations and phase-shift matrix of RIS, with the aim to maximize the sum rates of all small cells. Non-orthogonal multiple access (NOMA) is employed here to further enhance the system throughput albeit with increased complexity. An efficient sub-optimal algorithm is proposed which is based on alternating iterations to convert the original non-convex problem to a convex approximation. Simulations show that the proposed scheme can reduce the transmit power of small cell base stations by 40% compared to the case without RIS. The paper “Markovian Cascaded Channel Estimation for RIS Aided Massive MIMO Using 1-Bit ADCs and Oversampling” from University of Electronic Science and Technology of China extends the technique of uplink receiver-end oversampling with low-bit width to the scenario of cascaded user-to-RIS and RIS-to-base station where both base station and RIS are equipped with a large number of antenna elements. Bussgang decomposition is applied to handle the coarse quantization, while a Markov chain model is developed for the oversampling filter. The cascaded channels are estimated by the proposed approximate message-passing (AMP) based algorithm. It is shown from the simulation that the proposed 1-bit system with oversampling can deliver the comparable mean square error performance to the 2-bit systems, while requiring significantly less processing power at the receiver. The paper “RIS: Spatial-Wideband Effect Analysis and Off-Grid Channel Estimation” comes from ZTE Corporation, with a focus on the impact of spatial wideband effect on RIS operation. Such impact can be quite significant in millimeter wave deployment whose system bandwidth is typically ~400 MHz. The paper considers sparse channel parameters such as angles and gains, which can be estimated using compressed-sensing algorithm to achieve super resolution. The effectiveness of the algorithm is verified by the simulation results. This paper also discusses several challenges of RIS, for instance, channel modeling, near-field effect, multi-user transmission, and standardization.

The paper “Dual-Polarized RIS-Based STBC Transmission with Polarization Coupling Analysis” describes the recent work by researchers from Southeast University, China. Instead of using RIS as a passive relay, RIS in this paper is employed in the transmitter to replace some of radio-frequency processing modules in traditional base stations. A dual-polarized RIS transmitter is proposed, which can achieve 4-transmit space-time block coding (STBC) with proper partitions of RIS elements with different polarizations. The polarization coupling is thoroughly analyzed by a suitable model. A prototype system of RIS-based STBC is built using the fabricated dual-polarized phase adjustable RIS unit cells. Theoretical performance analysis and experimental results match closely, indicating the effectiveness of the model in characterizing the polarization coupling in RIS.

We would like to thank all the authors for their valuable contributions and all the reviewers for their timely and constructive comments on the submitted manuscripts. We hope that this special issue would be informative and useful for readers.



# Recent Progress in Research and Development of Reconfigurable Intelligent Surface

YUAN Yifei, GU Qi, WANG Anna, WU Dan, LI Ya  
(China Mobile Research Institute, Beijing 100053, China)

DOI: 10.12142/ZTECOM.202201002

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220228.2200.002.html>,  
published online March 4, 2022

Manuscript received: 2021-11-11

**Abstract:** We aim to provide a comprehensive overview of the progress in research and development of the reconfigurable intelligent surface (RIS) over the last 2–3 years in this paper, especially when the RIS is used as relays in next-generation mobile networks. Major areas of research in academia are outlined, including fundamental performance, channel estimation, joint optimization with antenna precoding at base stations, propagation channel modeling and meta-material devices of RIS elements. Development in industry is surveyed from the aspects of performance potentials and issues, realistic joint optimization algorithms, control mechanisms, field trials and related activities in standardization development organizations (SDOs). Our views on how to carry out the engineering-aspect study on RIS for 6G systems are also presented, which cover the realistic performance, the comparison with other topological improvements, approaches for channel modeling, factors for designing control mechanisms and the timeline for RIS standardization.

**Keywords:** 6G; reconfigurable intelligent surface; relay; meta-material; metasurface

**Citation** (IEEE Format): Y. F. Yuan, Q. Gu, A. N. Wang, et al. "Recent progress in research and development of reconfigurable intelligent surface," *ZTE Communications*, vol. 20, no. 1, pp. 3 – 13, Mar. 2022. doi: 10.12142/ZTECOM.202201002.

## 1 Introduction

Since 3G mobile communications, the spectrum limitation in licensed bands has pushed an expansion to the spatial domain to boost spectral efficiency. The multiple-input multiple-output (MIMO) technology in 4G and massive MIMO technology in 5G significantly improved the system throughput and played a major role in fulfilling the key performance requirements of IMT-Advanced and IMT-2020<sup>[1]</sup>, respectively. To further exploit the spatial dimension of propagation channels, the reconfigurable intelligent surface (RIS) has recently gained enormous attention both from academia and industry<sup>[2]</sup>, as a strong candidate technology for 6G mobile networks. Propelled by the advancement in manufacturing of meta-materials and devices, RIS is able to improve the radio propagation environment and reduce the interference by tuning the amplitude and phase of RIS elements. Compared to repeaters and relays, RIS can operate in a passive manner without the need for power amplifier, thus allowing more flexible deployment and opening a new paradigm for exploiting the spatial dimension for wireless communications.

There have been various activities across academia and industry on RIS. In academia, RIS workshops and panels are organized by the International Conference on Communications

(ICC), Global Communications Conference (GLOBECOM), etc. every year. In August 2020, an emerging technology initiative (ETI) on RIS was formed in IEEE<sup>[3]</sup>. In the industry, companies have submitted the proposals of studying RIS in the release time frame at 3GPP spanning January 2022 – June 2023. In June 2021, an industry specification group (ISG) for RIS was formed in ETSI to carry out the engineering related study and standardization of RIS. In China, task force groups on RIS were created in IMT-2030 (6G) Promotion Group and China Communications Standards Association (CCSA). Interested companies include mobile operators such as CMCC, NTT DOCOMO, China Unicom, China Telecom and KDDI, and vendors like ZTE, Huawei, vivo, CATT, etc. Some of them have already carried out preliminary field trials which demonstrated great potential of RIS in realistic environment. There have been many symposiums, seminars and webinars on RIS, participated by both academia researchers and practical engineers. RIS is an interdisciplinary effort which requires not only the expertise in mobile communications but also the deep knowledge in meta-surface material. Depending on the rate of amplitude and phase adjustment of RIS elements, RIS can be used either as relays deployed in the path connecting base stations and mobile terminals, or as part of antennas/radio fre-

quency components within base stations. Our focus in this paper is the relay scenario of RIS application.

To facilitate the discussion, we use Fig. 1 to represent a basic RIS-based downlink passive relay system where the base station (BS) is equipped with  $M$  transmit antennas. The user equipment (UE) has  $N$  receive antennas. The channel connecting the BS and the UE is denoted as  $H_d$ . The other route goes through a RIS panel that has  $K$  reflecting elements. The channel connecting the BS and RIS is denoted as  $H_1$ . The channel connecting the RIS and UE is denoted as  $H_2$ . The received signal at the UE can be represented as:

$$y = (H_d + H_2\Phi H_1)V s + n_1, \tag{1}$$

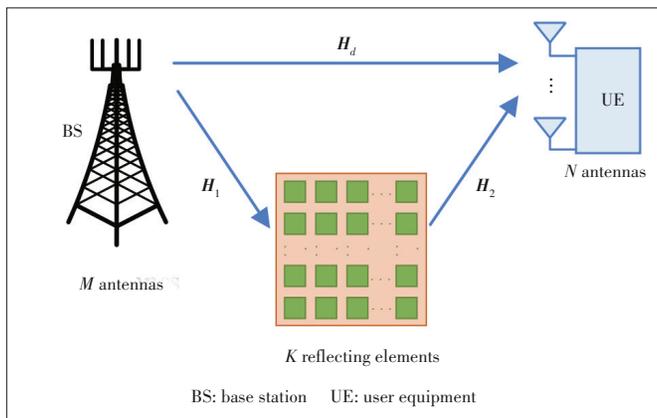
where  $\Phi$  is a diagonal matrix in which each diagonal element represents the amplitude and phase response of a RIS element. The matrix  $V$  is the precoder applied at the BS transmit antennas. The vector  $s$  contains the modulated symbols to be transmitted. The vector  $n_1$  is the noise vector received by  $N$  antennas at the UE. Note that this model is quite general in the sense that the structure of RIS, as well as the antenna configurations of BS and UE, is not explicitly modeled. More intricate models would be needed for more practical performance studies.

The paper is organized as follows. In Section 2, progress in RIS research in academia is reviewed, including several major areas of research. Recent development of RIS in industry is discussed in Section 3 from multiple aspects. Section 4 contains our views on some of the key components that would be important to future standardization and deployment of RIS. Section 5 concludes this paper.

## 2 Recent Research on RIS in Academia

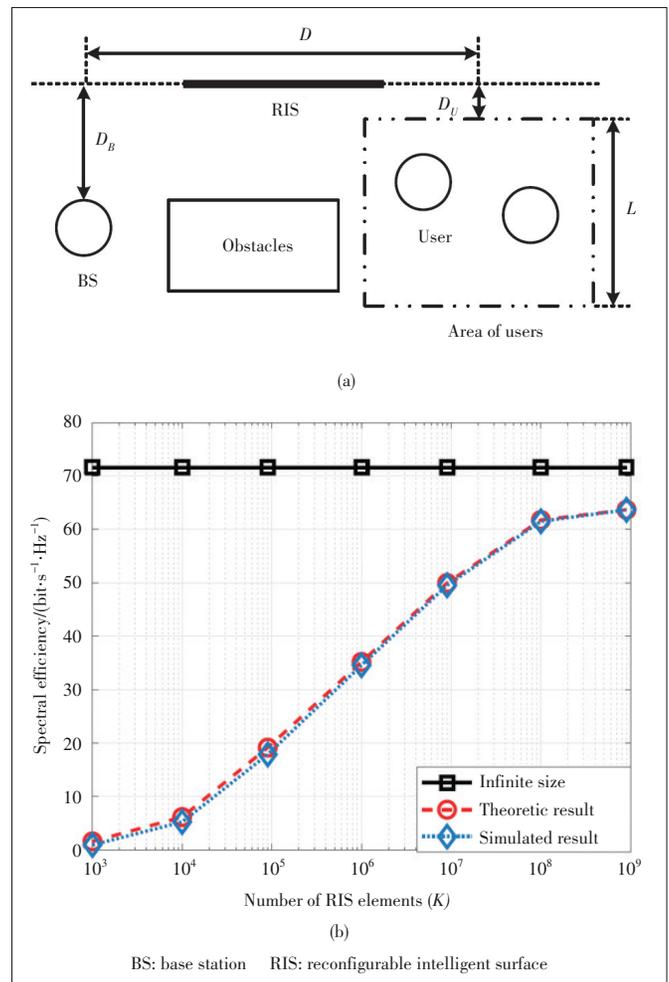
### 2.1 Fundamental Performance of RIS

As RIS represents a paradigm change of traditional MIMO technology, it is worthwhile to study its fundamental performance to develop some understanding of its potential, even



▲ Figure 1. A basic reconfigurable intelligent surface (RIS)-based relay setup with multiple transmit antennas at BS and multiple receive antennas at UE

though the initial assumptions are quite idealistic. In Ref. [4], the authors studied the number of RIS elements (i.e.,  $K$  in Fig. 1) needed to show the benefit of RIS. The study assumed a setup with a BS and multiple users located in a square-shaped area as illustrated in Fig. 2(a). There are obstacles between the BS and users. A RIS panel is placed between the BS and users to reflect the signal from the BS to the users, and vice versa. Closed-form formulae were derived for the system capacity of this setup, as well as the asymptotic capacity when  $K$  goes to infinity. It is seen from Fig. 2(b) that the spectral efficiency increases with the number of RIS elements, when the number of single-antenna users is five, the number of antennas at the BS is 10 (i.e.,  $M = 10$  in Fig. 1), the carrier frequency is 5.9 GHz and the path loss exponential is around 2.5. The trend shows sign of saturation when  $K$  goes beyond  $10^7$ , translated to the side of the RIS square of 56.7 m. With such an aperture size, the RIS is operating well within the near-field range. The study is very idealistic in the sense that the scenario is for single-cell without considering inter-cell interference or inter-



▲ Figure 2. Setup in Ref. [4]: (a) simulation layout for the RIS-based multi-user cellular network and (b) spectral efficiency vs. the number of RIS elements

RIS interference, that optimal phase adjustment is assumed without considering the phase granularity and control signaling overhead, that channel estimation is perfect, and that the channel state information (CSI) is precisely known. Nevertheless, the result does reveal the importance of having enough number of RIS elements in order to deliver enough good performance.

While increasing the number of RIS elements (i. e.,  $K$  in Fig. 1) is beneficial to the system capacity, the cost of manufacturing and deployment of RIS panels are increased as well. Besides, the power consumption would generally grow with larger panels of RIS even if RIS is a quasi-passive device. In Ref. [5], the optimum value of  $K$  is derived to balance the spectral efficiency and energy efficiency, for a single cell system equipped with single antenna at both the BS and the mobile terminal. Control overhead is also considered, which accounts for about 20% of total time/frequency resources. Under the normal setting of the parameters such as 30 – 45 dBm transmitting power of the BS and 110 dB of the path loss, it is found that the spectral efficiency of 20 – 25 bit · s<sup>-1</sup> · Hz<sup>-1</sup> can be achieved when  $K$  is around 130 – 170.

## 2.2 Channel Estimations

Compared to traditional BS and mobile connections, RIS when serving as relays introduces more hops along the track from the BS to a mobile terminal. Even for the single reflection scenario as shown in Fig. 1, two links need to be considered, ending up with a cascaded channel that is more complicated and harder to estimate accurately. Channel estimation becomes very challenging for RIS containing hundreds or thousands of elements, yet the overhead of signals to estimate RIS channels has to be kept small. Considering that a BS is usually equipped with multiple antennas and the link connecting BS and RIS (i. e.,  $H_1$ ) is changing slowly, a two-time scale channel estimation was proposed in Ref. [6] where the estimation of  $H_1$  is concentrated in the time slot with heavy density of pilot of both downlink and uplink, while the estimation of the link connecting RIS and the mobile (i. e.,  $H_2$ ) is carried out via the pilots residing in time scattered slots (Fig. 3). The pilot

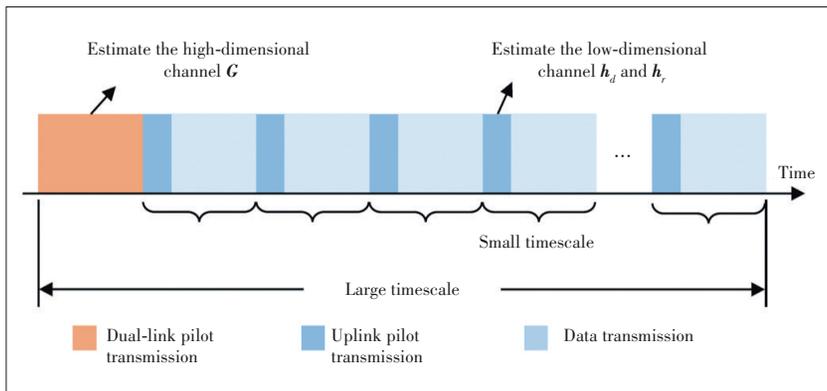
concentration ensures the accurate channel estimation of  $H_1$  that is usually high-dimensional because of large  $M$  and  $K$ . The relatively long period of concentrated pilot region can reduce the overhead without significantly affecting the estimation accuracy due to the slow fading in  $H_1$ . More frequent scattered pilots can better track the time evolution of RIS-mobile channel (i. e.,  $H_2$ ). The light density of pilots in scattered slots helps to reduce the overhead without degrading the channel estimation, since the number of receive antennas at the mobile (i. e.,  $N$  in Fig. 1) is usually small.

Assuming that mobiles are equipped with a single antenna, a three-phase channel estimation framework was proposed in Ref. [7]. In Phase 1, the direct link connecting BS and the mobile (i. e.,  $H_d$  in Fig. 1) is estimated while the RIS is switched off. In Phase 2, only RIS and the typical mobile are turned on, so that the cascaded channel  $H_2\Phi H_1$  can be estimated. In Phase 3, all other users are turned on to transmit the pilots. Assuming that the other users' reflected signals are scaled-version of the reflected signal of typical users, the cascaded channels of other users can be estimated efficiently.

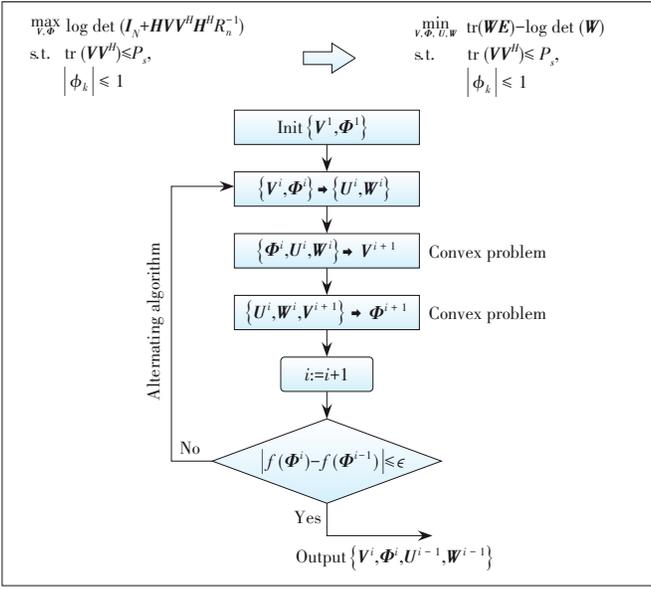
While the BS-RIS channel (i. e.,  $H_1$ ) is often considered high-dimensional, the location of RIS is expected to be optimized in practical deployment so that line-of-sight (LoS) propagation would be dominant. In this case, the cascaded channel would exhibit certain spatial sparsity especially in the millimeter band<sup>[8]</sup>. Compressive sensing-based algorithms can be used to improve the efficiency of channel estimation. More advanced channel estimation methods would involve the iterative process to estimate  $H_1$  and  $H_2$ , respectively. In general, RIS channel estimation algorithms, setting and configurations highly depend on channel models and operation modes.

## 2.3 Joint Optimization with BS Precoding

As Fig. 1 shows, both the BS and RIS would have multiple antennas or tunable elements, leading to very complicated joint optimization of BS precoding and phase control of RIS elements. In Ref. [9], an alternating weighted minimum mean squared error (WMMSE) algorithm was employed for MIMO of a RIS-aided single user. Here, the single user is in the sense that the BS precoder (i. e.,  $V$ ) and element phase of RIS (i. e.,  $\Phi$ ) are optimized to maximize the link capacity of BS to a specific mobile (including both the direct link and RIS-relayed link), without considering other users. The optimization for the capacity formula is non-convex. Instead of being solved directly, the problem is solved alternately and iteratively as shown in Fig. 4 where each sub-problem is convex. Here, matrix  $U$  is the receive beamforming matrix,  $E$  is the mean squared error (MSE) matrix, and  $W$  is the weighting matrix. As the iteration goes on, it would converge to the optimized  $V$  and  $\Phi$ . It is



▲ Figure 3. A two-time scale channel estimation for reconfigurable intelligent surface (RIS) aided system



▲ Figure 4. An alternating weighted minimum mean squared error (WMMSE) algorithm for joint optimization of reconfigurable intelligent surface (RIS) elements with base station (BS) precoding

seen that even convergence can be achieved by converting the non-convex problem to a convex problem, multiple iterations are required, each involving a bunch of matrix multiplications. The overall computation complexity is high. It is expected that for RIS-aided multi-user MIMO, the optimization would be even more complicated.

The difficulty and complexity shown above are partially due to the unconstrained phase shifts of RIS elements (i.e.,  $\Phi$ ) and BS precoder (i.e.,  $V$ ). Yet the practical RIS elements have limited resolution in phase adjustment. To reflect such limitation, the phase response of each RIS element is modeled as ON and OFF in Ref. [10]. The number of receive antennas at each mobile is assumed one (i.e.,  $N = 1$  in Fig. 1) to encourage more multi-user MIMO transmission. With these assumptions and setting, the complicity of the joint optimization can be significantly reduced, although an alternating type of algorithm is still used. The joint optimization problem can also be simplified by considering specific scenarios. For instance, millimeter wave (mmW) was the focus for RIS in Ref. [11]. Due to the less scattering in mmW propagation, the ranks of BS-RIS channel (i.e.,  $H_1$  in Fig. 1) and RIS-mobile channel (i.e.,  $H_2$ ) tend to be small. Hence, RIS element phase shifts (i.e.,  $\Phi$ ) can be optimized separately from the BS precoder (i.e.,  $V$ ).

In Ref. [12], the joint optimization problem also counts in the non-ideal channel estimation. More realistic BS precoders, such as Hadamard-matrix based and discrete Fourier transform (DFT) based codebooks, are considered. With these practical constraints, the optimization can be simplified to some extent without significantly degrading the performance. In Ref. [13], the joint optimizing problem between RIS and BS was extended to multi-cell environment where many assump-

tions had to be simplified to make the problem tractable.

## 2.4 Channel Modeling

As an emerging technology, RIS is bringing significantly new operations of wireless networks with many unknowns, one of which is the channel characteristics of the cascaded channel. To certain extent, the difficulty of modeling the RIS channel lies in the relatively large size of RIS panels (which is typically required to achieve enough good performance as indicated in Section 2.1), the large number of RIS elements (i.e.,  $K$  in Fig. 1), and the deterministic phase shifts of RIS elements (i.e.,  $\Phi$ ). Hence, a RIS panel cannot be treated as a cluster of scatterers as in traditional statistic channel models. The plane wave assumption which generally holds for traditional models may no longer be valid when the aperture of RIS panel is larger enough to meet the near-field criterion. There are some initial works<sup>[14-15]</sup> on RIS channel modeling based on Friis formula for line-of-sight propagation. For each RIS element (indexed by the  $n$ -th row and  $m$ -th column), the receive radiation pattern and transmit radiation pattern are included in the calculation as seen in the second and third terms.

$$F_{n,m}^{combine} = (\cos\theta_{n,m}^x)^{\frac{G_x}{2}-1} (\cos\theta_{n,m}^t) (\cos\theta_{n,m}^r) (\cos\theta_{n,m}^x)^{\frac{G_x}{2}-1} = \left( \frac{(d_1)^2 + (r'_{n,m})^2 - (d_{n,m})^2}{2d_1 r'_{n,m}} \right)^{\frac{G_x}{2}-1} \left( \frac{z_t}{r'_{n,m}} \right) \left( \frac{z_r}{r_{n,m}} \right) \left( \frac{(d_2)^2 + (r_{n,m}^r)^2 - (d_{n,m})^2}{2d_2 r_{n,m}^r} \right)^{\frac{G_x}{2}-1}, \quad (2)$$

where  $d_1$  is the distance between the transmitter and center of the RIS panel;  $d_2$  is the distance between the receiver and center of the RIS panel;  $d_{n,m}$  is the distance between the  $n$ -th row  $m$ -th column element and the center of the RIS panel;  $r'_{n,m}$  is the distance between the  $n$ -th row  $m$ -th column element and the transmitter;  $r_{n,m}^r$  is the distance between the  $n$ -th row  $m$ -th column element and the receiver. For simplicity, only elevation angle  $\theta$  is considered in this equation. The entire gain for the cascaded channel can be represented as:

$$P_r = P_t \frac{G_t G_r G_d d_x d_y \lambda^2}{64\pi^3} \left| \sum_{m=1}^M \sum_{n=1}^N \frac{\sqrt{F_{n,m}^{combine}} \Gamma_{n,m}}{r_{n,m}^t r_{n,m}^r} e^{-j2\pi(r'_{n,m} + r_{n,m}^r)/\lambda} \right|^2, \quad (3)$$

where  $d_x$  and  $d_y$  are the width and length of each RIS element, respectively;  $N$  and  $M$  are the total number of rows and columns of the RIS panel, respectively. While the above study is only applicable to simple environments, those formulae help build an important foundation for more sophisticated propagation scenarios. In Ref. [16], path-loss models were proposed for indoor and outdoor RIS, based on the real measurements for both 2.4 GHz and 28 GHz bands. It is seen that under the reasonable settings, the path loss of a cascaded channel sits between ITU-UMi (International Telecommunication Union-Urban Micro) LoS model and UMi NLoS (non-line-of-sight)

model. As the number of RIS elements (i. e.,  $K$  in Fig. 1) increases, the path loss becomes closer to that of the UMi LoS model. A 3D geometry-based non-stationary MIMO model was proposed for RIS-assisted system in Ref. [17]. An intricate cascaded model for RIS was proposed in Ref. [18] where the node pruning technique was used in the graph representation of the cascaded model.

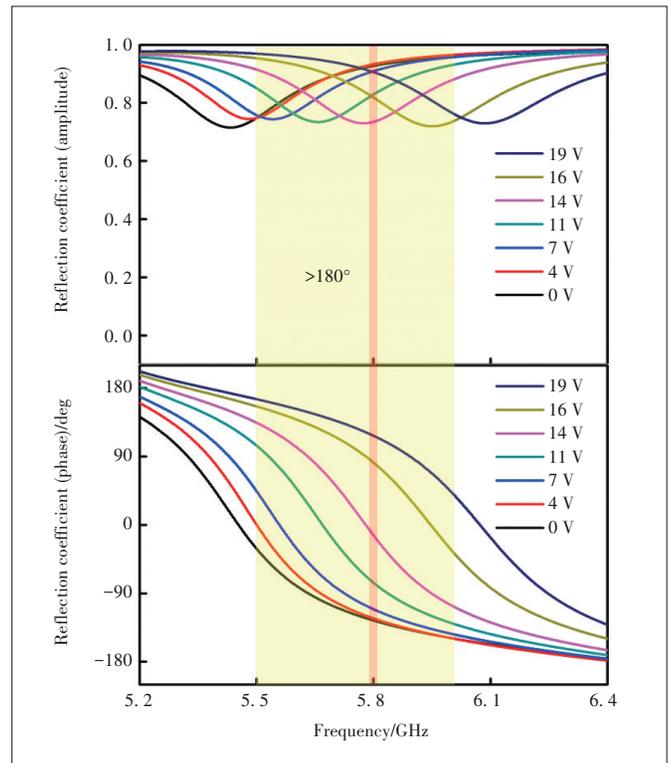
### 2.5 Meta-Material Devices

The characteristics of meta-material for RIS elements are crucial to the entire RIS system performance. Meta-material refers to a big family of man-made materials whose properties can be manipulated by introducing intricate wavelength-level or sub-wavelength-level structures at the surface. Note that the fundamental physical or chemical properties of the substances in meta-materials are not changed. For RIS elements, the most widely used devices are PN-diode or varactor based. With the fast development of computer simulation tools for electromagnetic field circuits, the maturity of RIS device manufacturing has been elevated dramatically over the last few years, especially in terms of cost and durability, which makes it possible to build less expensive large RIS panels with hundreds or thousands of elements (i. e.,  $K > 1\,000$  in Fig. 1). While the fine tuning of element phases is still difficult, granularities of 2 bits or 4 bits are well within the reach of practical engineering nowadays. For instance, a three-level phase tunable RIS device was designed in Ref. [19] for multi-beam antenna configuration at round 5 GHz band operation. In Ref. [15], two RIS panels were manufactured and employed for RIS channel model measurements at mmW bands. The first panel is rectangular of size  $7.8 \times 56 \text{ mm}^2$ , containing  $20 \times 56 = 1\,120$  elements operating at 27 GHz. The second RIS panel is a square of side length 152 mm that has  $40 \times 40 = 1\,600$  elements operating at 33 GHz. In the field trials carried in Ref. [19], the core part is a RIS panel designed by the authors. The rectangular panel operates at 5.8 GHz, built with varactor-based RIS devices. There are  $20 \times 55 = 1\,100$  elements and the phase of each element can be tuned to  $\sim 7$  levels by applying different voltages as illustrated in Fig. 5. No big fluctuation in the amplitude response is observed across the frequency range of interest and the applied voltages.

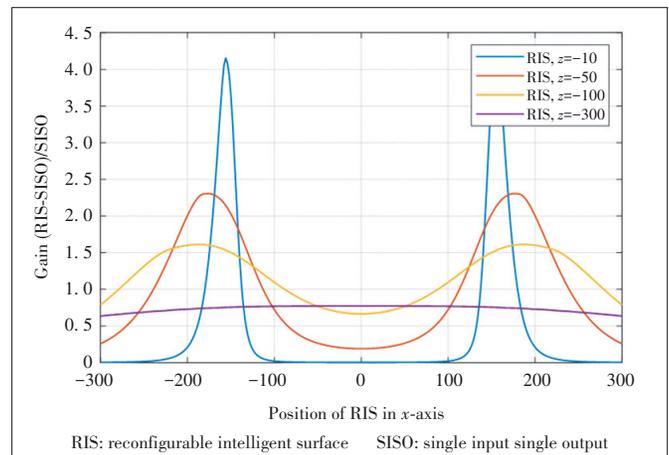
## 3 Development in Industry

### 3.1 Performance Potential and Issues

Performance evaluations have been carried out by a number of companies. With the basic setup as shown in Fig. 1, we investigated the link level performance gain of RIS compared to that of the direct link, as illustrated in Fig. 6. Single antenna BS and single antenna mobile are assumed here, e. g.,  $M = N = 1$ . The RIS is square-shaped, containing 400 phase-tunable elements (e. g.,  $K = 400 = 20 \times 20$ ) of half-wavelength spacing. The carrier frequency is 3 GHz. The antenna gains of BS, RIS element and



▲ Figure 5. Amplitude and phase response of reconfigurable intelligent surface (RIS) elements vs. frequency and applied voltages<sup>[19]</sup>



▲ Figure 6. RIS performance gains over direct link as a function of distance

UE are 5 dB, 5 dB and 0 dB, respectively. LoS propagation is assumed between BS and RIS, and between RIS and UE. For the direct link (e. g., BS and UE), the path loss is set so that the SNR equals 6 dB. Both mechanical tilt and phase adjustment are applied where the mechanical tilt (referenced at the center of the RIS panel) is determined by the Snell's Law of reflection, and the phase adjustment is calculated according to the distance of an element to the center of the RIS panel. It is seen that in order to maximize the performance, the RIS panel should be placed close to either the BS or the mobile. Such phenomenon was also observed by ZTE. It can be explained as

follows. When a RIS panel is close to a BS or a mobile, either BS-RIS link (e.g.,  $H_1$  in Fig. 1) or RIS-mobile link (e.g.,  $H_2$  in Fig. 1) is in the range of near field, as long as the RIS panel has a reasonable aperture size. In this case, the effective propagation distance of the cascaded channel tends to be the sum of propagation distances of  $H_1$  and  $H_2$ , rather than the multiplication of propagation distances of  $H_1$  and  $H_2$ , if element phases of RIS can be adjusted appropriately.

System level simulations have been carried out by a few companies and the very preliminary results show the performance benefits of RIS-aided relays under the multi-cell environment. Nevertheless, the passive nature of RIS elements would cause inadvertent disturbance to the networks of other operators if the allocated bands of different operators are adjacent to each other, because RIS usually does not have filters. Thus, it is challenging to confine the impact of phase tuning within an allocated spectrum.

### 3.2 Practical Beamforming Algorithms for RIS

Considering the rather complicated joint optimization of BS precoder and phases of RIS elements which may require many iterations to reach convergence, it is desirable to come up with engineering feasible beamforming algorithms for RIS. There are several aspects that may be simplified in order to focus on key characteristics of RIS. For example, single-user transmission instead of multi-user may be prioritized, without taking into account the grating lobes. While both LoS and NLoS components may be present in the channel, LoS may be emphasized, leading to low rank transmission and grid of beam (GoB) type of optimization. In another word, the signal to noise ratio (SNR) would be the main target for optimization, instead of the sum rate. In some situations, it may be wise to give up the full dynamic tuning of phases and opt for semi-static beamforming of RIS, as long as most of the performance of gains of RIS can be reaped.

### 3.3 Control Mechanisms of RIS

As mentioned previously, the RIS we considered in this paper serves as a passive relay without the power amplifier. Nevertheless, certain power consumption is still expected in order to adjust the phases of RIS element, to perform necessary digital processing and/or signal detection, etc. In such case, the RIS can be controlled in the following three ways.

The first one is through a wire connecting BS and the RIS controller. The advantage of wired connection is the simplified control protocol and lower power consumption. However, wires need to be reserved in this case, which would impose limitations on deployment. The wired connection can be either Xn protocol based which was designed specifically for backhaul of mobile networks and normally has higher quality, or via Internet Protocol (IP) routers that can be more easily plugged into the existing networks not necessarily built for radio access networks.

The second alternative is via wireless link using a Uu air interface (connecting BS and UE) or non-Uu air interface. The major advantage of wireless connection between BS and the RIS controller is the more flexible deployment of RIS. Nevertheless, the Uu interface for RIS control requires standardization work which would take time to study and specify in standardization development organizations (SDOs), and may not be able to reach consensus in short time. Power consumption is also assumed higher than the wired connection, given the higher propagation loss in the air and the less percentage of emitted electric-magnetic signal actually received by the mobile. There are also some issues with real-time control via wireless links.

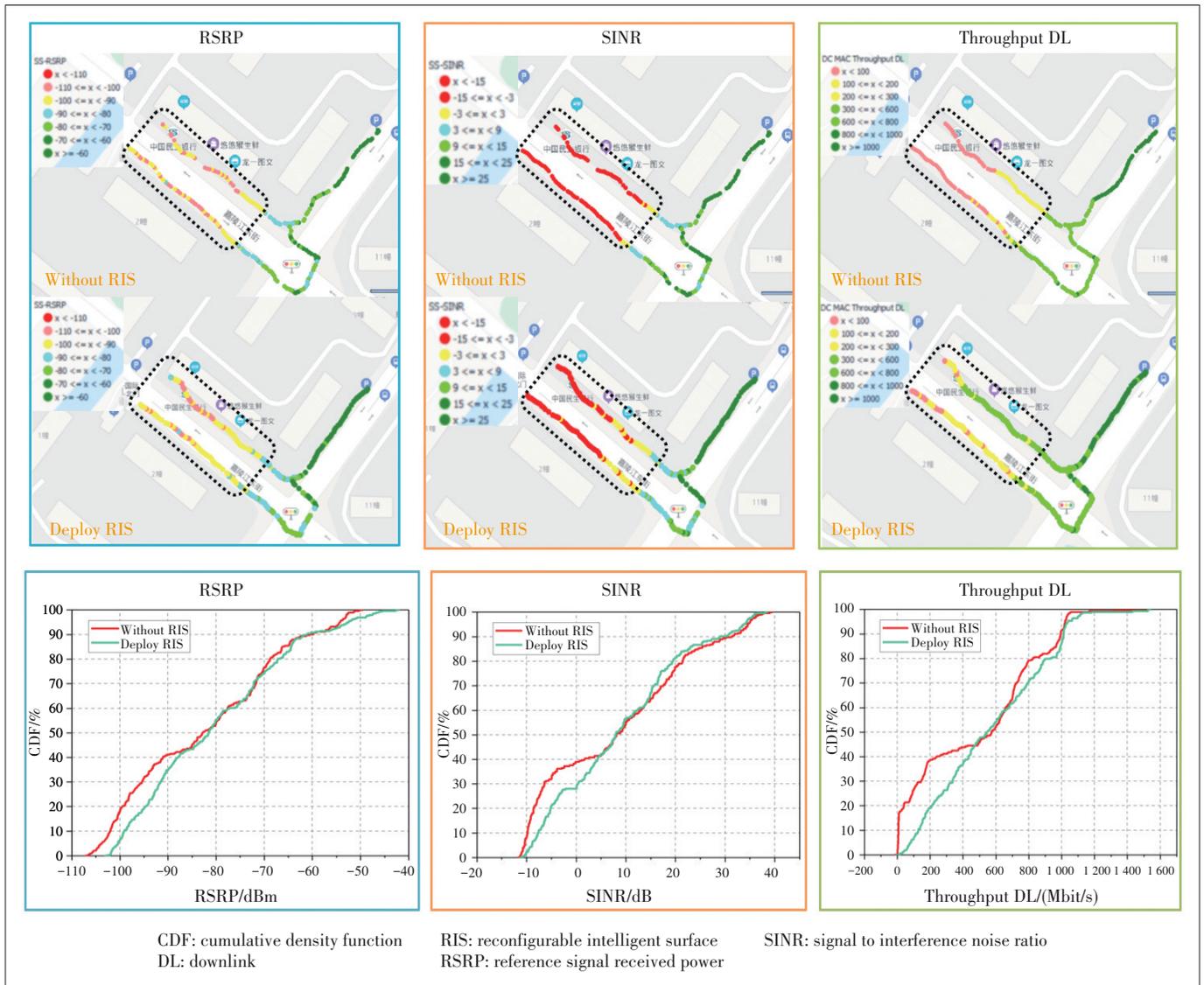
The third approach is self-control/sensing of RIS which has the advantage of very flexible deployment and self-reliance. The self-control would be based on information sensing to proactively measure the ambient environment, including the channels of BS-RIS and RIS-mobiles. Such a sensing based approach has the issues of high power consumption and performance sensitivity to sensing capability and accuracy. The joint optimization with BS would be more challenging.

### 3.4 Initial Field Trials

Several companies have carried out initial trials of RIS. One of them is China Mobile, teaming up with Southeast University. Their trial took place at Nanjing, in the networks of China Mobile Jiangsu Province, which represents a dense urban environment. The operating band is 2.6 GHz. The environment includes both outdoor and indoor. The RIS panel has the size of  $160 \times 80 \text{ cm}^2$  and contains  $32 \times 16 = 512$  elements ( $K = 512$  in Fig. 1). Both the azimuth and vertical directions of RIS beams can be tuned, with 1-bit phase resolution. The RIS beam is manually tuned according to the BS beams and geo-locations of the target UE, as well as the nearby propagation environment. Fig. 7 shows the reference signal received power (RSRP), signal to interference noise ratio (SINR) and down-link (DL) throughput measured along the driving path, with and without RIS. Cumulative density function (CDF) curves of RSRP, SINR and DL throughput are also compared between with and without RIS. It is observed from this trial that for users in good coverage, implementing RIS does not bring much benefit. However, for users in poor coverage, RIS can bring significant gains.

Since 2018, three trials of RIS have been conducted by NTT DOCOMO. The trial in 2018 showed magnitude-level throughput improvement when RIS was deployed along the city streets. In 2020, a prototype of transparent dynamic metasurface was demonstrated, which can be coated or inserted to window panels to improve the indoor coverage. In 2021, metasurface lens with the capability of focus control were presented.

Field trials of RIS by ZTE in Shanghai showed that RIS, if properly phase tuned, could drastically improve the received



▲ Figure 7. Initial trial results of RIS by China Mobile in Nanjing

SNR. However, if the RIS was not powered, e.g., no phase tuning, the SNR would be similar to that without RIS. This means that the random phase reflection like in scatterers, would not constitute coherent phases, resulting in little benefit in the gain of SNR.

### 3.5 Related Discussions in 3GPP

Due to the strong interest in RIS from both academia and industry, a few companies like ZTE, China Unicom, KDDI, etc. have proposed a study item of RIS in 3GPP for Release 18 which is being carried out from January 2022 to June 2023. The scope of the proposed RIS study item includes use scenarios, channel models and performance evaluations of RIS. If possible, high level designs would also be discussed in the proposed study item. Still, the majority of companies in 3GPP felt that it was too early to kick off RIS which is generally con-

sidered as a 6G technology rather than a 5G+ feature. There are many fundamental issues requiring more time for study in academia or in the industry. Many of them are not in the hand of experts in wireless baseband communications, for instance, the developments in device manufacturing and channel models of RIS, are also crucial. Once RIS is in the scope of 3GPP, the pace of its study would unavoidably be driven by tight schedules of delivering, which may not be beneficial to explore the full potential of RIS that would only be possible after a thorough analysis.

It is also noticed that in 3GPP Release 17 there was a work item in the RAN4 working group on the performance requirements for smart repeaters in 5G mobile networks. Hence, several companies like China Mobile, Qualcomm, etc. proposed a study item of smart repeaters in the RAN1 working group (responsible for physical layer specifications) to investigate the

control mechanisms for the repeaters. Different from the traditional repeaters that simply receive and amplify signals, smart repeaters can coordinate with the BS with regard to the time division duplex (TDD) subframes and beams for transmissions. In RAN1, system level simulations for smart repeaters are expected, together with the discussion on side control information between BS and RIS for the subframe and beam-forming coordination. Compared with RIS, the repeater is a more mature technology in the wireless. Smart repeater is just a small step further, with relatively small impact on the standards. Hence, it is likely that 3GPP Release 18 may include smart repeaters.

## 4 Views on RIS Study and Development

### 4.1 Performance Evaluation

It is expected that RIS panels would be deployed ubiquitously in future networks where multiple RIS panels would be employed in a cell to improve both the coverage and data throughput. The performance should be evaluated in a multi-cell (Fig. 8) and multi-user simulation setting, so that the inter-RIS and other cell interferences can be accurately captured. Considering that each RIS may contain hundreds of phase-tunable elements, the entire multi-cell and multi-RIS system level simulation can be extremely complex, even when a relatively simple channel model is assumed. Certain simplifications are needed to accelerate the computations in system simulations.

The simulations should reflect the coordination between the spatial precoding at BS in downlink (or at the mobile in uplink) and phase tuning at the RIS panel. At the link level, the joint optimization algorithm to be simulated should be engineering feasible, and can be modeled rather precisely in the

physical layer abstraction (also called link-to-system mapping). The optimization algorithms should consider different levels of control over the RIS panel, such as granularity of phases or amplitudes of RIS elements and the time duration for the control (dynamic or semi-static).

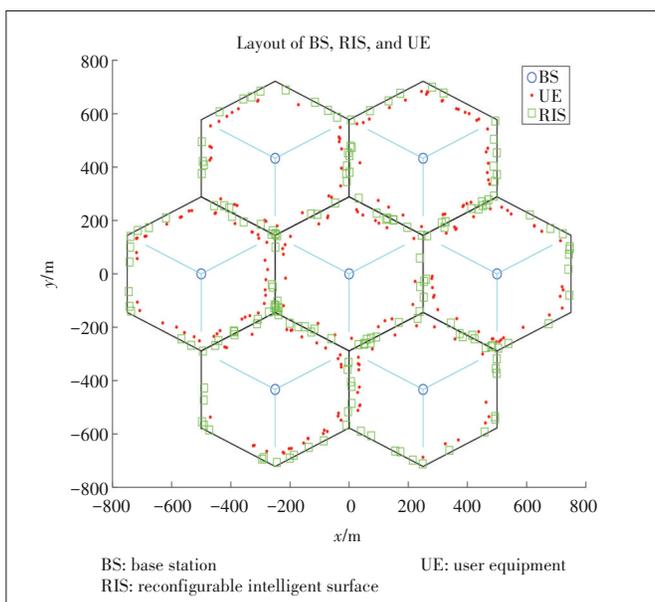
The performance evaluation is not only for downlink and uplink traffic channels, which is crucial to quantify the potential throughput gains of RIS. The performance of control channels, reflecting the basic coverage, should also be thoroughly simulated which include physical broadcast channel, primary and secondary synchronization channel, physical downlink control channel, preamble and messages for random access, etc.

Overall, a new set of simulation methodology is needed for RIS, especially at the system level, to comprehensively balance the simulation complexity and the unique characteristics of RIS systems. It should reflect typical deployment scenarios envisioned by operators, as well as suitable channel models.

### 4.2 Comparisons with Other Topological Improvements

In addition to the widely deployed homogeneous macro cell layout where the site-to-site distance, the antenna gains, the transmit power rating, etc. are more or less the same between different cells, heterogeneous networks (Het-Net) are often necessary to effectively serve sporadic hot-spot areas or fill the coverage holes, by adding stationary lower power nodes on top of the macro layer. Such Het-Net is sometimes called topological improvement which can take the forms of micro BS, pico-BS or femto-BS, repeater, relays, etc. All of them are active nodes, meaning that they have power amplifiers.

Since 2010, 3GPP has conducted extensive studies on these low power nodes, with abundant performance evaluation results which showed the significant gains in system throughput. The studies also revealed the inter-cell interference issues that require practical solutions either in pure implementation or with some impact on air interface standards. For instance, the introduction of micro/pico/femto cells causes severe interference from the macro cell to low power nodes and leads to significant coverage imbalance between the downlink and uplink. The coverage imbalance issue can be mitigated by specifying almost blank subframe (ABS). As one of the simplest low power nodes, a repeater can only amplify and then forward the signals from BS or mobiles, to mobiles or BS. Operations of repeaters are generally implementation specific and rarely require air interface changes. Relay, studied and specified in Release 10 Long-Term Evolution (LTE) of 3GPP, is essentially a micro/pico/femto node with wireless backhaul. The backhaul link (e.g.,  $H_1$  in Fig. 1) and access link (e.g.,  $H_2$  in Fig. 1) operate in the same band in time-division multiplexed (TDM) manner. Different from repeaters, a relay is typically able to decode the received signal, re-encode and send the sig-



▲ Figure 8. A multi-cell and multi-RIS layout for RIS system level simulation

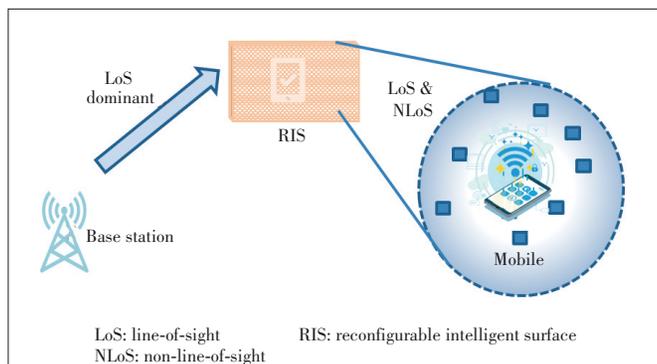
nal out, thus expected to have better performance. However, the standardization of the backhaul took a lot of effort, and the corresponding air interface specification was rather cumbersome, hampering its wide use. So far, very few networks have deployed relays.

While RIS is a passive device that does not have power amplifier and can be more flexibly deployed, RIS shares many deployment scenarios with relays and repeaters. For instance, all of them would likely be deployed by operators and their locations would be optimized to have higher probability of LoS propagation. Many experience and lessons from relay standardization can be helpful for RIS study and future standardization: 1) some simulation setup and evaluation methodology for relay may be reused for RIS; 2) technical solutions for RIS systems should be efficient and practical, e.g., to avoid complicated or redundant schemes.

The study on smart repeaters can provide a very good reference for future RIS standardization. Firstly, the deployment scenarios of smart repeaters would be overlapped with those of RIS. Hence, the channel characteristics of backhaul link of smart repeaters and RIS would share many similarities in terms of antenna height, LoS probability, path loss, shadow fading, etc.; Secondly, certain control signaling is expected between BS and smart repeaters to facilitate the coordination. This signaling is likely to be sent over the air. Some control signaling is also anticipated between BS and RIS; Thirdly, both smart repeaters and RIS have very limited baseband processing capability for baseband signals.

### 4.3 Practical Channel Models

RIS channel modeling requires extensive measurements in different scenarios with various operating bands from sub-6 GHz to mmW. Channel models generally have two categories: statistical models and ray-tracing based models. In 3GPP, statistical models<sup>[20]</sup> are more widely used, while the ray-tracing based models are optional. The reason is that statistical models tend to generate the channel coefficients with more consistent and universal characteristics, rather than targeting a specific layout of BS, scatterers and mobiles. The computation of statistical models is typically less burden-



▲ Figure 9. An example of statistical model for RIS cascaded channel

some. These traits are crucial for the research study in SDOs such as 3GPP, to 1) have a generic and comprehensive model to reflect all typical layout and settings; 2) facilitate the cross-checking of simulations and result comparisons between companies. In this sense, extension of statistical models to the RIS scenario may be a more practical approach, as illustrated in Fig. 9 where the backhaul link channel is simplified to be LoS dominant. It should be noted that many new parameters may need to be added, even if statistical models are to be used, considering the cascaded channel of BS-RIS-mobile. With hundreds or thousands of RIS elements in a panel, the computation complexity of the channel model can be daunting. More efficient characterization of RIS elements should be considered.

As discussed earlier, RIS performs well when it is placed close to BS or to mobiles. In this case, certain near field effect would be seen where the electromagnetic waves would not be strictly plane, e.g., the wave front becomes spherical. This would certainly increase the complexity of channel modeling and fading coefficient generation. Hence, some simplification methods seem necessary to efficiently model the near field effect.

### 4.4 Physical Layer and Higher Layer Control Aspects of RIS

The discussion of RIS control highly depends on the mode of operation and the nature of future RIS products. For RIS panels with fixed amplitude and phase responses, the control is actually reflected during the pre-configuration when operators first lay out the panels. Such configuration heavily involves proprietary solutions of network optimizations which are out of the scope of standardization. In such case, operators can directly purchase the static RIS panels from device manufacturers. For configurable RIS, it is a type of RIS-based relay node that has both a control unit and a RIS panel consisting of a large number of phase tunable elements. The potential manufacturers of this type of nodes may be traditional wireless equipment vendors or small vendors specialized in building low power nodes or passive devices, who may have different preferences about the functionalities of the RIS node. This would have big impact on the design of control mechanisms. One aspect is whether RIS panels are transparent or non-transparent to the mobiles being served. Transparent RIS requires less control signaling and thus less standardization work, albeit with smaller room for performance optimization. Non-transparent RIS needs more procedures, more complicated signaling and deeper involvement by mobiles, with the potential benefit of higher system throughput.

In order for RIS to work properly, four key technical aspects at the physical layer should have feasible solutions: 1) channel estimation; 2) CSI measurement and exchange; 3) joint spatial precoding between BS and RIS; 4) interference management. While many intricate schemes have been proposed

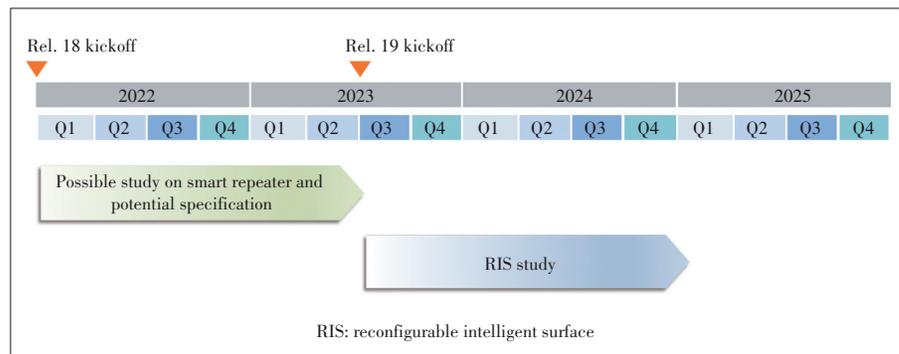
for RIS in academic literature, a large percentage of them can hardly be deployed in practical systems. They are simply too complicated and narrowly defined to be evaluated in SDOs such as 3GPP. Although the air interface standards would not specify the implementation-oriented techniques such as joint precoding and interference management, some basic or typical algorithms for implementation would be decided in SDOs to facilitate the simulations and cross-verification. The control of RIS also involves a higher layer, for instance, how to verify, register and configure a RIS node when first deployed in the network.

#### 4.5 Standardization Timeline in 3GPP

As the most important SDO in the world for mobile communications, 3GPP will be the most influential industrial platform to specify the necessary technology of RIS and foster an eco-system for wide deployment of RIS in future mobile networks across the globe. Nevertheless, as mentioned above, many companies in 3GPP, including China Mobile, have expressed views that it is immature to kick off the study on RIS in Release 18. On the other hand, we believe that it is beneficial to carry out the study on smart repeaters, as an intermediate stage for RIS, to get more understanding about the potential gain of topological enhancements and accumulate more knowledge prepared for RIS. As a revolutionary technology, RIS requires more time and experience for build-up. It should be emphasized the kickoff of RIS in 3GPP should not be too late either, considering that the enthusiasm on RIS by the academic community has already lasted for about 3 years. At a certain point, the industry community should be deeply involved, to provide clearer and consolidated feedback to academia, on which directions should have high priority. In our view, it is desirable to kick off a 3GPP study item of RIS in Release 19 starting from June 2023, as illustrated in Fig. 10. The study item may last over the entire duration of Release 19 (expected to end in December 2024) to have ample time for study and evaluation of RIS.

## 5 Conclusions

In this paper, we provide a picture of the research and development of RIS for relay use, both in academia and industry. Significant progress is observed in potential performance, channel estimation, joint precoding between BS and RIS, control mechanism, channel modeling, RIS device manufacturing, etc. Views on future work on RIS are also presented, including the comparison and experience sharing with other topological enhancement technologies, and the preferred timeline to kick off the RIS study in 3GPP.



▲ Figure 10. Time frame for possible kickoff of RIS study item in 3GPP

## References

- [1] LU Z H, YUAN Y F, WU H, et al. 5G massive MIMO enhancements [M]. Beijing, China: Telecom & Post Press, 2021
- [2] DI RENZO M, ZAPPONE A, DEBBAH M, et al. Smart radio environments empowered by reconfigurable intelligent surfaces: how it works, state of research and the road ahead [J]. IEEE journal on selected areas in communications, 2020, 38 (11): 2450 – 2525. DOI: 10.1109/JSAC.2020.3007211
- [3] IEEE. Emerging Technology Initiative of RIS [EB/OL]. [2021-11-11]. <https://www.comsoc.org/about/committees/emerging-technologies-initiatives/reconfigurable-intelligent-surfaces>
- [4] ZHANG H L, DI B Y, HAN Z, et al. Reconfigurable intelligent surface assisted multi-user communications: how many reflective elements do we need [J]. IEEE wireless communications letters, 2021, 10(5): 1098 – 1102. DOI: 10.1109/LWC.2021.3058637
- [5] ZAPPONE A, DI RENZO M, XI X J, et al. On the optimal number of reflecting elements for reconfigurable intelligent surfaces [J]. IEEE wireless communications letters, 2021, 10 (3): 464 – 468. DOI: 10.1109/LWC.2020.3034686
- [6] WEI X H, SHEN D C, DAI L L. Channel estimation for RIS assisted wireless communications—part I: fundamentals, solutions, and future opportunities [J]. IEEE communications letters, 2021, 25 (5): 1398 – 1402. DOI: 10.1109/LCOMM.2021.3052822
- [7] WANG Z R, LIU L, CUI S G. Channel estimation for intelligent reflecting surface assisted multiuser communications: framework, algorithms and analysis [J]. IEEE transactions on wireless communications, 2020, 19 (10): 6607 – 6620. DOI: 10.1109/TWC.2020.3004330
- [8] WANG P L, FANG J, DUAN H P, et al. Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems [J]. IEEE signal processing letters, 2020, 27 (5): 905 – 909. DOI: 10.1109/LSP.2020.2998357
- [9] GU Q, WU D, SU X, et al. Performance comparisons between reconfigurable intelligent surface and full/half-duplex relays [C]/94th Vehicular Technology Conference (VTC2021-Fall). IEEE, 2021. DOI: 10.1109/VTC2021-Fall52928.2021.9625201
- [10] LIN S, ZHENG B X, ALEXANDROPOULOS G C, et al. Reconfigurable intelligent surfaces with reflection pattern modulation: beamforming design and performance analysis [J]. IEEE transactions on wireless communications, 2021, 20 (2): 741 – 754. DOI: 10.1109/TWC.2020.3028198
- [11] WANG P L, FANG J, DAI L L, et al. Joint transceiver and large intelligent surface design for massive MIMO mmWave systems [J]. IEEE transactions on wireless communications, 2021, 20 (2): 1052 – 1064. DOI: 10.1109/TWC.2020.3030570
- [12] ZHOU Z Y, GE N, WANG Z C, et al. Joint transmit precoding and reconfigurable intelligent surface phase adjustment: a decomposition-aided channel estimation approach [J]. IEEE transactions on communications, 2021, 69 (2): 1228 – 1243. DOI: 10.1109/TCOMM.2020.3034259
- [13] PAN C H, REN H, WANG K Z, et al. Multicell MIMO communications relying on intelligent reflecting surfaces [J]. IEEE transactions on wireless communications, 2020, 19 (8): 5218 – 5233. DOI: 10.1109/TWC.2020.2990766
- [14] ELLINGSON S W, Path loss in reconfigurable intelligent surface-enabled channels [EB/OL]. (2019-12-14)[2021-11-11]. <https://arxiv.org/abs/1912.06759>
- [15] TANG W K, CHEN M Z, CHEN X Y, et al. Wireless communications with re-

configurable intelligent surface: path loss modeling and experimental measurement [EB/OL]. (2019-11-13)[2021-11-11]<https://arxiv.org/abs/1911.05326>

- [16] YILDIRIM I, UYRUS A, BASAR E. Modeling and analysis of reconfigurable intelligent surfaces for indoor and outdoor applications in future wireless networks [J]. *IEEE transactions on communications*, 2021, 69 (2): 1290 – 1301. DOI: 10.1109/TCOMM.2020.3035391
- [17] SUN G Q, HE R S, MA Z F, et al. A 3D geometry-based non-stationary MIMO channel model for RIS-assisted communications [C]//94th Vehicular Technology Conference (VTC2021-Fall). IEEE, 2021. DOI: 10.1109/VTC2021-Fall52928.2021.9625374
- [18] DOU J W, CHEN Y J, ZHANG N, et al. On the channel modeling of intelligent controllable electro-magnetic surface [J]. *Chinese journal of radio science*, 2021, 36 (3): 368 – 377 (in Chinese). DOI: 10.12265/j.cjors.2020195
- [19] PEI X L, YIN H F, TAN L, et al. RIS-aided wireless communications: prototyping, adaptive beamforming, and indoor/outdoor field trials [EB/OL]. (2019-11-13)[2021-11-11]. <https://arxiv.org/abs/2103.00534>
- [20] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz: TR 38.901 [S]. 2019

### Biographies

**YUAN Yifei** (yuanyifei@chinamobile.com) was with Alcatel-Lucent from 2000 to 2008. From 2008 to 2020, he was with ZTE Corporation as a technical director and Chief Engineer, responsible for standards and research of LTE-Advanced and 5G technologies. He joined the China Mobile Research Institute in 2020 as a Chief Expert, responsible for 6G. He has extensive publications, including seven books on LTE-Advanced and 5G. He has over 60 granted patents.

**GU Qi** received the B.E. degree from Civil Aviation University of China, China, in 2013, and the M.S. and Ph.D. degrees from Beijing Jiaotong University, China in 2015 and 2020, respectively. She is currently an engineer with the China Mobile Research Institute. Her research interests include reconfigurable intelligent surface.

**WANG Anna** received the B.Sc., M.Sc., and Ph.D. degrees from Southeast University, China in 2003, 2006, and 2010, respectively. In June 2010, she joined the China Mobile Research Institute, where she was involved in network optimization and the development of base station antennas, indoor distributed antennas and other types of antennas for specific scenarios, which are widely used in the commercial 4G and 5G networks of China Mobile. Since 2020, her research interests have been focused on the area of 6G networks, e.g., reconfigurable intelligent surface. She has authored 14 journal articles and holds over 30 patents.

**WU Dan** is a senior member of technical staff in the Future Research Laboratory, the China Mobile Research Institute. She received the B.S. and Ph.D. degrees from Beihang University, China in 2009 and 2015, respectively. Her research interests include wireless communications, MIMO, interference management, and reconfigurable intelligent surface.

**LI Ya** received the Ph.D. degree from Beihang University, China and joined the China Mobile Research Institute in 2020. He is currently the member of technical staff of the Future Research Laboratory. His research interests include reconfigurable intelligent surface and advanced materials.

## New Member of ZTE Communications Editorial Board >>>



**HE Yejun** received the Ph.D. degree in information and communication engineering from Huazhong University of Science and Technology (HUST), China in 2005. From 2005 to 2006, he was a research associate with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, China. From 2006 to 2007, he was a research associate with the Department of Electronic Engineering, Faculty of Engineering, The Chinese University of Hong Kong, China. In 2012, he was a visiting professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. From 2013 to 2015, he was an advanced visiting scholar (visiting professor) with the School of Electrical and Computer Engineering, Georgia Institute of Technology, USA.

Dr. HE has been a full professor with the College of Electronics and Information Engineering, Shenzhen University, China, where he is currently the director of the Guangdong Engineering Research Center of Base Station Antennas and Propagation, and the director of the Shenzhen Key Laboratory of Antennas and Propagation. He was listed as “World’s Top 2% Scientists 2021” (Career-long impact). He is the Principal Investigator for over 30 current or finished research projects, including NSFC of China and so on. His research interests include wireless communications, antennas, and radio frequency. He is an IET Fellow (2016), IEEE Senior Member (2009), and the Chair of the IEEE Antennas and Propagation Society-Shenzhen Chapter. He was selected as a Pengcheng Scholar Distinguished Professor, Shenzhen. He was the recipient of the Shenzhen Overseas High-Caliber Personnel Level B and Shenzhen High-Level Professional Talent. He is currently an associate editor of *IEEE Transactions on Antennas and Propagation*, *IEEE Antennas and Propagation Magazine*, *IEEE Network*, and so on. He was the General Chair for IEEE ComComAp2019 and the TPC Chair for IEEE ComComAp2021.

# Some Observations and Thoughts about Reconfigurable Intelligent Surface Application for 5G Evolution and 6G



HOU Xiaolin<sup>1</sup>, LI Xiang<sup>1</sup>, WANG Xin<sup>1</sup>, CHEN Lan<sup>1</sup>,  
SUYAMA Satoshi<sup>2</sup>

(1. DOCOMO Beijing Communications Laboratories Co., Ltd., Beijing  
100190, China;  
2. NTT DOCOMO, INC., Tokyo 239-8536, Japan)

DOI: 10.12142/ZTECOM.202201003

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220217.1159.002.html>,  
published online February 17, 2022

Manuscript received: 2021-12-14

**Abstract:** Reconfigurable intelligent surface (RIS) is one of the hottest research topics for 5G evolution and 6G. It is expected that RIS can improve the system capacity and coverage with low cost and power consumption. This paper first discusses typical applications of RIS for 5G evolution and 6G, including RIS-aided smart channels and RIS-aided mega multiple-input multiple-output (MIMO). Then, several observations from RIS trials and system-level simulations are presented, especially those on the deployment strategy and the potential performance gain of RIS for coverage enhancement. The near-field effect and a two-step dynamic RIS beamforming method are also discussed. Finally, we summarize the challenges and opportunities of the RIS technology for 5G evolution and 6G, including hardware design, system and channel modeling, algorithm design and optimization, and standardization. We also suggest a step-by-step commercialization strategy as a conclusion.

**Keywords:** 5G evolution; 6G; beamforming; coverage; RIS

**Citation** (IEEE Format): X. L. Hou, X. Li, X. Wang, et al., "Some observations and thoughts about reconfigurable intelligent surface application for 5G evolution and 6G," *ZTE Communications*, vol. 20, no. 1, pp. 14 - 20, Mar. 2022. doi: 10.12142/ZTECOM.202201003.

## 1 Introduction

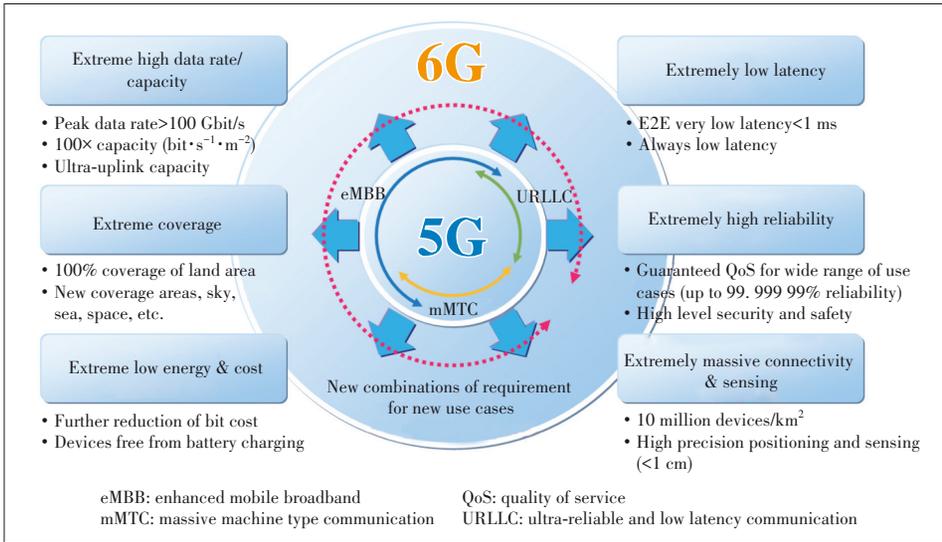
By 2020, the 5G mobile communication system has been commercialized in many countries, including Korea, China, Japan, the U. S., etc. 6G research and development activities have been initiated and are heating up worldwide. Many companies and research institutions have issued 6G white papers. It is expected that 6G standardization will be launched around 2025 and 6G commercialization will be achieved around 2030. Now it is still in an early stage of 6G research and development. Discussions are mainly related to use cases, requirements, and technology trends. For 6G, some new disruptive services and applications are foreseen with the fast development of technologies such as terahertz (THz), reconfigurable intelligent surface (RIS), satellite communications, artificial intelligence (AI)/deep learning (DL), etc. The most appealing 6G use cases include holographic-type communication, digital twin, tactile Internet, full coverage, ubiquitous intelligence, etc.

To realize the promising 6G use cases, NTT DOCOMO has proposed six extreme requirements, as shown in Fig. 1, for 6G radio access technologies<sup>[1]</sup>. These extreme requirements consist of extremely high data rates and high capacity, extremely

coverage extension, extremely low power consumption and cost reduction, extremely low latency, extremely high-reliable communication, and extremely massive connectivity and sensing. Note that these extreme requirements are driven not only by further extensions but also by new combinations of 5G use cases, including enhanced mobile broadband (eMBB), ultra-reliable and low latency communication (URLLC), and massive machine type communication (mMTC).

RIS is one of the hottest research topics on candidate technologies for 5G evolution and 6G. In recent years, RIS has emerged as a prominent technology in mobile communications and has attracted worldwide attention from both academia and industry. Intensive research has been conducted covering both theoretical and implementation aspects, e.g., hardware design and prototyping, system modeling and optimization, performance analysis and trials, as well as standardization activities<sup>[2-6]</sup>. The development of the RIS technology is originated from advances of meta material combined with the antenna and array theory, and is inclined towards flexibility, configurability, and intelligence. In principle, RIS can reradiate the incident electromagnetic wave as desired to some extent. This is achieved by pre- or reconfiguring the electromagnetic characteristics of RIS elements. The reconfigurability can be realized in different ways, including motors, micro-

Part of this paper was orally presented in 6G Seminar, Sep. 17, 2021.



▲ Figure 1. Extreme requirements of 6G

electromechanical systems (MEMS), PIN- or varactor-diodes, and functional materials such as liquid crystal and graphene. This enables the potential application of RIS in various fields, e.g., radar, communications, sensing, etc. Especially, 5G evolution and 6G could benefit from properties of RIS like low cost and power consumption, superior performance gain, and great flexibility in deployment. Hence, RIS is considered as a promising candidate technology to improve the system capacity and coverage with low cost and power consumption in the 5G evolution and 6G era.

The rest of this paper is organized as follows. Section 2 describes typical RIS applications for 5G evolution and 6G. In Section 3, several observations from trials and simulations are introduced, including the near-field effect and a two-step dynamic RIS beamforming method. Section 4 discusses the challenges and opportunities of RIS from four aspects, respectively. Finally, we conclude this paper in Section 5 with a step-by-step commercialization suggestion.

## 2 Typical RIS Applications for 5G Evolution & 6G

As for the 6G requirements, two RIS applications show great potential in future mobile communication systems, i.e., the RIS-aided smart channel and RIS-aided mega MIMO, as shown in Fig. 2.

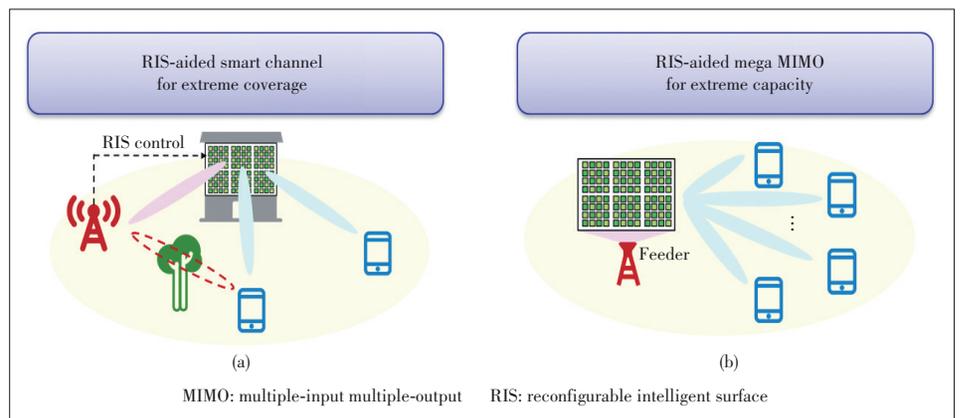
### 2.1 Application 1: RIS-Aided Smart Channel

Extreme coverage is expected for 6G, i.e., a high data rate connection shall be provided everywhere. In practical cellular systems, especially

in an urban scenario, fading and interferences in the radio environment make this difficult to achieve. Conventionally, additional base stations (BSs) shall be deployed and optimized to resolve this. Alternatively, relaying techniques may also be considered for some cases such as range extension in rural areas or an outdoor-to-indoor (O2I) scenario.

For coverage extension, RIS can be deployed in the wireless environment between BSs and user equipment (UE) to achieve a RIS-aided smart channel, as illustrated in Fig. 2(a). In this case, the RIS is responsible for redirecting the incident signals, under the control of the BS, towards the intended user devices.

As compared with the conventional relaying technique, RIS, as a new network topology technique, has the following advantages. Firstly, RIS operates in the full-duplex mode without significant self-interference. In contrast to this, conventional half-duplex relays suffer from a loss in spectral efficiency, while full-duplex relays require either well-isolated transmit and receive antennas or additional self-interference cancellation functionality. Secondly, RIS forwards negligible additive noise to the receiver, which, on the contrary, is a major performance limiting factor for conventional radio-frequency (RF) repeaters. Thirdly, thanks to the simple structure and low cost of RIS, the electrical size of a RIS panel can be much larger than that of a conventional relay, resulting in a very narrow RIS beam. Therefore, a RIS could better help focus the scattered energy on the intended UE while causing little interference to other pieces of UE. Finally, as the RIS technology develops, more dynamic and intelligent RIS can also be expected and exploited for applications such as inter-cell interference suppression and multi-layer enhancement in strong line-of-sight



▲ Figure 2. Typical RIS applications for 5G evolution & 6G

(LoS) scenarios.

### 2.2 Application 2: RIS-Aided Mega MIMO

The peak data rate of 6G is expected to be over 100 Gbit/s. To achieve this, it is necessary to exploit the spectrum resource in a higher frequency band like the sub-THz band. However, current transceiver architecture and the RF technology may not be well suited for such high frequencies and large bandwidth systems. Challenges such as the efficiency of power amplifiers, the resolution and bandwidth of analog-digital/digital-analog converters, the difficulties in realizing high-frequency phase shifters, and the much higher requirement on the die size will all influence the performance and the power- and cost-efficiency of future high frequency and large bandwidth systems.

To meet these challenges, implementing RIS on the BS-side is considered as an economical alternative to the massive MIMO architecture. This application is referred to as RIS-aided mega MIMO, as illustrated in Fig. 2(b). Specifically, modulated signals are fed onto a RIS either through a wired feed network or via space feeding, and the RIS shall beamform signals towards intended user devices. The RIS beam is expected to be highly directional due to the large electrical size of RIS. This helps compensate for the severe path loss of high frequency band radio channels and improves the received signal strength. In contrast to RIS-aided mega MIMO, expending a massive MIMO BS to the same array size is much more expensive. Furthermore, hybrid beamforming can also be adopted for RIS-aided mega MIMO to support multi-layer transmission and to achieve an even higher data rate. To this end, multiple feeders shall be used to feed pre-coded data streams to the RIS, which shall generate multiple narrow beams with low side lobe for interference suppression.

In the above case, the role of RIS can be regarded as a beamformer. Besides, RIS can also be utilized for direct signal modulation. For instance, unmodulated carrier signals can be fed onto the RIS, and the information to be transmitted shall

be modulated through a designated change of RIS state in the time domain<sup>[7]</sup>. This innovative technique largely simplifies the RF hardware in the transmitter and is suitable for low-cost and energy-limited devices. However, further investigation is needed to improve the limited data rate.

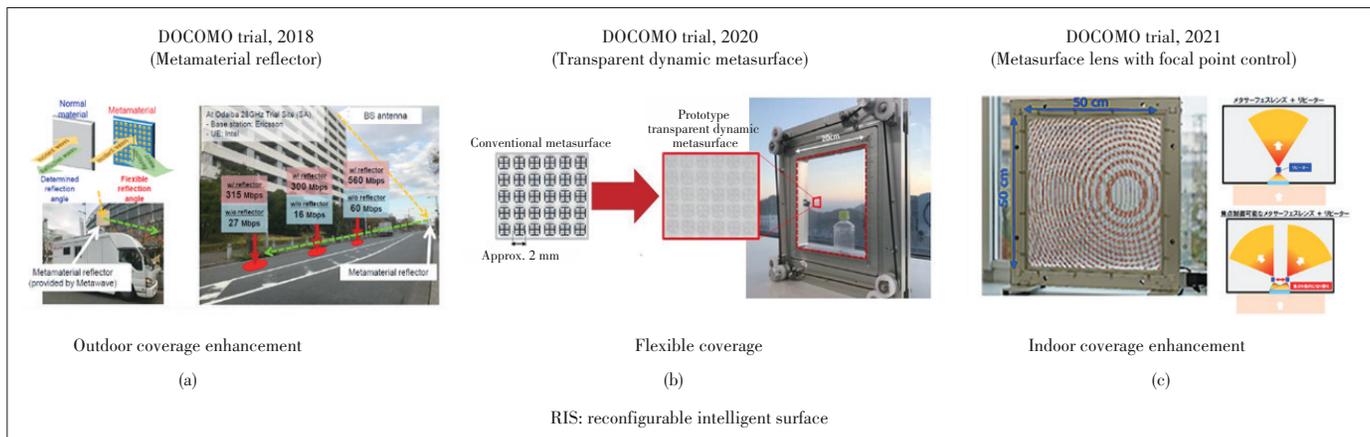
## 3 Observations from Trials and Simulations

In this section, we will introduce several observations from DOCOMO RIS trials and system-level computer simulations, especially the deployment strategy and potential performance gain for coverage enhancement.

### 3.1 RIS Trials

DOCOMO conducted several trials to test metamaterial surfaces from 2018 to 2021, including the first demonstration of a metamaterial reflector at 28 GHz for outdoor coverage expansion in 2018, the first trial of transparent dynamic metasurface capable of dynamically switching between full penetration, partial reflection, and full reflection modes in 2020, and a demonstration of metasurface lens to improve outdoor-indoor coverage with focal point control in 2021.

In 2018, DOCOMO conducted a field trial about a prototype of a metamaterial reflector working on a 28 GHz millimeter-wave (mmWave) band<sup>[8]</sup>. During the trial, the BS was mounted on the roof of a building which made the street under the building a blind spot of coverage, as shown in Fig. 3 (a). To cover this area, a metamaterial reflector was deployed and adjusted to reflect the beam to this area. After the metamaterial reflector was deployed, more than 15 dB signal-to-noise-ratio (SNR) enhancement and more than 500 Mbit/s throughput enhancement were observed, which demonstrated the effectiveness of the metamaterial reflector for coverage enhancement. In 2020, DOCOMO conducted trials with a transparent dynamic metasurface, as shown in Fig. 3(b)<sup>[9]</sup>. The transparent dynamic metasurface consists of two transparent base boards, on one of which a pre-designed pattern of meta-material is printed. The operation mode of the transparent dy-



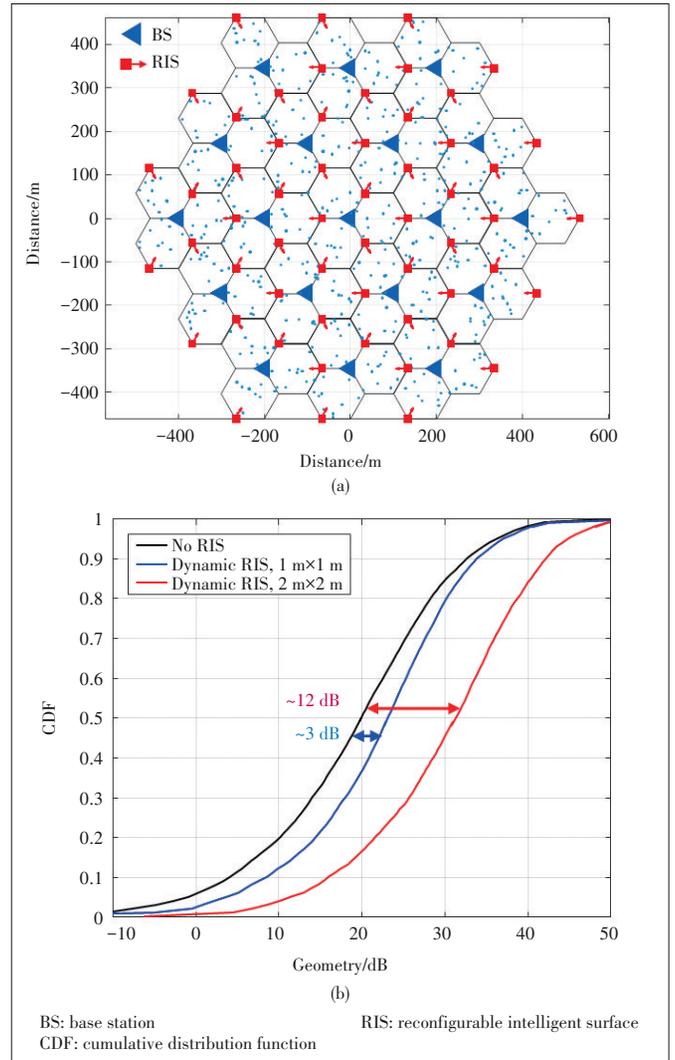
▲ Figure 3. DOCOMO RIS trials<sup>[8 - 10]</sup>

dynamic metasurface can be controlled by adjusting the distance of the two boards. One of three modes, i.e., full penetration, full reflection, and hybrid mode, can be configured. In the full penetration mode, the penetration loss of the metasurface is only about 1 dB, and more than 10 dB penetration loss is observed in the full reflection mode. Since the metasurface is transparent, it can be mounted on windows without impacts on the environment and can be used to control the signal/interference for a dedicated area. An enhanced version of the transparent dynamic metasurface was tested in 2021, which could focus the mmWave to a focal point<sup>[10]</sup>. On the focal point, the metasurface can improve the SNR by about 24 dB compared with normal glasses. The indoor coverage of mmWave can be enhanced if a repeater is deployed on the focal point of the metasurface as shown in Fig. 3(c). The O21 mmWave coverage can be enhanced with this solution, which provides a reliable link between outdoor mmWave BS and indoor UE.

### 3.2 Computer Simulations

Through a series of experiments, the coverage enhancement ability, dynamic control technology, near-field focusing technology, and typical scene application of RIS have been preliminarily verified. In order to test the effectiveness of large-scale deployment of RIS and compare the performance of different RIS functions, e.g., beamforming schemes, it is necessary to further explore the RIS deployment strategy with the help of simulation, such as the number, location, size and function of deploying RIS in cellular systems, so as to meet the performance requirements of real scenarios. To this end, we conducted several computer simulations, and the results will be presented in the following.

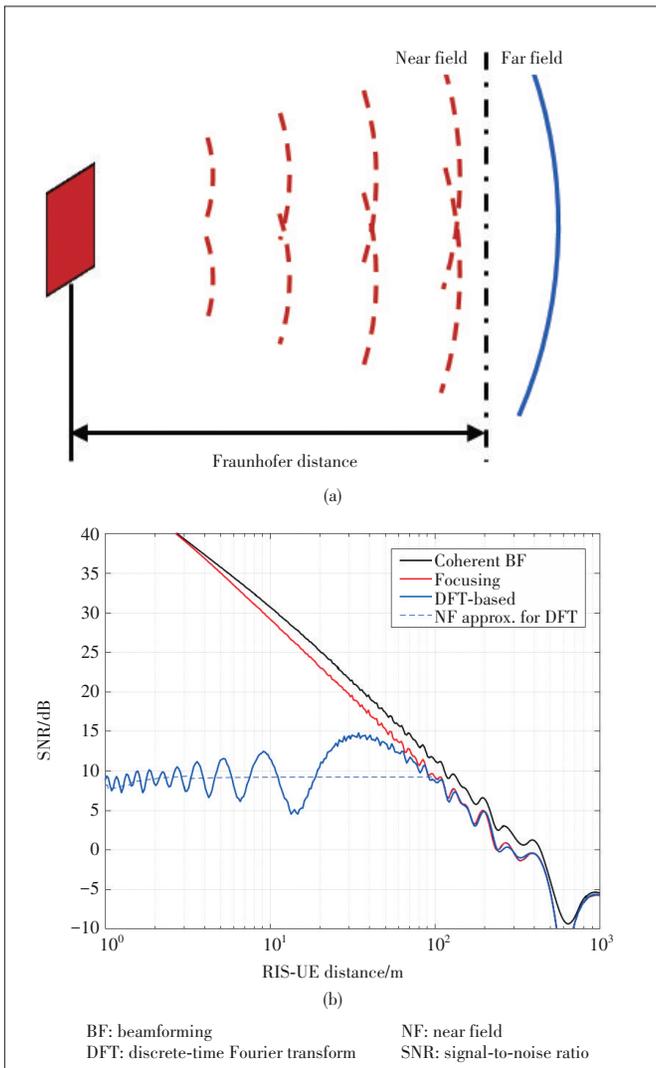
A system-level simulation (SLS) that performed in a multi-cell scenario for 28 GHz will be introduced first. The scenario, as shown in Fig. 4(a), consists of 57 explicitly modeled hexagonal cells following the Dense Urban-eMBB case in Ref. [11]. For simplicity, the user devices are assumed to have isotropic antennas and the fast fading channel component is not modeled. A RIS reflector is deployed in each cell at the cell corner facing the corresponding BS. An ideal case where both the BS-RIS link and the RIS-UE link are LoS is considered. Especially, the height of RIS reflectors is set to 10 m, and the UMi-Street Canyon path loss model is adopted for RIS-UE channels. Two different RIS sizes are considered in this simulation, i.e., a  $1 \times 1 \text{ m}^2$  and a  $2 \times 2 \text{ m}^2$  RIS. The RIS elements are assumed to have half-wavelength spacing. The RIS beamforming gain is modeled as the combined gain of a receiving beam towards the attached BS and a transmitting beam towards the serving UE. Furthermore, taking the large RIS size and the extended near field into account, the transmitting beam of RIS is assumed to focus on the UE to attain the maximum gain. The performance in terms of geometry is shown in Fig. 4(b). As compared with the baseline case without RIS, the observed performance gain of RIS is ca. 3 dB and 12 dB for  $1 \times 1 \text{ m}^2$  and



▲ Figure 4. System-level simulation (SLS) for RIS: (a) the scenario and (b) the geometry

$2 \times 2 \text{ m}^2$  RIS, respectively. Moreover, for the  $1 \times 1 \text{ m}^2$  RIS case, ca. 56% UE chooses to be served by the RIS link, while this number increases to 94% for the  $2 \times 2 \text{ m}^2$  RIS case. These results clearly demonstrate the potential gain of deploying RIS in mmWave systems. However, it shall also be noticed that, even under the idealized assumptions, the size of the required RIS for attaining a notable gain is remarkably large.

For a large-scale RIS, its radiative near field can be as far as hundreds of meters, as illustrated in Fig. 5(a). For instance, a  $1 \times 1 \text{ m}^2$  RIS operating at 30 GHz has a near field up to 400 m according to its Fraunhofer distance. In other words, in a typical cellular scenario, almost all user devices to be served by the RIS are located in its near field. However, conventional beamforming is designed for the far field of the antenna array and suffers from a noticeable performance loss in the near field due to diffraction. Given 400 MHz bandwidth at a 30 GHz band and  $1 \times 1 \text{ m}^2$  RIS deployed 150 m away from the BS, this effect is shown in Fig. 5(b), where the receiving SNR at UE is plot-



▲ Figure 5. Near-field effect

ted as a function of the distance between the RIS and the UE. A two-ray model with plane earth reflection is assumed for the RIS-UE channel. In Fig. 5(b), the black, red, and blue solid curves represent the performance of coherent, focused, and collimated (DFT-based) beamforming, respectively. The blue dotted curve represents the approximated average performance of collimated beamforming in the near field of RIS. It can be noticed that in this simple scenario, focused beamforming achieves near-optimal performance. However, the performance of collimated beamforming oscillates around a constant value for UE within 100 m. This damped oscillating behavior follows the well-known Fresnel diffraction. The above results reveal both potential

possibilities and challenges for RIS beamforming design. On the one hand, large-scale RIS can provide sufficient gain to combat the path loss of the two hops via beamforming. On the other hand, a new distance-dependent beamforming scheme is necessary to exploit the full potential of RIS, especially in the near field.

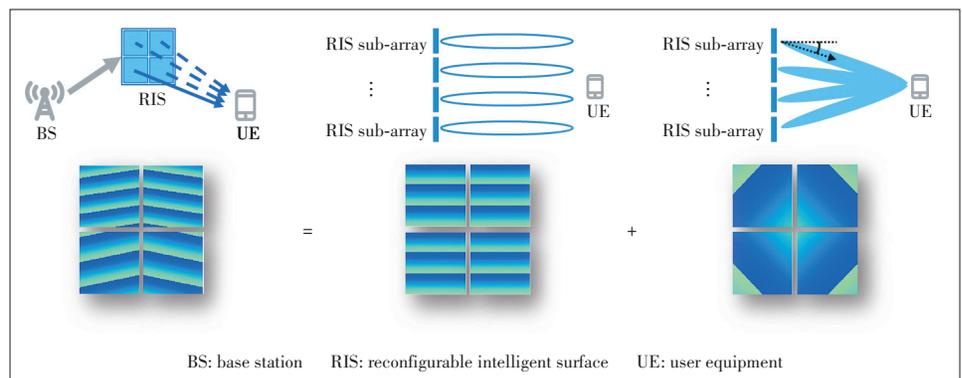
To accommodate the near-field effect, we are designing a RIS beamforming scheme with a tradeoff between complexity and performance. The principle of the RIS beamforming scheme is to divide a large RIS into multiple smaller sub-arrays, whose size shall be dynamically chosen according to the RIS-UE distance. The scheme includes two beamforming steps, as shown in Fig. 6. In the first step, let a collimated/DFT-based beam be selected for all RIS sub-arrays based on the UE direction. Then, in the second step, each sub-beam of the RIS sub-arrays is steered by an individual angle so that the sub-beams are focused on the UE. Preliminary simulation shows that this two-step dynamic RIS beamforming method can improve the received SNR at the near field UE by about 10 - 20 dB compared with the traditional collimated/DFT-based beamforming method.

#### 4 Challenges and Opportunities of RIS

Though the performance of RIS has been demonstrated by simulations and verified by prototyping and trials, a lot of challenges still need to be addressed before this technology can be implemented in practical systems. These challenges can be, roughly speaking, viewed from four aspects, i. e., hardware, modeling, algorithm, and standardization.

##### 4.1 Hardware

For the time being, different RIS prototypes vary in the material, element design and arrangement, tuning method, and operating frequency. As for the performance, both the gain to the in-band signal and the attenuation to the out-band signal are of particular importance. For the in-band signal, a uniform gain over large bandwidth is more favorable in terms of improving the system capacity. For the out-band signal, deeper attenuation and sharper edge shall better help interference suppression in practical systems. Besides, the accuracy and



▲ Figure 6. Two-step RIS beamforming method

speed of tuning the RIS element also largely influence the performance and application. In terms of the cost, both manufacturing and maintaining costs shall be considered, which depend on the material, processing, package, durability, power consumption, etc. Finally, an outlook design that fits the application and the environment is always preferred.

#### 4.2 Modeling

To study the performance of RIS in a more realistic scenario, sufficiently accurate modeling with acceptable complexity for both the RIS element and the RIS channel is needed. For the RIS element, both the spatial and the spectral characteristics shall be modeled. Furthermore, depending on the hardware implementation, imperfections such as amplitude-phase coupling, mutual coupling, and angle-dependent response need to be taken into consideration. For the modeling of RIS channel, intensive work has been conducted. However, a unified and reliable RIS channel model is still missing. Issues such as the modeling methodology, the near field channel model, and modeling for specific scenarios are still open.

#### 4.3 Algorithm

Concerning the schemes and algorithms for RIS-aided transmission, the following topics are closely related to the application of this technology. Firstly, due to the passiveness and the large size of typical RIS, the channel estimation is rather challenging. Therefore, a practical RIS channel estimation mechanism with acceptable signaling overhead is needed. Secondly, adaptive RIS beamforming is crucial for achieving the promised gain. Besides the performance, other aspects such as the complexity, the CSI requirement, and the robustness shall be considered. Thirdly, the cooperation between multiple BSs and RIS, as well as the scheduling of both RIS and UE, is an interesting topic related to the future RIS-aided wireless systems.

#### 4.4 Standardization

3GPP officially identified 5G-Advanced as the name of 5G evolution standard, marking the opening of the post 5G era. Some preliminary RIS use cases which are transparent to UE can be considered now for the 5G-Advanced standardization stage. Some companies have already proposed to study RIS during the discussion about 3GPP Release 18 study and work items. There are also proposals on studying smart repeater for Release 18 during the discussion, which has tight connections with the RIS use cases. However, many challenges are still to be addressed to specify RIS for current NR systems. Current RIS prototypes have different physical dimensions, scales, and materials of different RF characteristics. It is a great challenge to model and accommodate such variations in the specifications. We may need a set of RF requirements and a flexible baseband specification to support all these kinds of RIS devices. A unified model and evaluation methodology should be discussed for RIS. Although RIS can be viewed as a kind of smart repeaters and could inherit many features that would be

specified for smart repeaters. There will still be differences which should be addressed in the future after the smart repeater is specified. For example, the beam control of RIS is more complex than the smart repeater. There are a lot of differences in terms of beam number, beamwidth, timing, and power control between RIS and smart repeater, which have not been included in the current working scope of smart repeaters. In our view, a comprehensive study should be done before specifying RIS for 5G-Advanced. A set of RF models should be investigated and formulated in order to clearly describe the characteristic of RIS, based on which we can further study its potential performance gain and specification impacts.

## 5 Conclusions

In this paper, we mainly discuss some observations and thoughts about RIS applications for 5G evolution and 6G. As an emerging technology, RIS has promising advantages of high integration, low cost and low power consumption, but its commercial maturity still needs time. At present, we should seize the time window of 5G evolution and 6G to realize the preliminary commercial application of RIS. In terms of 5G evolution and 6G time scales, the most attractive RIS applications include blind-spots coverage, indoor coverage enhancement, MIMO transmitter enhancement, etc. Meanwhile, RIS is expected to become the infrastructural platform for integrated communication, sensing, and AI, and to realize the intelligent channel of independent perception of the environment and real-time parameter optimization. Therefore, we suggest that the RIS application can take a step-by-step strategy and gradually expand its application scope. The successful application of RIS in 5G evolution, 6G, and future mobile systems calls for tight collaboration between academia and industry.

## References

- [1] NTT DOCOMO. DOCOMO 6G White Paper-5G Evolution and 6G [R]. Tokyo, Japan: NTT DOCOMO, 2021
- [2] DI RENZO M, ZAPPONE A, DEBBAH M, et al. Smart radio environments empowered by reconfigurable intelligent surfaces: how it works, state of research, and the road ahead [J]. *IEEE journal on selected areas in communications*, 2020, 38(11): 2450 - 2525. DOI: 10.1109/JSAC.2020.3007211
- [3] GONG S M, LU X, HOANG D T, et al. Toward smart wireless communications via intelligent reflecting surfaces: a contemporary survey [J]. *IEEE communications surveys & tutorials*, 2020, 22(4): 2283 - 2314. DOI: 10.1109/COMST.2020.3004197
- [4] YUAN X J, ZHANG Y J A, SHI Y M, et al. Reconfigurable-intelligent-surface empowered wireless communications: challenges and opportunities [J]. *IEEE wireless communications*, 2021, 28(2): 136 - 143. DOI: 10.1109/MWC.001.2000256
- [5] WU Q Q, ZHANG S W, ZHENG B X, et al. Intelligent reflecting surface-aided wireless communications: a tutorial [J]. *IEEE transactions on communications*, 2021, 69(5): 3313 - 3351. DOI: 10.1109/TCOMM.2021.3051897
- [6] LIU Y W, LIU X, MU X D, et al. Reconfigurable intelligent surfaces: principles and opportunities [J]. *IEEE communications surveys & tutorials*, 2021, 23(3): 1546 - 1577. DOI: 10.1109/COMST.2021.3077737
- [7] TANG W K, DAI J Y, CHEN M Z, et al. MIMO transmission through reconfigurable intelligent surface: system design, analysis, and implementation [J]. *IEEE journal on selected areas in communications*, 2020, 38(11): 2683 - 2699. DOI:

- 10.1109/JSAC.2020.3007055
- [8] KITAYAMA D, KURITA D, MIYACHI K, et al. 5G radio access experiments on coverage expansion using metasurface reflector at 28 GHz [C]/2019 IEEE Asia-Pacific Microwave Conference. IEEE, 2019: 435 - 437. DOI: 10.1109/APMC46564.2019.9038267
- [9] NTT DOCOMO. DOCOMO Conducts World's First Successful Trial of Transparent Dynamic Metasurface [EB/OL]. [2021-12-14]. [https://www.nttdocomo.co.jp/english/info/media\\_center/pr/2020/0117\\_00.html](https://www.nttdocomo.co.jp/english/info/media_center/pr/2020/0117_00.html)
- [10] NTT DOCOMO. DOCOMO and AGC use metasurface lens to enhance radio signal reception indoors [EB/OL]. [2021-12-14]. [https://www.nttdocomo.co.jp/english/info/media\\_center/pr/2020/0117\\_00.html](https://www.nttdocomo.co.jp/english/info/media_center/pr/2020/0117_00.html)
- [11] ITU-R. Guidelines for evaluation of radio interface technologies for IMT-2020: M.2412 [S]. 2017

### Biographies

**HOU Xiaolin** (hou@docomolabs-beijing.com.cn) received the B.S. degree in telecommunication engineering and the Ph.D. degree in communication and information system from Beijing University of Posts and Telecommunications, China in 2000 and 2005, respectively. Now he is the director of wireless technology department of DOCOMO Beijing Laboratories. He has a strong academic and industrial background and has been actively contributing to 4G and 5G research, standardization, and trials. He is now focusing on the research and standardization towards 5G evolution and 6G, including massive MIMO, mmWave, RIS, flexible duplex, NOMA, waveform, and the applications of artificial intelligence in physical layer design.

**LI Xiang** received the B.S. degree in communication engineering from Wuhan University, China, the M.S. degree in electrical engineering, and the Ph.D. degree in communications engineering from the University of Rostock, Germany in 2005, 2012, and 2018, respectively. In 2019, he joined DOCOMO Beijing Laboratories as an advanced researcher in wireless technology department. His research interests include multi-antenna systems, interference management, wireless relaying, and cooperative communications. He is currently focusing on the physical layer technologies towards 6G, including Mega MIMO, RIS, and

Terahertz communications.

**WANG Xin** received the B.S. degree in electronics engineering and the Ph.D. degree in signal and information processing from Beihang University, China in 2001 and 2008, respectively. He is now a senior researcher in wireless technology department of DOCOMO Beijing Laboratories. His current research interests include Mega-MIMO (ultra-massive MIMO) technologies on high frequency bands, reconfigurable intelligent surfaces, deep learning for physical layer design, and Terahertz communications.

**CHEN Lan** received the M.E. and Ph.D. degrees from the Department of Communications and Computer Engineering of Kyoto University, Japan. She joined the NTT DOCOMO in 1999. Since then, she has been involved in research on LTE and LTE-Advanced and 3GPP RAN standardization. From 2013 to 2016, she was involved in the development for the enhancement of commercial radio NW. Since July 2016, she has been leading DOCOMO Beijing Laboratories on the research and standardization on 5G, evolution and beyond, and deep-learning based AI innovation. She received the Best Paper Awards in ICT2002 and ChinaCom2012, and Encouragement Award and Accomplishment Award in field of ICT from ITU Association of Japan in 2014 and 2018, respectively.

**SUYAMA Satoshi** received the B.S. degree in electrical and electronic engineering, the M.S. degree in information processing, and the Dr. Eng. degree in communications and integrated systems, all from Tokyo Institute of Technology, Japan, in 1999, 2001, and 2010, respectively. From 2001 to 2013, he was an assistant professor in the Department of Communications and Integrated Systems, the Tokyo Institute of Technology, Japan. He has been engaged in research on OFDM mobile communications systems and applications of the adaptive signal processing, including turbo equalization, interference cancellation and channel estimation. Since April 2013, he has joined NTT DOCOMO, Inc. and has been involved in research and development of 5G mobile communications system. He is currently the manager of 6G Laboratories in NTT DOCOMO. He received the Best Paper Prize from the European Wireless Technology Conference (EuWiT) in 2009, Best Paper Award from International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) in 2016, and the Paper Award from IEICE in 2012.

## New Member of ZTE Communications Editorial Board >>>



**WANG Chengxiang** received the B.S. and M.E. degrees in communication and information systems from Shandong University, China in 1997 and 2000, respectively, and the Ph.D. degree in wireless communications from Aalborg University, Denmark in 2004.

He was a research assistant with Hamburg University of Technology, Germany, from 2000 to 2001, a visiting researcher with Siemens AG Mobile Phones, Germany, in 2004, and a research fellow with the University of Agder, Norway from 2001 to 2005. He has been with Heriot-Watt University, U.K. since 2005, where he was promoted to a professor in 2011. From 2018, he has been with Southeast University, China as a professor. He is also a part-time professor with Purple Mountain Laboratories, China. He has authored four books, three book chapters, and more than 450 papers in refereed journals and conference proceedings, including 25 highly cited papers. He has also delivered 23 invited keynote speeches/talks and nine tutorials in international conferences. His current research interests include wireless channel measurements and modeling, 6G wireless communication networks, and electromagnetic information theory.

Dr. WANG is a member of the Academia Europaea (The Academy of Europe), a Fellow of the Royal Society of Edinburgh (RSE), IEEE, IET, and China Institute of Communications (CIC), an IEEE Communications Society Distinguished Lecturer in 2019 and 2020, and a highly-cited researcher recognized by Clarivate Analytics in 2017–2020. He is currently an executive editorial committee member of *IEEE Transactions on Wireless Communications*, and has served as an editor/guest editor for about 20 international journals. He has served as a General Chair, TPC Chair, or Symposium Chair for more than 30 international conferences. He received 14 best paper awards from international conferences, and received the 2020 & 2021 “AI 2000 Most Influential Scholar Award Honorable Mention”.



# Recent Developments of Transmissive Reconfigurable Intelligent Surfaces: A Review

TANG Junwen, XU Shenheng, YANG Fan, LI Maokun

(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

DOI: 10.12142/ZTECOM.202201004

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220305.1951.004.html>, published online March 8, 2022

Manuscript received: 2022-01-24

**Abstract:** Reconfigurable intelligent surface (RIS) is considered as one of the key technologies for the next-generation mobile communication systems. The transmissive RIS is able to achieve dynamic beamforming capability while transmitting an in-band RF signal through its aperture, and has promising prospects in various practical application scenarios. This paper reviews some of the latest developments of the transmissive RIS. The approaches for transmissive RIS designs are classified and described briefly. Numerous designs with different phase resolutions, such as 1-bit, 2-bit or continuous  $360^\circ$  phase shifts, are presented, with detailed discussions on their operating mechanisms and transmission performances. The design solutions for various transmissive RIS elements are summarized and compared.

**Keywords:** antenna; mobile communications; phase shifter; RIS; transmitarray

**Citation** (IEEE Format): J. W. Tang, S. H. Xu, F. Yang, et al, "Recent developments of transmissive reconfigurable intelligent surfaces: a review," *ZTE Communications*, vol. 20, no. 1, pp. 21 – 27, Mar. 2022. doi: 10.12142/ZTECOM.202201004.

## 1 Introduction

As one of the key technologies for the next-generation mobile communication systems, the reconfigurable intelligent surface (RIS)<sup>[1-3]</sup> has attracted great attentions recently. On the contrary to the commonly used reflection mode, the transmissive RIS is able to achieve dynamic beamforming capability while transmitting an in-band RF signal through its aperture. This distinctive feature allows much improved direct indoor coverage by the outdoor cellular signals. It can also be employed as a reconfigurable radome for 5G base station antennas or terminal antennas. The large-scale RIS array can effectively boost the coverage of the medium-sized 5G massive multiple-input multiple-output (MIMO) antennas, and the analog-digital hybrid structure can keep the overall cost and power consumption low.

Solid-state electronic devices are commonly integrated in each constituent RIS element to achieve phase reconfigurability. Continuous phase shifts can be realized using analog-type devices like varactor diodes<sup>[4-13]</sup>. The phase shift range usually exceeds  $360^\circ$  so that the phase errors are negligible. However, the transmission insertion loss of the RIS element is relatively high and for instance, can be up to 5.7 dB<sup>[7]</sup>, considerably reducing the aperture efficiency of the RIS. Switch-type devices, such as p-i-n diodes and radio frequency (RF) microelectromechanical system (MEMS) switches, are widely used to pro-

duce discrete phase shifts<sup>[14-31]</sup>, especially at millimeter wave frequencies. Most designs focus on 1-bit phase reconfigurability because the complexity in design and fabrication is more manageable<sup>[14-27]</sup>. The phase quantization errors associated with discrete phase shifts inevitably introduce performance degradation<sup>[32-34]</sup>. It is observed that the 1-bit designs suffer from 3 – 4 dB loss attributed to the coarse phase resolution, resulting in low aperture efficiencies. The sidelobe levels are usually much higher, sometimes even causing unwanted grating lobes. The 2-bit design is considered a well-balanced choice between the design complexity and the element performance. The phase quantization loss can be greatly reduced to less than 1 dB, and the sidelobe envelop is much improved as well<sup>[28-31]</sup>.

Most transmissive RIS designs are based on the receiver-transmitter (Rx-Tx) structure<sup>[4-9, 14-22, 25-31]</sup>, and microstrip patch antennas<sup>[4-9, 14-18, 20-21, 25-26, 30]</sup>, slot antennas<sup>[19, 28-29]</sup>, vivaldi antennas<sup>[22]</sup> and dipole antennas<sup>[27, 31]</sup> are employed as receivers or transmitters. There are a handful of designs using the stacked frequency selective surface (FSS) structures<sup>[10-13, 23-24]</sup>. Each FSS layer provides a finite-degree phase shift, and a wide phase shift range is achieved by stacking multiple FSS layers. The structure is relatively more complicated and more solid-state electronic devices are needed in each unit cell.

This paper provides a review of some recent transmissive

RIS studies and is organized as follows. The configuration classification of the transmissive RIS is introduced in Section 2. Various transmissive RIS designs with different phase resolutions or polarizations are described in Section 3. Finally, conclusions are given in Section 4.

## 2 Design Approaches

There are two widely used design approaches for the transmissive RIS, which are the receiver-transmitter (Rx-Tx) structure and the stacked FSS structure. Generally, a typical Rx-Tx structure consists of three components: the receiving (Rx) cell, the transmitting (Tx) cell, and the power transmission structure. A ground plane separates the Rx from the Tx, so that they can operate independently, avoiding the mutual interference with each other. The power transmission structure between them is usually realized by a metallic via<sup>[7-9, 14-18, 20-21, 25-27, 30]</sup> or a slot-coupling structure<sup>[4-6]</sup>. When the RIS is illuminated by an incident wave, the Rx on one side of the RIS first receives the wave and converts it into a guided wave. This wave is then conveyed through the power transmission structure to the Tx deployed on the other side of the RIS, and re-radiated to the free space. The phase shift can be realized on either the Rx, the Tx, or the power transmission structure.

The stacked FSS structure type is inspired by the classic FSS designs. It consists of multiple FSS layers of metallic and dielectric substrates. A phase shift with finite degrees is obtained when a wave propagates through each FSS layer, and by cascading multiple layers, a larger phase shift can usually be realized. The separation between layers is also a crucial parameter. The overall thickness is relatively large. Unlike the Rx-Tx type, there is no ground plane in the stacked FSS structure type (Fig. 1).

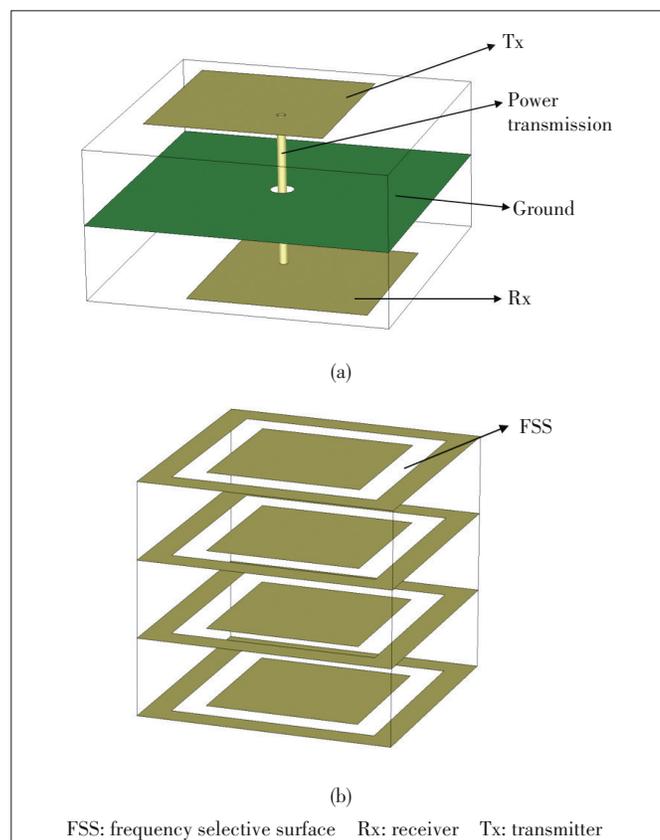
## 3 Representative Transmissive RIS Designs

### 3.1 Single-Polarized 1-Bit Designs

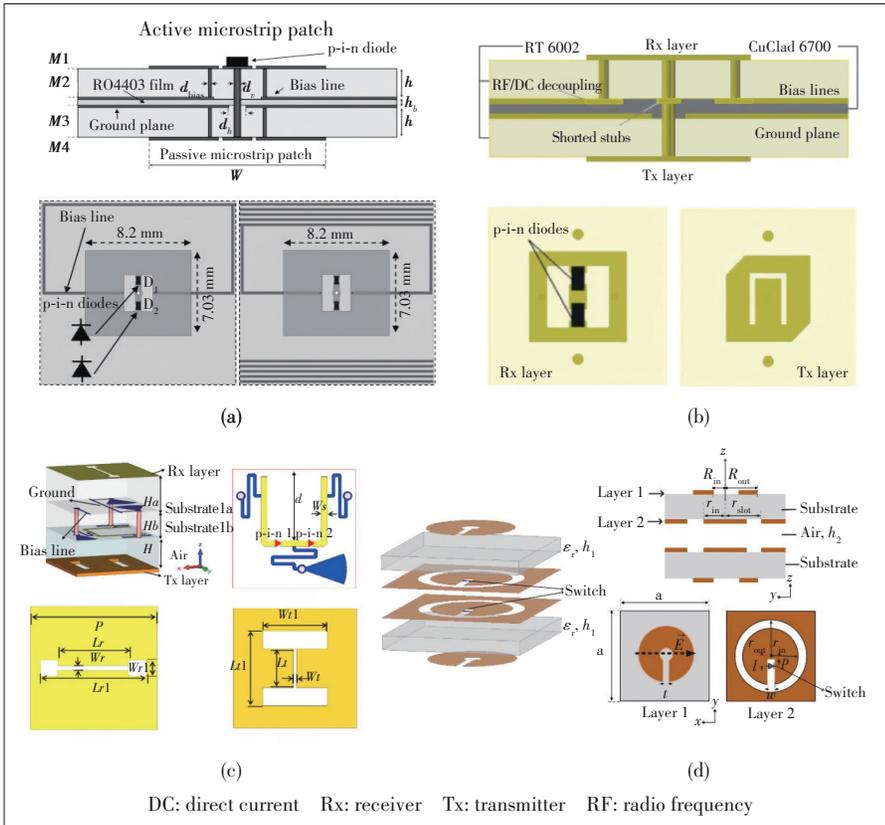
A 1-bit transmissive RIS element can realize two phase states with a phase difference of  $180^\circ$ . Generally, there are two operating mechanisms in the element designs, which are the current reversal method and the variable resonance method. The current reversal method based on the Rx-Tx structure is most popular for 1-bit designs in literature. The p-i-n diode is usually chosen as the controlling device due to its small volume and stable performance especially at high frequencies. The p-i-n diode can work at “ON” or “OFF” states by supplying a different DC biasing signal. In practice, two p-i-n diodes are usually embedded in parallel and in opposite directions in each element, and a single positive or negative biasing signal can switch them between the “ON-OFF” and “OFF-ON” states. Hence, the excitation current on the radiating element flows in opposite directions, and the radiated field is reversed, resulting in a stable phase shift of

$180^\circ$ . The stacked FSS type structure is usually based on the variable resonance method. The resonant characteristic of the element structure is manipulated by switching the integrated p-i-n diode “ON” or “OFF”, and consequently, a phase difference in its radiated field is caused. It is noted that the phase difference varies with frequency, and its bandwidth is usually narrower than that using the current reversal method.

Single polarized 1-bit RIS has been widely investigated. A representative linearly polarized design with an O-slot patch and two p-i-n diodes based on the Rx-Tx structure and the current reversal mechanism is proposed in Ref. [14], as shown in Fig. 2(a). The element evolves from the U-slot microstrip patch antenna. It is modified into an O-slot so that two p-i-n diodes placed across it and biased in opposite states can effectively form a U-slot patch with reversible direction, thereby achieving a phase difference of  $180^\circ$  due the current reversal mechanism. A metallic via connecting the Rx and the Tx is used to transmit the signal from the incident side to the re-radiating side. The measured transmission insertion loss is 1.7/1.9 dB for two states and the 3-dB transmission bandwidth is 14.7%. Moreover, the DC biasing circuit is elaborately designed. The positive end of one p-i-n diode and the negative end of the other are connected together through two symmetrical metallic vias. Only a single DC biasing signal is required



▲ Figure 1. Design approaches: (a) Rx-Tx structure; (b) stacked FSS structure



▲ Figure 2. Schematics of single-polarized 1-bit transmissive reconfigurable intelligent surface (RIS) elements: (a) linear polarization in Ref. [14]; (b) linear-circular polarization conversion in Ref. [16]; (c) slot coupling structure in Ref. [19]; (d) variable resonance design in Ref. [23]

to control them working at opposite states. The biasing lines are arranged close to the ground, thus minimizing its interference on the RF performance. It is also worth mentioning that by truncating the corners of the U-slot patch or modifying the U-slot patch, the 1-bit linearly polarized element can be readily developed into a single linear-circular polarization conversion<sup>[16, 18]</sup> or polarization switching designs<sup>[17]</sup>. The sketch of the element in Ref. [16] is shown in Fig. 2(b). A minimum transmission insertion loss of 1.59/1.70 dB at Ka band with a 3-dB bandwidth of more than 12% is validated in measurement.

The 1-bit element presented in Ref. [19] exploits two slot antennas as Rx and Tx. As Fig. 2(c) shows, two H-shaped slots are placed orthogonally to receive and re-radiate electromagnetic waves, with high polarization isolation due to the polarization orthogonality. Two p-i-n diodes are integrated in the U-shaped transmission line to control the excitation current direction. By switching them working at “ON-OFF” or “OFF-ON” state, the electric field coupled to the Tx H-slot is reversed, resulting in a 180° phase shift. The operating frequency of the element is 12.5 GHz, and the simulated transmission insertion loss is 0.86/0.76 dB.

A 1-bit element design based on the variable resonance method is proposed in Ref. [23], and the element configura-

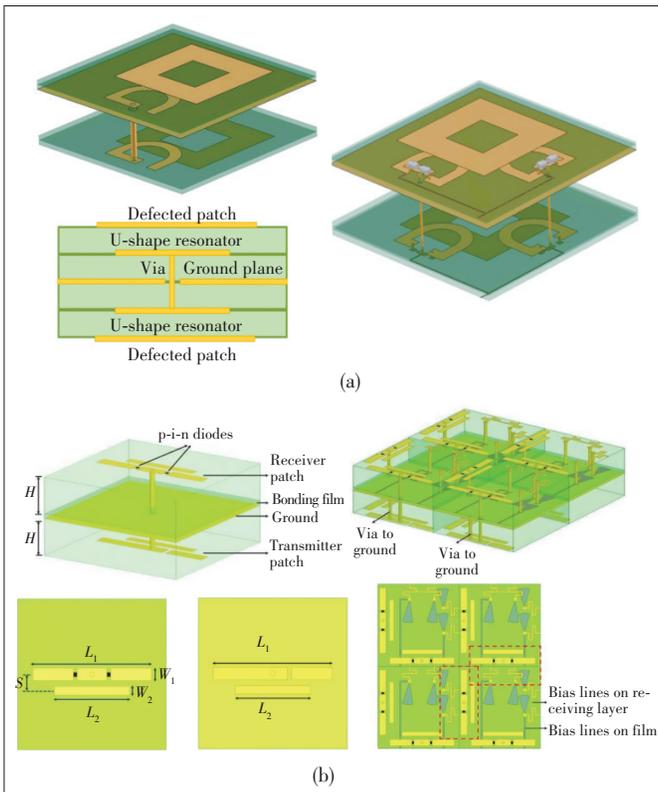
tion is shown in Fig. 2(d). It is composed of four metallic layers and two layers of dielectric substrates. The upper and lower C-shaped patches are utilized for the transmission of the electromagnetic waves, with two ring slots loaded by a rectangular gap used as the phase shifter. Two p-i-n diodes are positioned across the two rectangular gaps and work at “ON-ON” or “OFF-OFF” state. They change the widths of the gaps, and consequently, the equivalent resonance length varies, thus shifting the transmission phase of the element. The measurement shows that the element transmission insertion losses for the two phase states at 11.5 GHz are 0.6 dB and 2.9 dB, respectively. Note that the transmission insertion loss increases due to the detuning of the resonant characteristic of the element structure when the 180° phase shift is realized.

### 3.2 Dual-Polarized 1-Bit Designs

The dual-polarized transmissive RIS designs mostly focus on 1-bit phase resolution. In general, there are two solutions. One is to excite two orthogonal radiating modes of the same radiator<sup>[25]</sup>, and the other is to arrange two sets of single-polarized 1-bit RIS elements orthogonally to achieve the dual-polarized function<sup>[24, 26–27]</sup>. Note that there are some designs with the polarization manipulation capability<sup>[7, 17–18]</sup>; however, they are unable to operate in two polarization directions simultaneously and hence excluded in this section.

The dual-linear polarized 1-bit element proposed in Ref. [25] exploits a five-layer-stacked configuration. The element configuration is illustrated in Fig. 3(a). For the realization of the single-linear polarized 1-bit phase controlling, it relies on the U-shaped feeding line integrated with two p-i-n diodes. By alternatively switching the working states, the radiated fields in opposite direction are excited on the top metallic square ring, and a 180° phase difference can be generated. A dual-linear polarized design can be achieved by employing two orthogonally placed U-shaped feeding lines to excite the same metallic square ring. The measured results show that this element can achieve a transmission insertion loss of 1.7/1.9 dB at 5.9 GHz, with an operating bandwidth of 2.8%.

Ref. [27] presents a dual-linear polarized 1-bit RIS element based on a via-fed dipole with a parasitic dipole. Fig. 3(b) describes the fundamental single-polarized 1-bit RIS element. The element employs the Rx-Tx structure, both based on di-



▲ Figure 3. Schematics of the dual-linear polarized 1-bit transmissive reconfigurable intelligent surface (RIS) element designs: (a) two orthogonal radiating modes of the same radiator in Ref. [25]; (b) two sets of viadipoles with parasitic dipoles in Ref. [27]

pole configurations. Two p-i-n diodes are placed on the arms of the receiver dipole, and a parasitic dipole is added near the dipole to reduce the loss. The mechanism can be explained as the current is mostly coupled to the parasitic dipole, reducing the loss that is caused by the current flowing through the p-i-n diodes. Therefore, the simulated element transmission insertion loss is only 0.8 dB at 12.25 GHz. The design of the dual-linear polarized element can be accomplished by arranging two sets of single-linear polarized elements orthogonally.

### 3.3 2-Bit Designs

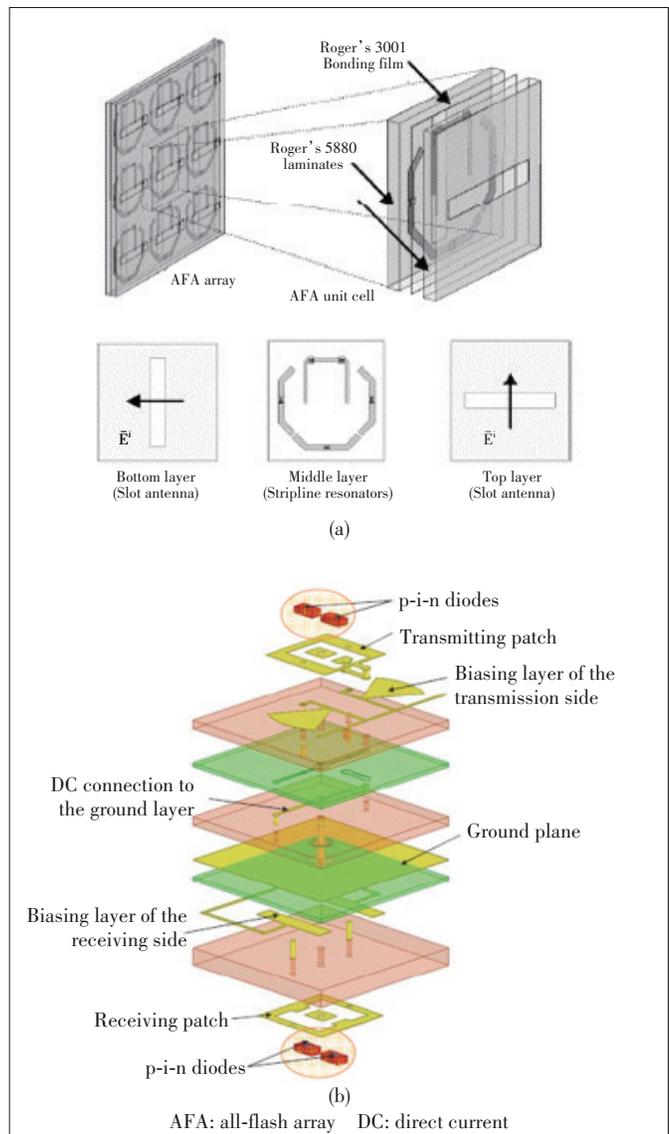
Although considerable developments of the 1-bit transmissive RIS designs have been accomplished, they still suffer from suboptimal radiation behavior caused by the large phase quantization error. In order to increase the phase accuracy so as to improve the radiation performance, the 2-bit design is investigated.

Another 2-bit RIS element at Ka band is proposed in Refs. [28–29]. As illustrated in Fig. 4(a), the Rx and Tx slots are fed with striplines, and five RF MEMS switches are loaded on the striplines to switch the four operating modes alternatively. The measured element transmission insertion loss, however, reaches up to 4.2–9.2 dB due to the challenging fabrication process. A Ka-band linearly polarized 2-bit RIS element is

presented in Ref. [30]. The configuration of the element is depicted in Fig. 4(b). It is composed of six metallic layers and three dielectric substrates, with four integrated p-i-n diodes. The 180° phase shift is realized by alternating the two p-i-n diodes loaded on the receiving O-slot patch. Two more p-i-n diodes are loaded on the transmitting patch, which contains a delay line to achieve the additional 0°/90° phase switching. Thus, the 2-bit phase tuning capability can be obtained. The measured element insertion loss is 1.5 – 2.3 dB for the four phase states at 29 GHz, with the measured 3-dB transmission bandwidth of 10.1% – 12.1%.

### 3.4 Designs with Continuous Phase Shift

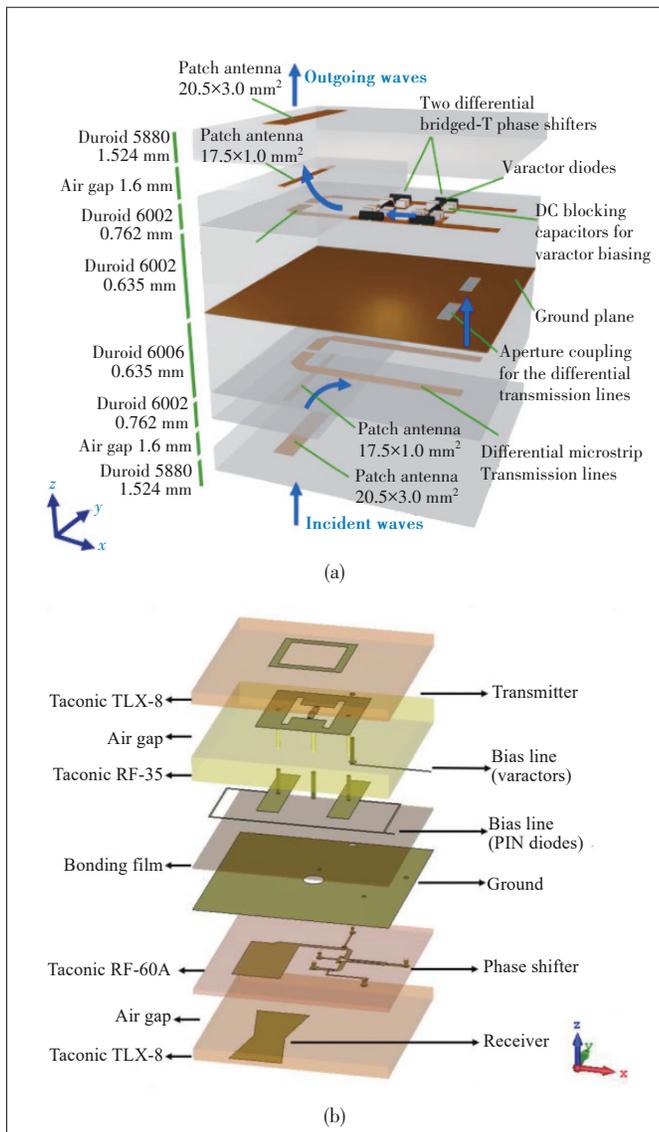
To achieve a continuous 360° phase shift, varactor diodes



▲ Figure 4. Schematics of the 2-bit transmissive reconfigurable intelligent surface (RIS) element designs: (a) dual polarization with five radio frequency (RF) microelectromechanical system (MEMS) switches in Refs. [28–29]; (b) O-slot patches loaded with four p-i-n diodes in Ref. [30]

are widely utilized in most transmissive RIS element designs since the capacitance of the varactor diodes can be changed by applying a DC biasing voltage. One phase tuning approach is based on the microstrip phase shifter<sup>[4-5, 7-9]</sup>. By employing the Rx-Tx structure, the varactor diodes are placed on the phase shifter to control the transmission phase, ensuring the stability of the transmission insertion loss in the operating frequency band when the DC biasing voltage changes. Thus, most wideband RIS designs explore this approach to tune phase shift. The other approach is to change the resonant characteristic of the element structure<sup>[6, 10-13]</sup>, and it is widely used in the stacked FSS structures.

An element with continuous phase controlling can be found in Ref. [5], and the element structure is exhibited in Fig. 5(a).



▲ Figure 5. Transmissive reconfigurable intelligent surface (RIS) element designs with  $360^\circ$  phase controlling: (a) two cascaded bridged-T phase shifters in Ref. [5]; (b) a  $180^\circ$  analog phase shifter combined with 1-bit phase shifter in Ref. [9]

The microstrip-slot coupling structure is adopted in this design to transfer energy between Rx and Tx, which are both stacked dipole patches. Two cascaded bridged-T phase shifters loaded with six varactor diodes are selected to reconstruct the transmission phase. It is worth mentioning that the stacked dipole patches are fed by the differential microstrip transmission lines, and two rectangular gaps are opened in the ground to transfer the differential signals. The simulated results demonstrate that the element is capable of a phase tuning range of over  $400^\circ$ , with the transmission loss maintaining below 3.6 dB.

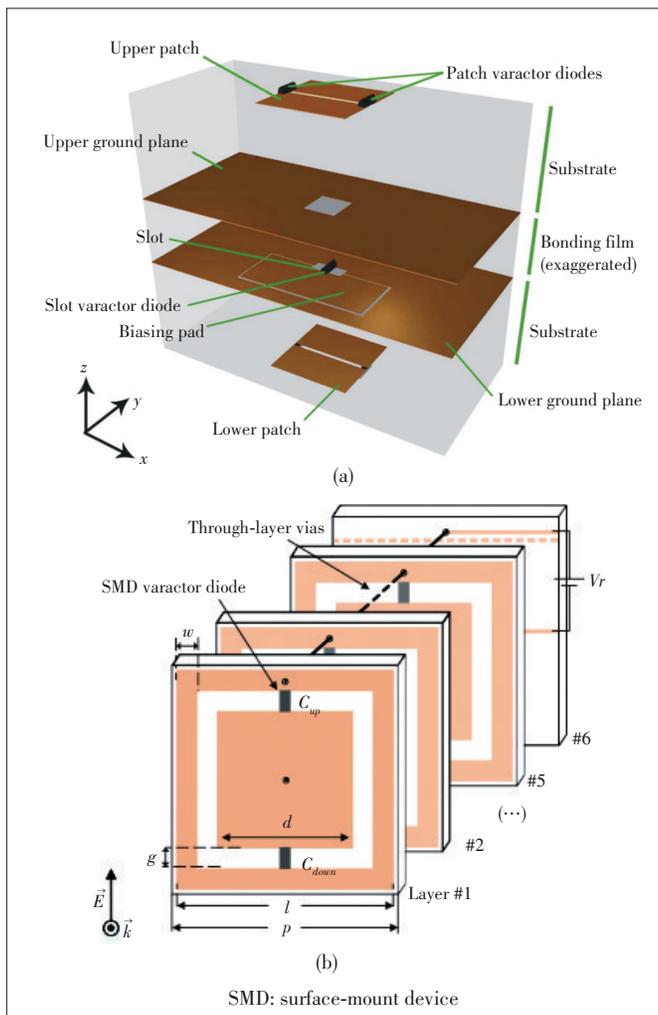
Some attempts have been done in Ref. [9] to reduce the element transmission insertion loss, and the schematic of the proposed element is displayed in Fig. 5(b). By combining a  $180^\circ$  analog phase shifter loaded with varactor diodes and a 1-bit phase shifter loaded with p-i-n diodes, the element is able to shift its transmission phase over  $360^\circ$ . By utilizing the simple current reversal mechanism, the 1-bit phase shifter circuit is integrated with the Tx patch instead of using a complex phase shifter, and the element insertion loss is effectively reduced. It is indicated that the simulated transmission insertion loss ranges from 0.95 dB to 1.35 dB at 5.6 GHz, with a 3-dB transmission loss bandwidth of 16.7%.

In Ref. [6], an element consisting of three resonant structures is exhibited. As displayed in Fig. 6(a), two microstrip patches embedded with varactor diodes operate as two resonant structures. They are positioned on the two sides of a ground plane and coupled by a slot aperture. A varactor diode is inserted into the slot to form the third resonant structure. Therefore, the resonant frequency of the radiating patch could be manipulated by changing the biasing voltages applied on the varactor diodes. The measured phase shift ranges at 5.5 GHz is only  $245^\circ$  while the transmission insertion loss reaches up to 2.7 – 5.7 dB.

Based on the stacked FSS structure, a continuous phase controlling RIS element at C band is detailed in Ref. [13]. As illustrated in Fig. 6(b), it is made of a five-layer stacked square slot and an extra layer to accommodate the biasing circuits. Each square slot FSS is loaded with two surface-mounted device (SMD) varactors and functions as a band-pass filter. This means if the frequency of the incident wave is close to the resonant frequency of the FSS, it can propagate through FSS. By adjusting the values of the SMD varactors, the resonant frequency of the square slot varies, and thus the phase response of the propagating wave can be controlled. Thanks to the cumulative effects of the phase response of each FSS, the stacked FSS element can generate a phase tuning range of  $360^\circ$  when the transmission insertion loss is within 3 dB in simulations. However, the operating bandwidth is only 90 MHz.

## 4 Conclusions

Recently, the RIS technology develops rapidly as the demands of low-cost and low power consumption beamforming systems increase. Various application scenarios of the RIS



▲ Figure 6. Transmissive reconfigurable intelligent surface (RIS) element designs with 360° phase controlling: (a) an element with three resonant structures in Ref. [6]; (b) 5-layer stacked frequency selective surface (FSS) structure in Ref. [13]

have been explored, such as wireless power delivery<sup>[35]</sup>, wireless coverage extension<sup>[36]</sup> and efficient channel estimation<sup>[37]</sup>. Specifically, the transmissive RIS has found its place in the indoor signal coverage. This paper reviews some of the latest developments of the transmissive RIS designs. The design methods utilized in most transmissive RIS designs are briefly introduced. Several representative designs with different phase resolution and polarization capabilities are presented and discussed. Currently, great efforts are being made to push the frequency frontier of the RIS technology towards the W and THz bands. Active devices are being incorporated to achieve signal boosting capability. Various RIS-aided systems are being developed to demonstrate the potential applications of the RIS technology in future wireless communications. It is believed that the RIS has a promising future.

## References

- [1] LIASKOS C, NIE S, TSIOLIARIDOU A, et al. A new wireless communication paradigm through software-controlled metasurfaces [J]. *IEEE communications magazine*, 2018, 56(9): 162 – 169. DOI: 10.1109/MCOM.2018.1700659
- [2] TANG W, LI X, DAI J, et al. Wireless communications with programmable metasurface: transceiver design and experimental results [J]. *China communications*, 2019, 16(5): 46 – 61. DOI: 10.23919/J.CC.2019.05.004
- [3] DAI L, WANG B, WANG M, et al. Reconfigurable intelligent surface-based wireless communications: antenna design, prototyping, and experimental results [J]. *IEEE access*, 2020, 8: 45913 – 45923. DOI: 10.1109/ACCESS.2020.2977772
- [4] LAU J Y, HUM S V. A wideband reconfigurable transmitarray element [J]. *IEEE transactions on antennas and propagation*, 2011, 60(3): 1303 – 1311. DOI: 10.1109/TAP.2011.2180475
- [5] LAU J Y, HUM S V. Reconfigurable transmitarray design approaches for beam-forming applications [J]. *IEEE transactions on antennas and propagation*, 2012, 60(12): 5679 – 5689. DOI: 10.1109/TAP.2012.2213054
- [6] LAU J Y, HUM S V. Analysis and characterization of a multipole reconfigurable transmitarray element [J]. *IEEE transactions on antennas and propagation*, 2010, 59(1): 70 – 79. DOI: 10.1109/TAP.2010.2090456
- [7] HUANG C, PAN W, MA X, et al. Using reconfigurable transmitarray to achieve beam-steering and polarization manipulation applications [J]. *IEEE transactions on antennas and propagation*, 2015, 63(11): 4801 – 4810. DOI: 10.1109/TAP.2015.2479648
- [8] HUANG C, PAN W, LUO X. Low-loss circularly polarized transmitarray for beam steering application [J]. *IEEE transactions on antennas and propagation*, 2016, 64(10): 4471 – 4476. DOI: 10.1109/TAP.2016.2586580
- [9] TANG J, XU S, YANG F, et al. Design and measurement of a reconfigurable transmitarray antenna with compact varactor-based phase shifters [J]. *IEEE antennas and wireless propagation letters*, 2021, 20(10): 1998 – 2002. DOI: 10.1109/LAWP.2021.3101891
- [10] FRANK M, LURZ F, WEIGEL R, et al. Electronically reconfigurable 6×6 element transmitarray at K-band based on unit cells with continuous phase range [J]. *IEEE antennas and wireless propagation letters*, 2019, 18(4): 796 – 800. DOI: 10.1109/LAWP.2019.2903338
- [11] BOCCIA L, RUSSO I, AMENDOLA G, et al. Multilayer antenna-filter antenna for beam-steering transmit-array applications [J]. *IEEE transactions on microwave theory and techniques*, 2012, 60(7): 2287 – 2300. DOI: 10.1109/TMTT.2012.2195673
- [12] REIS J R, COPNER N, HAMMOUDEH A, et al. FSS-inspired transmitarray for two-dimensional antenna beamsteering [J]. *IEEE transactions on antennas and propagation*, 2016, 64(6): 2197 – 2206. DOI: 10.1109/TAP.2016.2543802
- [13] REIS J R, CALDEIRINHA R F S, HAMMOUDEH A, et al. Electronically reconfigurable FSS-inspired transmitarray for 2-D beamsteering [J]. *IEEE transactions on antennas and propagation*, 2017, 65(9): 4880 – 4885. DOI: 10.1109/TAP.2017.2723087
- [14] CLEMENTE A, DUSSOPT L, SAULEAU R, et al. 1-bit reconfigurable unit cell based on PIN diodes for transmit-array applications in X-band [J]. *IEEE transactions on antennas and propagation*, 2012, 60(5): 2260 – 2269. DOI: 10.1109/TAP.2012.2189716
- [15] CLEMENTE A, DUSSOPT L, SAULEAU R, et al. Wideband 400-element electronically reconfigurable transmitarray in X band [J]. *IEEE transactions on antennas and propagation*, 2013, 61(10): 5017 – 5027. DOI: 10.1109/TAP.2013.2271493
- [16] DI PALMA L, CLEMENTE A, DUSSOPT L, et al. Experimental characterization of a circularly polarized 1 bit unit cell for beam steerable transmitarrays at Ka-band [J]. *IEEE transactions on antennas and propagation*, 2018, 67(2): 1300 – 1305. DOI: 10.1109/TAP.2018.2880095
- [17] PAN W, HUANG C, MA X, et al. A dual linearly polarized transmitarray element with 1-bit phase resolution in X-band [J]. *IEEE antennas and wireless propagation letters*, 2014, 14: 167 – 170. DOI: 10.1109/LAWP.2014.2358267
- [18] HUANG C, PAN W, MA X, et al. 1-bit reconfigurable circularly polarized transmitarray in X-band [J]. *IEEE antennas and wireless propagation letters*, 2015, 15: 448 – 451. DOI: 10.1109/LAWP.2015.2451697
- [19] WANG M, XU S, YANG F, et al. Design and measurement of a 1-bit reconfigurable transmitarray with subwavelength H-shaped coupling slot elements [J]. *IEEE transactions on antennas and propagation*, 2019, 67(5): 3500 – 3504.

- DOI: 10.1109/TAP.2019.2902676
- [20] WANG Y, XU S, YANG F, et al. A novel 1-bit wide-angle beam scanning reconfigurable transmitarray antenna using an equivalent magnetic dipole element [J]. *IEEE transactions on antennas and propagation*, 2020, 68(7): 5691 – 5695. DOI: 10.1109/TAP.2020.2964954
- [21] WANG M, XU S, YANG F, et al. A novel 1-bit reconfigurable transmitarray antenna using a C-shaped probe-fed patch element with broadened bandwidth and enhanced efficiency [J]. *IEEE access*, 2020, 8: 120124 – 120133. DOI: 10.1109/ACCESS.2020.3004435
- [22] XIAO Y, YANG F, XU S, et al. Design and implementation of a wideband 1-bit transmitarray based on a yagi-ivaldi unit cell [J]. *IEEE transactions on antennas and propagation*, 2021 69(7): 4229 – 4234. DOI: 10.1109/TAP.2020.3048496
- [23] NGUYEN B D, PICHOT C. Unit-cell loaded with PIN diodes for 1-bit linearly polarized reconfigurable transmitarrays [J]. *IEEE antennas and wireless propagation letters*, 2018, 18(1): 98 – 102. DOI: 10.1109/LAWP.2018.2881555
- [24] WANG M, XU S, YANG F, et al. A novel 1-bit dual-linear polarized reconfigurable transmitarray element using double-layer dipoles [C]//*IEEE Asia-Pacific Microwave Conference (APMC)*. IEEE, 2019: 685 – 687. DOI: 10.1109/APMC46564.2019.9038225
- [25] MUNINA I, TURALCHUK P, VEREVKIN A, et al. A study of C-band 1-bit reconfigurable dual-polarized transmitarray [C]//*13th European Conference on Antennas and Propagation (EuCAP)*. EurAAP, 2019: 1 – 5
- [26] PHAM K, SAULEAU R, CLEMENTE A, et al. Electronically reconfigurable unit-cell and transmitarray in dual-linear polarization at Ka-band [C]//*13th European Conference on Antennas and Propagation (EuCAP)*. EurAAP, 2019: 1 – 4
- [27] WANG Y, XU S, YANG F, et al. 1-bit dual-linear polarized reconfigurable transmitarray antenna using asymmetric dipole elements with parasitic bypass dipoles [J]. *IEEE transactions on antennas and propagation*, 2021, 69(2): 1188 – 1192. DOI: 10.1109/TAP.2020.3005713
- [28] CHENG C C, ABBASPOUR-TAMIJANI A. Study of 2-bit antenna-filter-antenna elements for reconfigurable millimeter-wave lens arrays [J]. *IEEE transactions on microwave theory and techniques*, 2006, 54(12): 4498 – 4506. DOI: 10.1109/TMTT.2006.885993
- [29] CHENG C C, LAKSHMINARAYANAN B, ABBASPOUR-TAMIJANI A. A programmable lens-array antenna with monolithically integrated MEMS switches [J]. *IEEE transactions on microwave theory and techniques*, 2009, 57(8): 1874 – 1884. DOI: 10.1109/TMTT.2009.2025422
- [30] DIABY F, CLEMENTE A, DI PALMA L, et al. Design of a 2-bit unit-cell for electronically reconfigurable transmitarrays at Ka-band [C]//*47th European Microwave Conference (EuMC)*. EuMA, 2017: 1321 – 1324. DOI: 10.23919/EuMC.2017.8231095
- [31] TANG J, XU S, YANG F. Design of a 2.5-D 2-bit reconfigurable transmitarray element for 5G mmwave applications [C]//*IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting*. IEEE, 2020: 631 – 632. DOI: 10.1109/IEEECONF35879.2020.9330153
- [32] WU B, SUTINJO A, POTTER M E, et al. On the selection of the number of bits to control a dynamic digital MEMS reflectarray [J]. *IEEE antennas and wireless propagation letters*, 2008, 7: 183 – 186. DOI: 10.1109/LAWP.2008.920908
- [33] YANG H, YANG F, XU S, et al. A study of phase quantization effects for reconfigurable reflectarray antennas [J]. *IEEE antennas and wireless propagation letters*, 2017, 16: 302 – 305. DOI: 10.1109/LAWP.2016.2574118
- [34] WU Q, ZHANG R. Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts [J]. *IEEE transactions on communications*, 2020, 68(3): 1838 – 1851. DOI: 10.1109/TCOMM.2019.2958916
- [35] BUCKMASTER J G, LEE T H. An electronically steerable millimeter-wave reflectarray for wireless power delivery [C]//*50th European Microwave Conference (EuMC)*. EuMA, 2021: 514 – 517. DOI: 10.23919/EuMC48046.2021.9338208
- [36] ZENG S, ZHANG H, DI B, et al. Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization [J]. *Communications letters*, 2020, 25(1): 269 – 273. DOI: 10.1109/LCOMM.2020.3025345
- [37] MILLER C L, DMOCHOWSKI P A, SMITH P J. Efficient channel estimation for RIS [C]//*IEEE International Conference on Communications*. IEEE, 2021: 1 – 6. DOI: 10.1109/ICC42927.2021.9500469

## Biographies

**TANG Junwen** (tjw17@mails.tsinghua.edu.cn) received the B.S. degree in electronic information engineering from Xidian University, China in 2017, and she is currently pursuing the Ph.D. degree at Department of Electronic Engineering, Tsinghua University, China. Her current research interests include antenna theory and reconfigurable antenna.

**XU Shenheng** (shxu@tsinghua.edu.cn) received the B.S. and M.S. degrees from Southeast University, China in 2001 and 2004, respectively and the Ph.D. degree in electrical engineering from the University of California at Los Angeles (UCLA), USA in 2009. From 2000 to 2004, he was a research assistant with the State Key Laboratory of Millimeter Waves, Southeast University, China. From 2004 to 2011, he was a graduate student researcher and then became a post-doctoral researcher with the Antenna Research, Analysis, and Measurement Laboratory, UCLA. In 2012, he joined the Department of Electronic Engineering, Tsinghua University, China, as an associate professor. His current research interests include novel designs of high-gain antennas for advanced applications, artificial electromagnetic structures, and electromagnetic and antenna theories.

**YANG Fan** received the B.S. and M.S. degrees from Tsinghua University, China and the Ph.D. degree from the University of California at Los Angeles (UCLA), USA. He joined the Electrical Engineering Department, University of Mississippi, USA, as an assistant professor and was promoted to an associate professor in 2009. In 2011, he joined the Electronic Engineering Department, Tsinghua University, as a professor. He has authored or coauthored more than 300 journal articles and conference articles, six book chapters, and five books. Dr. YANG is an ACES Fellow, IEEE Fellow, and a recipient of several prestigious awards and recognition, including the Young Scientist Awards of the 2005 URSI General Assembly and of the 2007 International Symposium on Electromagnetic Theory, the 2008 Junior Faculty Research Award of the University of Mississippi, the 2009 Inaugural IEEE Donald G. Dudley Jr. Undergraduate Teaching Award, and the 2011 Recipient of Global Experts Program of China. He was the Technical Program Committee (TPC) Chair of 2014 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting. He is also an IEEE APS Distinguished Lecturer from 2018 to 2020. His current research interests include antennas, surface electromagnetics, computational electromagnetics and applied electromagnetic systems.

**LI Maokun** received the B.S. degree in electronic engineering from Tsinghua University, China in 2002, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign, USA in 2004 and 2007, respectively. After graduation, he worked as a senior research scientist at Schlumberger-Doll Research, USA. Since 2014, he has joined the Department of Electronic Engineering, Tsinghua University. He has authored one book chapter, 50 journal articles and 120 conference proceedings, and held three patent applications. His research interests include fast algorithms in computational electromagnetics and their applications in antenna modeling, electromagnetic compatibility analysis, and inverse problems. Dr. LI was a recipient of the China National 1 000 Plan in 2014 and the 2017 IEEE Ulrich L. Rohde Innovative Conference Paper Award. He also serves as the Associate Editor for the *IEEE Journal on Multiscale and Multiphysics Computational Techniques* and *Applied Computational Electromagnetic Society Journal*, and the Guest Editor for the special issue on “Electromagnetic Inverse Problems for Sensing and Imaging” in *IEEE Antennas and Propagation Magazine*.

# IRS-Enabled Spectrum Sharing: Interference Modeling, Channel Estimation and Robust Passive Beamforming



GUAN Xinrong<sup>1</sup>, WU Qingqing<sup>2</sup>

(1. College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China;

2. State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau 999078, China)

DOI: 10.12142/ZTECOM.202201005

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220303.1145.004.html>, published online March 6, 2022

Manuscript received: 2021-11-11

**Abstract:** Intelligent reflecting surface (IRS), with its unique capability of smartly reconfiguring wireless channels, provides a new solution to improving spectrum efficiency, reducing energy consumption and saving deployment/hardware cost for future wireless networks. In this paper, IRS-enabled spectrum sharing is investigated, from the perspectives of interference modeling, efficient channel estimation and robust passive beamforming design. Specifically, we first characterize the interference in a spectrum sharing system consisting of a single primary user (PU) pair and a single secondary user (SU) pair, and extend it to the large-scale network by leveraging the Poisson point process (PPP). Then, we propose an efficient channel estimation framework based on decoupling the cascaded IRS channels. Moreover, the tradeoff between spectrum efficiency and energy efficiency is derived from the view of channel estimation accuracy. Finally, we discuss the robust passive beamforming design in presence of imperfect channel estimation and nonideal/discrete phase shifts. It is hoped that this paper provides useful guidance for unlocking the full potential of IRS for achieving efficient spectrum sharing for future wireless networks.

**Keywords:** intelligent reflecting surface; spectrum sharing; channel estimation; passive beamforming

**Citation** (IEEE Format): X. R. Guan and Q. Q. Wu, "IRS-enabled spectrum sharing: interference modeling, channel estimation and robust passive beamforming," *ZTE Communications*, vol. 20, no. 1, pp. 28 - 35, Mar. 2022. doi: 10.12142/ZTECOM.202201005.

## 1 Introduction

Over the past few decades, the rapid development of wireless applications and exponentially grown wireless devices have led to an ever-increasing demand for a high data rate and seamless coverage. To fulfill these requirements, a variety of wireless technologies have been proposed by academics and engineers, such as ultra-dense network (UDN), massive multiple-input multiple-output (MIMO), millimeter wave (mmWave) communication, and so on. However, for the future beyond fifth-generation (5G) wireless network, it becomes insufficient to rely solely on the above technologies, when considering various emerging services and applications, such as 3D holographic imaging and presence, 5D communications (sight, hearing, touch, smell and taste), and the Internet of Everything (IoE), with more challeng-

ing performance requirements, e. g., 100 Gbit/s to 1 Tbit/s peak data rates,  $100/\text{m}^3$  device density, and 10 times more energy efficient than 5G<sup>[1]</sup>. For example, seeking more spectrum resources in the mmWave and even terahertz (THz) bands inevitably suffers from path loss and blockage. Though densely deployed base stations (BSs) and/or substantially increased antennas at the BSs can help to overcome the above problems, they are faced with practical challenges in terms of energy consumption, hardware cost and signal processing.

Recently, due to the developments in metamaterials/meta-surfaces, intelligent reflecting surface (IRS) has emerged as an energy-efficient and cost-effective technology to solve the above problems<sup>[2-4]</sup>. Specifically, IRS is a planar surface consisting of a large number of passive reflecting elements, each of which can independently reflect the incident signal with a tunable phase and/or amplitude. By dynamically adjusting the reflecting coefficients to adapt to the wireless propagation environment, the signal can be enhanced/suppressed in desired/undesired directions. Since IRS requires no power-hungry components, e. g., radio frequency (RF) chains, it has the potential to achieve near-zero power consumption by exploiting wireless energy harvesting module<sup>[5]</sup>. Moreover, IRS can be integrated into existing wireless networks without changes at the transceivers, thus providing great flexibility in practical de-

Corresponding author: Wu Qingqing

The work of GUAN Xinrong was supported by the National Natural Science Foundation of China under Grant No. 62171461 and Natural Science Foundation on Frontier Leading Technology Basic Research Project of Jiangsu under Grant No. BK20212001. The work of WU Qingqing was supported by the Macau Science and Technology Development Fund, Macau SAR, under Grant Nos. 0119/2020/A3, SKL-IOTSC-2021-2023 and 0108/2020/A, the Guangdong NSF under Grant No. 2021A1515011900, and the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University, under Grant No. 2021D15.

ployment. Owing to the above advantages, IRS has been extensively studied in various wireless systems, such as non-orthogonal multiple access (NOMA)<sup>[6-7]</sup>, simultaneous wireless information and power transfer (SWIPT)<sup>[8-9]</sup>, secrecy communications<sup>[10-12]</sup>, and so on.

On the other hand, spectrum sharing has been thoroughly investigated in the literature as an efficient way to solve the spectrum scarcity problem, by fully exploiting the opportunities within the limited spectrum resource rather than resorting to higher frequency bands. Provided that the quality of service (QoS) for primary users (PUs) is ensured, secondary users (SUs) can share the same spectrum thus the spectrum efficiency can be significantly improved. However, the SU performance largely decreases in the presence of strong cross-link interference. For instance, considering a spectrum sharing communication system consisting of a pair of PU and a pair of SU, the SU rate would be constrained by the strong interference from PT if the primary transmitter is located near the secondary receiver. Alternatively, if the secondary transmitter is located near the primary receiver, the SU rate would also be constrained since the transmission power of the secondary transmitter should be tuned at an extremely low level for ensuring the reception at the primary receiver. As such, the spectrum sharing efficiency largely depends on the interference distribution. Note that with increased device density, the future wireless network will take the interference challenge to a new level, which will thus limit the application potential of spectrum sharing.

With capability for signal enhancement as well as interference suppression, IRS is a promising solution to tackling the above spectrum sharing challenges. Fig. 1 shows three typical setups for IRS-enabled spectrum sharing. In particular, in Setup 1, by deploying an IRS nearby the SU cluster, the interference from the primary BS (PBS) can be efficiently reduced at the SUs, while the desired signal from the secondary BS (SBS) can be enhanced. As such, the transmission power at the SBS can be largely decreased such that the QoS at the PUs can be guaranteed, even if PUs are located near the SBS. In

Setup 2, an IRS is deployed nearby the SBS. In this case, the transmission of SBS can be properly tuned so that the signal/interference at the SUs/PUs can be enhanced/reduced. In Setup 3, considering that an IRS is deployed nearby the PU cluster. Then, by mitigating the SBS-PU interference, the SBS can enlarge its transmission power for combating the PBS's interference at the SUs.

However, most existing works focus on the IRS-aided point-to-point transmission or broadcast system, while its promising channel reconfiguration ability for spectrum sharing, i.e., interference channel, has not been well studied. Motivated by this, we investigate the IRS-enabled spectrum sharing, from the perspectives of interference modeling, efficient channel estimation and robust passive beamforming design. The main contributions of this work are summarized as follows. 1) An interference-based performance analysis model is established to evaluate the passive beamforming gain of IRS assisted spectrum sharing system. 2) An efficient channel estimation method based on decoupling the cascaded reflecting channel is proposed. Moreover, we derive the tradeoff between energy efficiency and spectrum efficiency from the perspective of channel estimation accuracy. 3) A robust passive beamforming design framework is proposed while considering the imperfect channel estimation and nonideal/discrete reflection coefficients in practice.

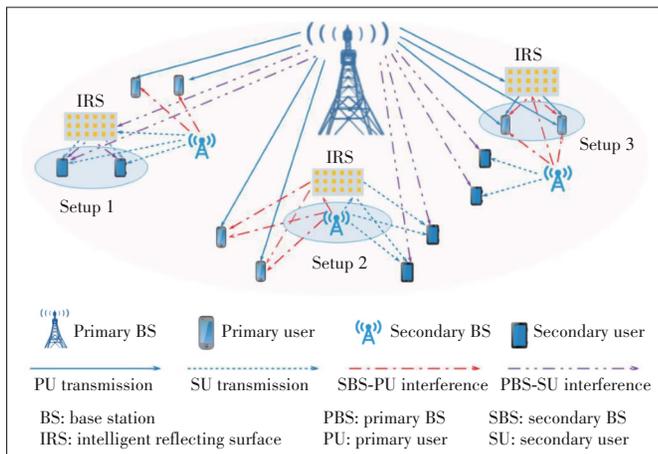
The rest of this paper is organized as follows. Section 2 presents an interference-based performance analysis model. Section 3 discusses the efficient channel estimation and Section 4 investigates the robust passive beamforming design. Section 5 provides the numerical results to evaluate the performance of the proposed designs. Finally, we conclude this paper in Section 6.

## 2 Interference Modeling

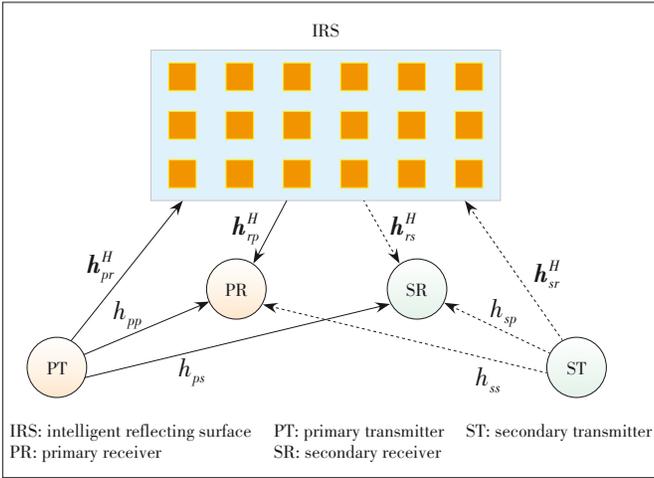
It is not easy to evaluate the passive beamforming gain of IRS in spectrum sharing systems due to the complicated links among PUs and SUs. However, considering that the bottleneck of spectrum sharing is the interference, starting from interference modeling is expected to provide some useful insights. Thus, in this section, we aim to characterize the performance gain of IRS from the view of interference. In particular, we first consider the IRS-enabled spectrum sharing system consisting of a single PU pair and a single SU pair, and then extend the analysis to the more general and practical large-scale network setup with massive users.

### 2.1 Case 1: Single PU Pair and Single SU Pair

Fig. 2 shows an IRS-enabled spectrum sharing system, where an SU link consisting of a secondary transmitter (ST) and a secondary receiver (SR) coexists with a PU link consisting of a primary transmitter (PT) and a primary receiver (PR), and an IRS is deployed to assist in the spectrum sharing. Assume that all nodes are equipped with a single antenna, while the number of reflecting elements at the IRS is denoted by  $N$ . The baseband equivalent channels from the PT(ST) to the PR,



▲ Figure 1. IRS-enabled spectrum sharing with different setups



▲ Figure 2. IRS-enabled spectrum sharing with single primary-user/secondary-user (PU/SU) pair

SR and IRS are denoted by  $h_{pp}$ ,  $h_{ps}$ ,  $\mathbf{h}_{pr} \in \mathbb{C}^{N \times 1}$ ,  $h_{sp}$ ,  $h_{ss}$  and  $\mathbf{h}_{sr}^H \in \mathbb{C}^{1 \times N}$ , respectively, while those from the IRS to the PR and SR are denoted by  $\mathbf{h}_{rp}^H \in \mathbb{C}^{1 \times N}$  and  $\mathbf{h}_{rs}^H \in \mathbb{C}^{1 \times N}$ , respectively. Let  $\mathbf{v}^H = [v_1, v_2, \dots, v_N]$  represent the passive beamforming vector of the IRS, where  $v_n = e^{j\theta_n}$ ,  $\theta_n \in [0, 2\pi)$  is the phase shift on the combined incident signal by its  $n$ -th element. The composite PT/ST-IRS-PR/SR channel is then modeled as a concatenation of three components, namely, the PT/ST-IRS link, the IRS reflecting with phase shifts, and the IRS-PR/SR link. The quasi-static flat-fading model is assumed for all channels. We assume that the channel state information (CSI) of all channels involved is perfectly known at the ST/IRS for the joint power control and passive beamforming design.

Assume that the transmission power at the PT and ST is given by  $p_p$  and  $p_s$ , respectively, while the complex additive white Gaussian noise (AWGN) at the PR and SR is denoted by  $n_p \sim CN(0, \sigma_p^2)$  and  $n_s \sim CN(0, \sigma_s^2)$ , respectively. As a result, the received interference power at the PR and SR is respectively given by

$$\begin{cases} I_p = p_s \left| h_{sp} + \mathbf{v}^H \mathbf{h}_{srp} \right|^2 \\ I_s = p_p \left| h_{ps} + \mathbf{v}^H \mathbf{h}_{prs} \right|^2 \end{cases} \quad (1)$$

where  $\mathbf{h}_{srp} = \text{diag}(\mathbf{h}_{rp}^H) \mathbf{h}_{sr}^*$  and  $\mathbf{h}_{prs} = \text{diag}(\mathbf{h}_{rs}^H) \mathbf{h}_{pr}^*$ . Traditionally, since the interference at the PR, i.e.,  $I_p$  should be constrained below a certain threshold such that the QoS at the PR can be ensured, the transmission power of ST, i.e.,  $p_s$  should be very low in the presence of strong ST-PR link, which thus results in an extremely low SU data rate. Fortunately, by deploying an IRS, the above strong interference can be effectively mitigated with that from the reconfigurable reflecting channel. As such, the transmission power at the ST can be largely improved as well as the SU data rate.

Motivated by the above, we focus on the ST-PR interference

$I_p$ , of which the minimum value can be given by

$$\begin{aligned} I_p^{\min} &= \min_{\mathbf{v}} p_s \left| h_{sp} + \mathbf{v}^H \mathbf{h}_{srp} \right|^2, \\ \text{s.t. } &|v|_n = 1. \end{aligned} \quad (2)$$

It can be observed that the maximum power gain of the reflected channel can be expressed by

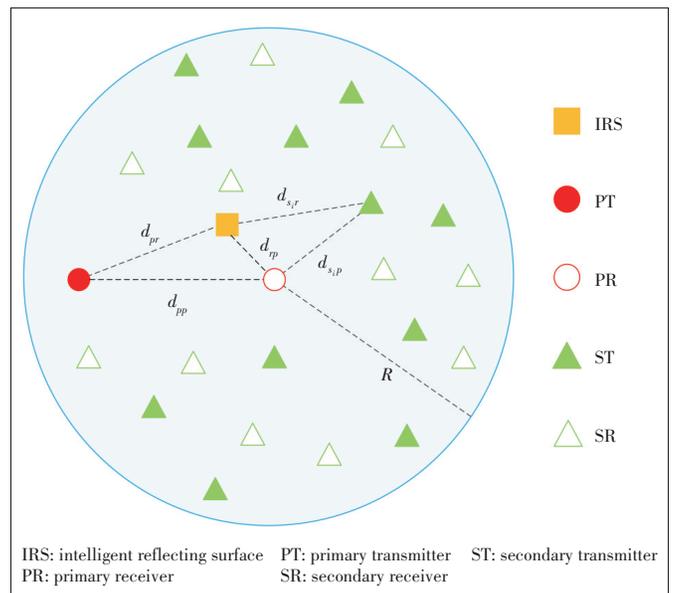
$$I_R = \max_{\mathbf{v}} \left| \mathbf{v}^H \mathbf{h}_{srp} \right|^2 = \left| \sum_{n=1}^N \mathbf{h}_{srp}(n) \right|^2, \quad (3)$$

which represents the maximum capability of the IRS for interference mitigating. In particular, if  $I_R \geq |h_{sp}|^2$ , the interference at the PR can be completely reduced, i.e.,  $I_p^{\min} = 0$ , via properly designed passive beamforming. Otherwise, the minimum interference at the PR is then given by

$$I_p^{\min} = p_s \left( \left| h_{sp} \right| - \left| \sum_{n=1}^N \mathbf{h}_{srp}(n) \right| \right)^2. \quad (4)$$

## 2.2 Case 2: Large-Scale Network with Massive Users

In practical systems, the interference modeling would be much more complex due to massive users. In this subsection, we still focus on the interference at the PR side, which is crucial to the maximum allowed transmission power at STs. Fig. 3 shows IRS-enabled spectrum sharing in a large-scale network, where an IRS is deployed nearby the PR to reduce the interference from multiple STs. Denote the distances from PT to PR and IRS by  $d_{pp}$  and  $d_{pr}$ , respectively, that from IRS to PR by  $d_{rp}$ , and those from the  $i$ -th ST to PR and IRS by  $d_{s_i p}$  and  $d_{s_i r}$ , respectively. Assuming that the distribution of STs in the considered circle area with radius  $R$  follows a Poisson point process (PPP) with density  $\lambda_s$ . Denoting the number of STs by  $K$ ,



▲ Figure 3. IRS-enabled spectrum sharing in large-scale network

the probability of  $K=k$  is then expressed by

$$\Pr(K = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \geq 0. \quad (5)$$

Moreover, considering both the path loss and small-scale fading, the channels are given by

$$\mathbf{h}_{ij} = L_{ij} \mathbf{g}_{ij}, \quad i \in \{p, s_i, r\}, \quad j \in \{r, p\}, \quad i \neq j, \quad (6)$$

where  $L_{ij} = L_0 d_{ij}^{-c}$  denotes the path loss,  $L_0$  is the path loss at a reference distance of 1 m, and  $c$  is the corresponding path loss exponent, while  $\mathbf{g}_{ij}$  follows complex Gaussian distribution with zero means and unit variances. Based on the above, the power of total interference at the PR can be written as

$$I_p = \sum_{i=1}^k p_{s_i} \left| L_{s_i p} \mathbf{g}_{s_i p} + \sum_{n=1}^N L_{s_r} L_{r p} \mathbf{g}_{s_r} (n) \mathbf{g}_{r p} (n) v_n \right|^2. \quad (7)$$

It can be observed that the interference power  $I_p$  largely depends on a variety of distances, which are random variables themselves. To obtain the statistical characteristics of  $I_p$ , we first derive its eigenfunction by

$$\varphi_{I_p}(t) = \sum_{k=0}^{\infty} \Pr\{K = k\} \mathbb{E}\left(e^{jtI_p} | k\right). \quad (8)$$

Then, by exploiting the Fourier transformation, the probability density function (PDF) of  $I_p$  can be obtained. Note that without IRS, the interference power  $I_p$  in Eq. (7) would be  $I_p = \sum_{i=1}^k p_{s_i} \left| L_{s_i p} \mathbf{g}_{s_i p} \right|^2$ . Comparing the PDF of  $I_p$  in two cases, some useful insights can be drawn into the performance gain of IRS for spectrum sharing.

### 3 Efficient Channel Estimation

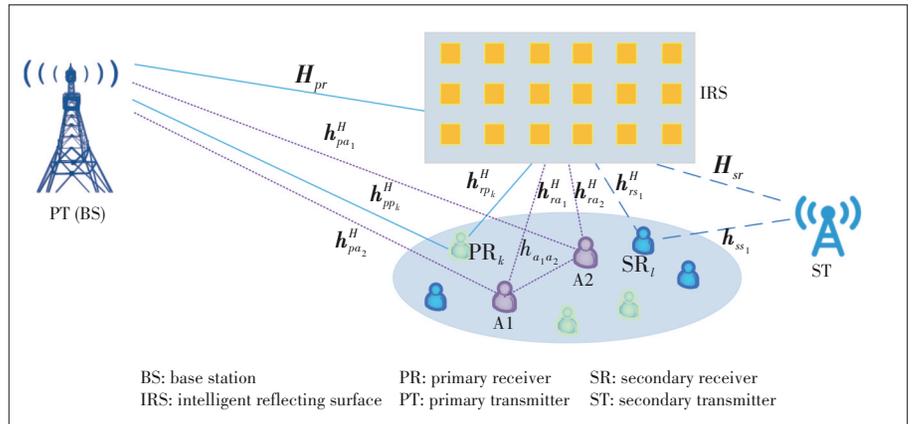
In IRS-aided wireless systems, channel estimation is fundamentally challenging due to the following reasons<sup>[13]</sup>. First, the passive IRS elements can only reflect signals without the capability of signal transmission/reception, which thus makes the separate estimation of the BS-IRS and IRS-user channels infeasible. Alternatively, a practical approach is to estimate the cascaded BS-IRS-user channels based on the training signals from the BS/users with properly designed IRS reflection pattern over time<sup>[14]</sup>. Second, since the number of IRS reflecting elements is generally very large, the training overhead required for cascaded channel estimation becomes prohibitively high. To reduce the complexity, an element-grouping based channel estimation strategy was proposed in Ref. [15]. In particular, by grouping the adja-

cent IRS elements into a sub-surface, only the cascaded user-IRS-BS channel associated with each sub-surface needs to be estimated. Moreover, another approach to reduce the pilot overhead is exploiting the IRS channel sparsity, which however is usually applicable in the IRS-assisted communication system operating at high-frequency bands<sup>[16]</sup>. It should be noted that the above works cannot be straightforwardly extended to the IRS-aided multiuser system, since they mainly focus on channel estimation for the IRS-aided single-user system. As the number of users increases, the required pilot overhead becomes unaffordable by exploiting the above user-by-user successive channel estimation methods. Especially, for the IRS-enabled spectrum sharing system, in addition to increased transmission links, more interference links are involved, which further makes the channel estimation even more challenging.

In this section, we first propose an efficient channel estimation method for the IRS-enabled spectrum sharing system, which is motivated by the cascaded channel decoupling<sup>[17]</sup>. Then, we discuss the tradeoff between spectrum efficiency and energy efficiency.

#### 3.1 Cascaded Channel Decoupling Based Channel Estimation

Take the IRS-enabled spectrum sharing system shown in Fig. 4 as an example, which consists of a multi-antenna PT, a multi-antenna ST, an IRS and multiple single-antenna PRs and SRs. It can be observed that: 1) All cascaded PT-IRS-PR/ST-IRS-SR channels share the common PT-IRS/ST-IRS channels. If such common channels can be obtained, only IRS-PR/IRS-SR channels need to be estimated for recovering the desired cascaded channels. Meanwhile, the training overhead can be largely reduced since the IRS-PR/IRS-SR channels are of much lower dimension as compared to the cascaded ones. 2) Since the locations of PT/ST and IRS are fixed, the above common channels usually vary much more slowly than the IRS-PR/IRS-SR channels, which thus only need to be esti-



▲ Figure 4. Cascaded channel decouple based channel estimation for IRS-enabled spectrum sharing system

mated offline. 3) Given the estimated PT-IRS/ST-IRS and IRS-SR/IRS-PR channels, the interference channels, i.e., PT-IRS-SR and ST-IRS-PR channels, can be directly obtained via computation, without additional training. Motivated by the above, we propose an efficient channel estimation method based on the cascaded channel decoupling, which is specified as follows and where the estimation of direct channels is omitted for brevity since the focus is the IRS involved channels.

### 3.1.1 Off-Line Estimation

As shown in Fig. 4, two anchor nodes, namely A1 and A2, are deployed nearby the IRS to assist in channel estimation in the considered IRS-enabled spectrum sharing system. First, A1 transmits pilot symbols, while PT, ST and A2 estimate  $\mathbf{H}_{pra_1} = \text{diag}(\mathbf{h}_{ra_1}^H)\mathbf{H}_{pr}$ ,  $\mathbf{H}_{sra_1} = \text{diag}(\mathbf{h}_{ra_1}^H)\mathbf{H}_{sr}$  and  $\mathbf{h}_{a_1ra_2} = \text{diag}(\mathbf{h}_{ra_1}^H)\mathbf{h}_{ra_2}^*$ , respectively. Second, A2 feeds back the estimated  $\mathbf{h}_{a_1ra_2}$  to PT and ST. Third, A2 transmits pilot symbols while PT and ST estimate  $\mathbf{H}_{pra_2} = \text{diag}(\mathbf{h}_{ra_2}^H)\mathbf{H}_{pr}$  and  $\mathbf{H}_{sra_2} = \text{diag}(\mathbf{h}_{ra_2}^H)\mathbf{H}_{sr}$ , respectively. At last, based on the above estimated  $\mathbf{H}_{pra_1}$ ,  $\mathbf{h}_{a_1ra_2}$  and  $\mathbf{H}_{pra_2}$ , the separate PT-IRS, IRS-A1 and IRS-A2 channels can be obtained at PT<sup>[13]</sup>. Similarly, the ST-IRS channel can be obtained at ST based on the estimated  $\mathbf{H}_{sra_1}$ ,  $\mathbf{h}_{a_1ra_2}$  and  $\mathbf{H}_{sra_2}$ . Note that all of the above channels need to be estimated only once over a long time due to the fixed locations of the PT/ST, IRS and A1/A2.

### 3.1.2 On-Line Estimation

First, PRs consecutively transmit pilot symbols, while the PT estimates the IRS-PR channels efficiently by exploiting the offline estimated common PT-IRS channel. Then, the PT computes the desired PT-IRS-PR channels based on the offline estimated PT-IRS channel and the online estimated IRS-PR channel. Second, SRs consecutively transmit pilot symbols while the ST estimates the IRS-SR channels efficiently by exploiting the offline estimated common ST-IRS channel. After that, the ST can compute the desired ST-IRS-SR channels. Third, the PT and ST exchange the estimated IRS-PR and IRS-SR channels, such that they can obtain the CSI of the interference links, i.e., PT-IRS-SR/ST-IRS-PR without additional pilot training.

## 3.2 Tradeoff Between Spectrum Efficiency and Energy Efficiency

Obviously, a training-transmitting tradeoff is involved: too few training results in a coarsely estimated channel at the PT/ST and hence a reduced passive beamforming gain, whereas too much training costs excessive energy and also leaves less time for data transmission. To show the essential insight, we consider the IRS-enabled spectrum sharing system in Fig. 2 with single PU/SU pair for brevity. As shown in Fig. 5, assuming a channel coherence block of  $T_c$  (normalized by symbol duration), the time

allocated for channel estimation and data transmission is denoted by  $T_e$  and  $T_d$ , respectively. Denoting the pilot power at PT/ST and the noise power by  $p_0$  and  $\sigma_0^2$ , the mean-squared error (MSE) by applying the channel estimation method in Ref. [14] is given by  $\sigma_0^2/(p_0T_e)$ . Thus, the channel estimation accuracy improves with larger  $p_0$  and  $T_e$ , which thus leads to increased signal-to-interference-plus-noise ratio (SINR) at the PR and SR during the data transmission phase, i.e.,

$$\gamma_p(p_0, T_e), \gamma_s(p_0, T_e) \propto p_0, T_e. \quad (9)$$

Note that in spectrum sharing systems, PUs would not adapt their QoS requirement to SUs, while SUs should adjust their transmission strategy even towards a low data rate. If a QoS constraint on PU transmission as  $\gamma_p(p_0, T_e) \geq \gamma_{th}$ , where  $\gamma_{th}$  is the PU SINR target and corresponds to a fixed PU data rate  $\log_2(1 + \gamma_{th})$ , the outage probability of PU transmission is obtained by  $\Pr\{\gamma_p(p_0, T_e) < \gamma_{th}\}$ . Based on the above, the average throughput of the spectrum sharing system during each  $T_c$  can be expressed by

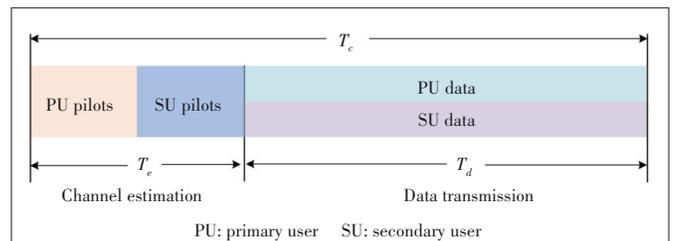
$$\bar{\eta} = \left\{ \Pr\{\gamma_p(p_0, T_e) \geq \gamma_{th}\} \log_2(1 + \gamma_{th}) + \log_2(1 + \gamma_s(p_0, T_e)) \right\} (1 - T_e/T_c). \quad (10)$$

One can observe that larger training time  $T_e$  on one hand results in lower outage probability for PU transmission and a higher data rate for SU transmission, but on the other hand renders a lower spectrum efficiency for leaving less time for data transmission.

Next, we consider the energy efficiency of the above channel estimation and data transmission protocol. The transmission power of the PT and ST in the data transmission phase is denoted by  $p_p$  and  $p_s$ , respectively, while the circuit power consumption in  $T_c$  (including that consumed in PT/PR/ST/SR/IRS) denoted by  $p_c$ . Thus, the system energy consumption in each  $T_c$  is given by

$$E = p_0T_e + T_d(p_p + p_s) + T_cp_c. \quad (11)$$

Based on Eqs. (10) and (11), the energy efficiency can be obtained by



▲ Figure 5. A channel estimation and data transmission protocol for the IRS-enabled spectrum sharing system in Fig. 2

$$\lambda = \frac{\bar{\eta}}{E} = \frac{\left\{ \Pr\left\{ \gamma_p(p_0, T_0) \geq \gamma_{th} \right\} \log_2(1 + \gamma_{th}) + \log_2(1 + \gamma_s(p_0, T_0)) \right\} (1 - T_c/T_e)}{p_0 T_e + T_d(p_p + p_s) + T_c p_c}. \quad (12)$$

Next, the tradeoff between spectrum efficiency and energy efficiency can be derived by

$$f = f(\bar{\eta}, E), \quad (13)$$

which can further be exploited to optimize the parameters settings, such as time/power allocation between the channel estimation phase and data transmission phase.

## 4 Robust Passive Beamforming Design

Most of the existing passive beamforming designs are based on the assumption of perfect CSI, which is however difficult to obtain in practice due to inevitable estimation errors arising from noise, feedback latency, limited pilot power, and so on<sup>[18]</sup>. Another assumption made in most existing works is that the IRS phase shifts optimized are continuous values, which is also challenging to implement in practice due to hardware limitations and cost control. Thus, implementing discrete and finite reflection phase shifts is more cost-effective. In this section, we propose robust passive beamforming designs with imperfect CSI and discrete phase shifts, respectively.

### 4.1 With Imperfect CSI

Considering the IRS-enabled spectrum sharing system shown in Fig. 2, we assume imperfect CSI for passive beamforming design. Denote  $\bar{\mathbf{h}}_{ij}^T = [h_{ij}, h_{ij}]$ ,  $i, j \in \{p, s\}$  and  $\bar{\mathbf{v}}^H = e^{j\omega} [\mathbf{v}^H, 1]$ , while other notations are the same as those in Section 2. We assume that the CSI errors are bounded by some possible values, thus the channels are modeled by

$$\bar{\mathbf{h}}_{ij} = \hat{\mathbf{h}}_{ij} + \Delta \mathbf{h}_{ij}, \left\| \Delta \mathbf{h}_{ij} \right\|^2 \leq \xi, \quad (14)$$

where  $\xi$  is the radius of the uncertainty region. We aim to maximize the achievable SU rate via jointing the power control at the ST and the passive beamforming at the IRS, subject to the SINR constraint at the PR. Thus, the optimization problem is formulated as

$$\begin{aligned} \text{(P1): } \max_{p_s, \bar{\mathbf{v}}} & \frac{p_s \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{ss} + \Delta \mathbf{h}_{ss}) \right|^2}{p_p \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{ps} + \Delta \mathbf{h}_{ps}) \right|^2 + \sigma_s^2} \\ \text{s.t. } & \frac{p_p \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{pp} + \Delta \mathbf{h}_{pp}) \right|^2}{p_s \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{sp} + \Delta \mathbf{h}_{sp}) \right|^2 + \sigma_p^2} \geq \gamma_{th}, \\ & p_s \leq P_{\max}, \left\| \Delta \mathbf{h}_{ij} \right\|^2 \leq \xi, i, j \in \{p, s\}, \\ & \left| \bar{v}_n \right| = 1, n = 1, \dots, N + 1. \end{aligned} \quad (15)$$

An alternating optimization (AO) based algorithm can be used to solve (P1) sub-optimally, by iteratively optimizing one of  $p_s$  and  $\bar{\mathbf{v}}$  with the other being fixed at each iteration until the convergence is reached. Specifically, for given  $\bar{\mathbf{v}}$ , the optimal  $p_s$  can be easily obtained, whereas optimizing  $\bar{\mathbf{v}}$  for given  $p_s$  is more challenging due to the channel estimation errors. By introducing variables  $t$ ,  $\alpha$  and  $\beta$ , (P1) can be rewritten as

$$\begin{aligned} \text{(P2): } \max_{\bar{\mathbf{v}}} & t \\ \text{s.t. } & p_s \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{ss} + \Delta \mathbf{h}_{ss}) \right|^2 \geq t\alpha, \\ & p_p \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{ps} + \Delta \mathbf{h}_{ps}) \right|^2 + \sigma_s^2 \leq \alpha, \\ & p_p \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{pp} + \Delta \mathbf{h}_{pp}) \right|^2 \geq \gamma_{th}\beta, \\ & p_s \left| \bar{\mathbf{v}}^H (\hat{\mathbf{h}}_{sp} + \Delta \mathbf{h}_{sp}) \right|^2 + \sigma_p^2 \leq \beta, \\ & \left\| \Delta \mathbf{h}_{ij} \right\|^2 \leq \xi, i, j \in \{p, s\}, \left| \bar{v}_n \right| = 1, n = 1, \dots, N + 1. \end{aligned} \quad (16)$$

Then, by exploiting S-procedure, the constraints involved with CSI errors can be re-expressed by a variety of linear matrix inequalities. Further, by leveraging the technique of semi-definite relaxation (SDR), the optimization problem (P2) can be solved.

### 4.2 With Discrete Phase Shifts

To focus on the discrete passive beamforming design, here we assume perfect CSI for brevity. Denoting the number of tunable IRS phase shift by  $Q$ , the set of discrete reflection coefficients for each IRS element can be written as  $v_n \in \{0, e^{j\Delta\theta}, \dots, e^{j(Q-1)\Delta\theta}\}$ ,  $n = 1, \dots, N$ . We aim to maximize the achievable SU rate via jointly optimizing  $p_s$  and  $\bar{\mathbf{v}}$ , subject to the SINR constraint at the PR. Thus, the optimization problem can be modeled by

$$\begin{aligned} \text{(P3): } \max_{p_s, \bar{\mathbf{v}}} & \frac{p_s \left| h_{ss} + \mathbf{v}^H \mathbf{h}_{srs} \right|^2}{p_p \left| h_{ps} + \mathbf{v}^H \mathbf{h}_{prs} \right|^2 + \sigma_s^2} \\ \text{s.t. } & \frac{p_p \left| h_{pp} + \mathbf{v}^H \mathbf{h}_{ppp} \right|^2}{p_s \left| h_{sp} + \mathbf{v}^H \mathbf{h}_{spp} \right|^2 + \sigma_p^2} \geq \gamma_{th}, \\ & p_s \leq P_{\max}, \\ & v_n \in \{0, e^{j\Delta\theta}, \dots, e^{j(2^Q-1)\Delta\theta}\}, n = 1, \dots, N. \end{aligned} \quad (17)$$

Similar to (P2), the above optimization problem can be solved by the AO based algorithm, and we also focus on optimizing  $\bar{\mathbf{v}}$  for given  $p_s$ . Since an IRS usually consists of a large number of reflecting elements, exhaustive search becomes infeasible. By denoting  $\bar{\mathbf{v}}^H = e^{j\omega} [\mathbf{v}^H, 1]$ ,  $\bar{\mathbf{h}}_{ij}^T = [h_{ij}, h_{ij}]$ , and  $\mathbf{H}_{ij} =$

$p_i \bar{\mathbf{h}}_i \bar{\mathbf{h}}_i^H, i, j \in \{p, s\}$ , (P3) can be rewritten as

$$\begin{aligned}
 \text{(P4): } & \max_{\bar{\mathbf{v}}} t \\
 \text{s.t. } & \bar{\mathbf{v}}^H (\mathbf{H}_{ss} - t \mathbf{H}_{ps}) \bar{\mathbf{v}} \geq t \sigma_s^2, \\
 & \bar{\mathbf{v}}^H (\mathbf{H}_{pp} - t \mathbf{H}_{sp}) \bar{\mathbf{v}} \geq \gamma_{th} \sigma_p^2, \\
 & v_n \in \{0, e^{j\Delta\theta}, \dots, e^{j(2^Q-1)\Delta\theta}\}, n = 1, \dots, N. \quad (18)
 \end{aligned}$$

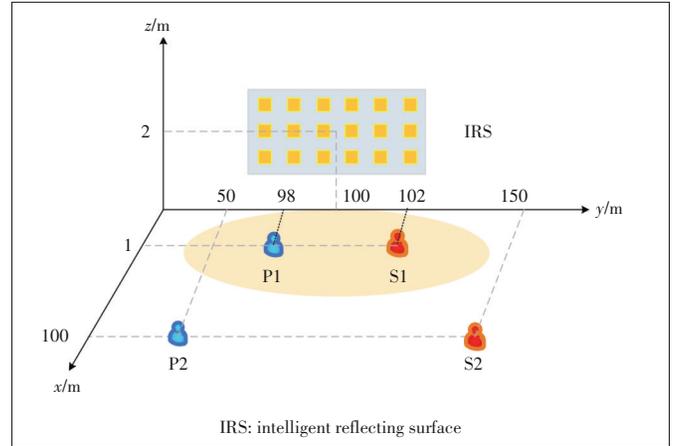
This is an integer nonlinear program, which is still difficult to solve. Then, by denoting  $\mathbf{G}_s = \mathbf{H}_{ss} - t \mathbf{H}_{ps}$ ,  $\mathbf{G}_p = \mathbf{H}_{pp} - t \mathbf{H}_{sp}$ ,  $f_s = \bar{\mathbf{v}}^H \mathbf{G}_s \bar{\mathbf{v}} - |h_{ss}|^2 + t |h_{ps}|^2$ , and  $f_p = \bar{\mathbf{v}}^H \mathbf{G}_p \bar{\mathbf{v}} - |h_{pp}|^2 + t |h_{sp}|^2$ , we have  $f_s = \sum_{i=1}^{N-1} \sum_{j=i+1}^N 2 [\text{Re}(\mathbf{G}_s(i, j)) \cos(\theta_j - \theta_i) - \text{Im}(\mathbf{G}_s(i, j)) \sin(\theta_j - \theta_i)] + \sum_{i=1}^N \mathbf{G}_s(i, i) + 2 \sum_{i=1}^N |\mathbf{G}_s(i, N+1)| [\cos(\angle \mathbf{G}_s(i, N+1)) \cos \theta_i - \sin(\angle \mathbf{G}_s(i, N+1)) \sin \theta_i]$  and  $f_p = \sum_{i=1}^{N-1} \sum_{j=i+1}^N 2 [\text{Re}(\mathbf{G}_p(i, j)) \cos(\theta_j - \theta_i) - \text{Im}(\mathbf{G}_p(i, j)) \sin(\theta_j - \theta_i)] + \sum_{i=1}^N \mathbf{G}_p(i, i) + 2 \sum_{i=1}^N |\mathbf{G}_p(i, N+1)| [\cos(\angle \mathbf{G}_p(i, N+1)) \cos \theta_i - \sin(\angle \mathbf{G}_p(i, N+1)) \sin \theta_i]$ . As such, the integer quadratic constraints in Eq. (17) can be transformed into integer linear constraints, i.e., (P4) can be re-expressed into an integer linear program (ILP), for which the globally optimal solution can be obtained by applying the branch-and-bound method.

## 5 Numerical Results

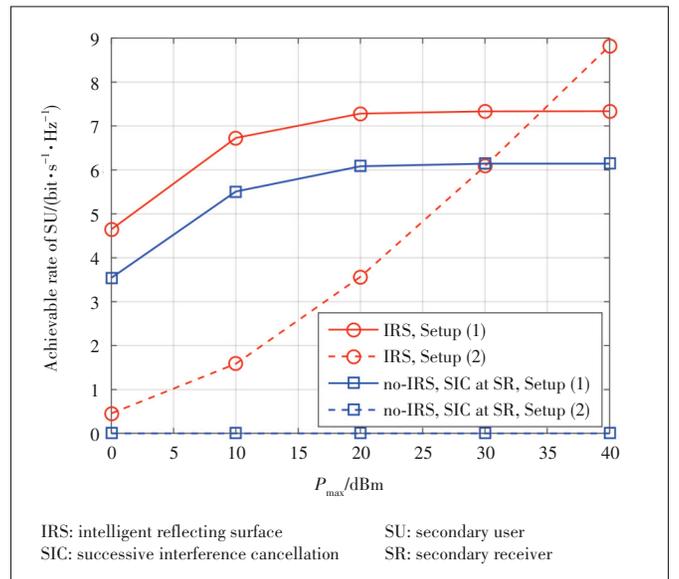
As shown in Fig. 6, to study the effectiveness of deploying IRS in a spectrum sharing system, we consider two different setups with strong cross link interference. Specifically, in Setup 1, P1 and P2 work as PT and PR, and S1 and S2 work as SR and ST (thus incurring strong PT-SR interference), while in Setup 2, P1 and P2 work as PR and PT, and S1 and S2 work as ST and SR (thus incurring strong ST-PR interference). We assume that the system operates on a carrier frequency of 750 MHz with the wavelength  $\lambda_c = 0.4$  m and the path loss at the reference distance  $d_0 = 1$  m is given by  $L_0 = -30$  dB. Suppose that the IRS is equipped with a uniform planar array with 6 rows and 10 columns, and the element spacing is  $\lambda_d = 3\lambda_c/8$ . The noise power is set as  $-105$  dBm. The channels between PT/ST and IRS/PR/SR are assumed to be Rayleigh fading with the path loss exponents set as 3, whereas the channels between IRS and PR in Setup 1 and SR in Setup 2 are line of sight (LoS) with the path loss exponents set as 2. Besides, the case without IRS but with successive interference cancellation (SIC) at the SR for mitigating the interference from PT, i.e., (no-IRS, SIC at SR), is also considered for performance comparison and showing the benefit of using IRS.

Fig. 7 shows the achievable SU rate versus the maximum transmit power of the ST in the considered two setups. It can be observed that the case with IRS always achieves a higher

SU rate than that of the (no-IRS, SIC at SR) design. Specifically, in Setup 1, using IRS not only cancels the interference from the PT(P1) to the SR(S1), but also enhances the desired signal from the ST(S2) to the SR(S1); while applying SIC at the SR can only achieve the former. Moreover, the SU rate for the case with IRS eventually saturates. The reason is that both the ST(S2) and the PR(P2) are not in the coverage of IRS and thus the interference from the ST(S2) to the PR(P2) cannot be reduced, which then constraints the transmitting power at the ST(S2). In Setup 2, the no-IRS design becomes ineffective, though SIC is applied at the SR(S2). This is because given the severe interference from the ST(S1) to the PR(P1), the former should keep silent to guarantee the reception of the latter. However, by applying the IRS, the above ST-PR interference can be efficiently reduced, thus the SU can access the spectrum and achieve a higher rate with increased transmit power.



▲ Figure 6. Two simulation setups with strong cross link interference



▲ Figure 7. Achievable SU rate versus maximum transmitting power at secondary transmitter (ST)

## 6 Conclusions

IRS has been considered as a promising technology to achieve highly spectral and energy efficient wireless communication systems. In this paper, we investigate the IRS-enabled spectrum sharing system and highlight several important issues including the modeling of interference, estimation of channels and design of robust passive beamforming. Our simulations show the effectiveness of employing IRS to improve the SU rate and its advantages in dealing with highly challenging interference scenarios in conventional spectrum sharing systems without the IRS. We hope that this paper would provide a useful guidance for future research into this emerging and promising area.

## References

- [1] 6G Flagship. Key drivers and research challenges for 6G ubiquitous wireless intelligence [EB/OL]. (2019-09-09) [2021-10-14]. <http://jultika.oulu.fi/Record/isbn978-952-62-2354-4>
- [2] WU Q Q, ZHANG R. Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network [J]. *IEEE communications magazine*, 2020, 58(1): 106 – 112. DOI: 10.1109/MCOM.001.1900107
- [3] WU Q Q, ZHANG S W, ZHENG B X, et al. Intelligent reflecting surface aided wireless communications: a tutorial [J]. *IEEE transactions on communications*, 2021, 69(5): 3313–3351. DOI: 10.1109/TCOMM.2021.3051897
- [4] LIU H, ZHANG J Y, WU Q Q, et al. RIS-aided next-generation high-speed train communications: challenges, solutions, and future directions [EB/OL]. (2021-03-17) [2021-10-14]. <https://arxiv.org/abs/2103.09484>
- [5] WU Q, GUAN X, ZHANG R. Intelligent reflecting surface-aided wireless energy and information transmission: an overview [J]. *Proceedings of the IEEE*, 2021, 110(1): 150–170. DOI: 10.1109/JPROC.2021.3121790
- [6] MU X, LIU Y, GUO L, et al. Exploiting intelligent reflecting surfaces in NOMA networks: joint beamforming optimization [J]. *IEEE transactions on wireless communications*, 2020, 19(10): 6884–6898. DOI: 10.1109/JPROC.2021.3121790
- [7] DING Z, POOR H V. A simple design of IRS-NOMA transmission [J]. *IEEE Communications Letters*, 2020, 24(5): 1119–1123. DOI: 10.1109/LCOMM.2020.2974196
- [8] WU Q Q, ZHANG R. Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints [J]. *IEEE journal on selected areas in communications*, 2020, 38(8): 1735 – 1748. DOI: 10.1109/JSAC.2020.3000807
- [9] PAN C H, REN H, WANG K Z, et al. Intelligent reflecting surface enhanced MIMO broadcasting for simultaneous wireless information and power transfer [J]. *IEEE journal on selected areas in communications*, 2020, 38(8): 1719 – 1734. DOI: 10.1109/JSAC.2020.3000802
- [10] CUI M, ZHANG G C, ZHANG R. Secure wireless communication via intelligent reflecting surface [J]. *IEEE wireless communications letters*, 2019, 8(5): 1410–1414. DOI: 10.1109/LWC.2019.2919685
- [11] SHEN H, XU W, GONG S L, et al. Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications [J]. *IEEE communications letters*, 2019, 23(9): 1488–1492. DOI: 10.1109/LCOMM.2019.2924214
- [12] ZHANG J Y, DU H Y, SUN Q, et al. Physical layer security enhancement with reconfigurable intelligent surface-aided networks [J]. *IEEE transactions on information forensics and security*, 2021, 16: 3480 – 3495. DOI: 10.1109/TIFS.2021.3083409
- [13] JIN Y, ZHANG J, ZHANG X, et al. Channel estimation for semi-passive reconfigurable intelligent surfaces with enhanced deep residual networks [J]. *IEEE transactions on vehicular technology*, 2021, 70(10): 11083 – 11088. DOI: 10.1109/TVT.2021.3109937
- [14] JENSEN T L, DE CARVALHO E. An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator [C]// *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020: 5000–5004. DOI: 10.1109/ICASSP40776.2020.9053695
- [15] ZHENG B, ZHANG R. Intelligent reflecting surface enhanced OFDM: channel estimation and reflection optimization [J]. *IEEE wireless communications letters*, 2020, 9(4): 518–522. DOI: 10.1109/LWC.2019.2961357
- [16] LIU H, ZHANG J, WU Q, et al. ADMM based channel estimation for RISs aided millimeter wave communications [J]. *IEEE communications letters*, 2021, 25(9): 2894–2898. DOI: 10.1109/LCOMM.2021.3095218
- [17] GUAN X R, WU Q Q, ZHANG R. Anchor-assisted channel estimation for intelligent reflecting surface aided multiuser communication [J]. *IEEE transactions on wireless communications*, 2021, early access. DOI: 10.1109/TWC.2021.3123674
- [18] ZHANG J, ZHANG Y, ZHONG C, et al. Robust design for intelligent reflecting surfaces assisted MISO systems [J]. *IEEE communications letters*, 2020, 24(10): 2353–2357. DOI: 10.1109/LCOMM.2020.3002557

## Biographies

**GUAN Xinrong** received both the B.Eng. degree in communications engineering and Ph.D. degree in communications and information systems from the College of Communications Engineering, PLA University of Science and Technology, China in 2009 and 2014, respectively. From 2014, he worked as a lecturer at the College of Communications Engineering, Army Engineering University of PLA, China. His current research interests include physical layer security, wireless key generation, and intelligent reflecting surfaces. He has coauthored more than 40 IEEE papers with one ESI highly cited paper. He has served as a reviewer of several IEEE journals.

**WU Qingqing** (qingqingwu@um.edu.mo) received the B.Eng. and Ph.D. degrees in electronic engineering from South China University of Technology and Shanghai Jiao Tong University (SJTU), China in 2012 and 2016, respectively. He is currently an assistant professor with the State Key Laboratory of Internet of Things for Smart City, University of Macau, China. From 2016 to 2020, he was a research fellow in the Department of Electrical and Computer Engineering, National University of Singapore. His current research interests include intelligent reflecting surfaces, unmanned aerial vehicle (UAV) communications, and MIMO transceiver design. He has coauthored more than 100 IEEE papers with 23 ESI highly cited papers and eight ESI hot papers, which have received more than 9 000 Google citations. He was listed as a World's Top 2% Scientist by Stanford University in 2020 and a Clarivate ESI Highly Cited Researcher in 2021.

# Resource Allocation for Two-Tier RIS-Assisted Heterogeneous NOMA Networks



XU Yongjun<sup>1</sup>, YANG Zhaohui<sup>2</sup>, HUANG Chongwen<sup>3</sup>,  
YUEN Chau<sup>4</sup>, GUI Guan<sup>5</sup>

(1. School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. Department of Electronic and Electrical Engineering, University College London, London WC1E 6BT, UK;

3. College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China;

4. Engineering Product Development Pillar, Singapore University of Technology and Design, Singapore 487372, Singapore;

5. College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

DOI: 10.12142/ZTECOM.202201006

<https://kns.cnki.net/kcms/detail/34.1294.TN.34.1294>

TN.20220311.1430.002.html, published online March 14, 2022

Manuscript received: 2022-01-04

**Abstract:** Reconfigurable intelligent surface (RIS) as a promising technology has been proposed to change weak communication environments. However, most of the current resource allocation (RA) schemes have focused on RIS-assisted homogeneous networks, and there is still no open works about RA schemes of RIS-assisted heterogeneous networks (HetNets). In this paper, we design an RA scheme for a RIS-assisted HetNet with non-orthogonal multiple access to improve spectrum efficiency and transmission rates. In particular, we jointly optimize the transmit power of the small-cell base station and the phase-shift matrix of the RIS to maximize the sum rates of all small-cell users, subject to the unit modulus constraint, the minimum signal-to-interference-plus-noise ratio constraint, and the cross-tier interference constraint for protecting communication quality of microcell users. An efficient suboptimal RA scheme is proposed based on the alternating iteration approach, and successive convex approximation and logarithmic transformation approach. Simulation results verify the effectiveness of the proposed scheme in terms of data rates.

**Keywords:** heterogeneous networks; non-orthogonal multiple access; reconfigurable intelligent surface; resource allocation

**Citation** (IEEE Format): Y. J. Xu, Z. H. Yang, C. W. Huang, et al., "Resource allocation for two-tier RIS-assisted heterogeneous NOMA networks," *ZTE Communications*, vol. 20, no. 1, pp. 36 - 47, Mar. 2022. doi: 10.12142/ZTECOM.202201006.

## 1 Introduction

Recently, reconfigurable intelligent surface (RIS) as a promising technology has been proposed to improve energy efficiency (EE) and transmission quality by reconfiguring the wireless propagation environment<sup>[1-4]</sup>. Particularly, a RIS has a large number of low-cost passive reflecting elements which can reflect the incident signals from the transmitters by changing the phase shift (PS) of the RIS in a passive way. Since the RIS is connected with a base station (BS) by a RIS controller, the reflected signal can be smartly configured to strengthen the desired signal and suppress the undesired signal for meeting the transmission requirements of communication systems.

The work was partially supported by the China National Key R&D Program under Grant No. 2021YFA1000502, National Natural Science Foundation of China under Grant No. 62101492, Zhejiang Provincial Natural Science Foundation of China under Grant No. LR22F010002, Distinguished Young Scholars of the National Natural Science Foundation of China, Ng Teng Fong Charitable Foundation in the form of ZJU-SUTD IDEA Grant, Zhejiang University Education Foundation Qizhen Scholar Foundation, and Fundamental Research Funds for the Central Universities under Grant No. 2021FZZX001-21.

### 1.1 Related Work

According to the above advantages of RIS, resource allocation (RA) problems for RIS-assisted networks have been concerned by many scholars. For example, the authors in Ref. [5] proposed a low-complexity iteration algorithm to maximize the spectrum efficiency (SE) by jointly optimizing the beamformer at the BS and the PS of the RIS in a point-to-point RIS-assisted multiple-input single-output (MISO) communication system. But only one user is considered. To overcome the effect of imperfect channel state information (CSI), a weighted sum-rate maximization problem for the multiuser scenario was investigated by jointly designing the active beamforming of the access point (AP) and the passive beamforming of the RIS under perfect/imperfect CSI in Ref. [6], where stochastic successive convex approximation (SCA) was used to obtain the solution. But the minimum signal-to-interference plus-noise ratio (SINR) constraint is not considered, and it fails to meet the required quality of service (QoS) of each user. Considering the same communication network as in Ref. [6], the transmit power and the PS of the RIS were jointly designed in Ref. [7] to maximize the total EE by using an alternat-

ing iteration approach and the sequential fractional programming. Moreover, in Ref. [8], the total transmit power of the AP was minimized by jointly optimizing the transmit beamforming of the AP and the reflect beamforming of the RIS, subject to the SINR of each user. In Ref. [9], a particle swarm optimization method was used to minimize the total transmit power. However, the above works<sup>[5-9]</sup> consider the RIS-assisted MISO system. For a RIS-assisted multiple-input multiple-output (MIMO) network<sup>[10]</sup>, a weighted sum-rate maximization RA problem was studied by jointly optimizing the precoding matrices of the BS and the PS of the RIS, subject to the maximum power constraint of each BS and the unit modulus constraint of the RIS, where a RIS at the cellular edge was developed to assist the downlink transmission for cell-edge users. The classical block coordinate descent (BCD) algorithm, the Majorization-Minimization algorithm and the complex circle manifold method were used to deal with the non-convex problem. The authors in Ref. [11] proposed a local search method to maximize the sum rate in a RIS-assisted Terahertz communication system. For a RIS-assisted simultaneous wireless information and power transfer (SWIPT) system<sup>[12]</sup>, the authors studied the RA problem of the weighted sum power maximization of multiple energy-harvesting receivers. In Ref. [13], a theoretical outage performance comparison between non-orthogonal multiple access (NOMA) and orthogonal multiple access (OMA) in the RIS-assisted downlink communication was analyzed. Moreover, the power minimization problem and performance analysis for RIS-aided NOMA systems have been studied in Ref. [14] and Refs. [15 - 17], respectively. The performance comparisons between the new RIS system and the traditional decode-and-forward relaying system were studied in terms of total transmit power minimization and the EE maximization<sup>[18]</sup>. Moreover, other related works on the RIS-assisted scenario have been studied from space shift keying<sup>[19]</sup>, channel estimation<sup>[20]</sup>, etc.

## 1.2 Motivation and Contributions

However, the above works<sup>[6-20]</sup> have not considered the RA problem in heterogeneous networks (HetNets). HetNet as an important technique in 6G communications<sup>[21]</sup> can obviously improve communication coverage and network capacity by deploying small cells into the macrocell<sup>[22-23]</sup>. However, the QoS of indoor users is heavily degraded by interior walls and strong fading channels. Thus, it is a natural idea to combine RIS and HetNet to exemplify received signals and expand communication coverage. Besides, the introduction of NOMA can improve SE and allow more accessing users. Different from the RA problem in traditional NOMA systems, our work needs to simultaneously optimize transmit power and phase shifts to further improve SE and signal received quality at the receivers.

In this paper, we study the RA problem for a downlink RIS-assisted heterogeneous NOMA network, where a small-cell BS (SBS) transmits wireless signals to the targeted users whose signals are enhanced with the help of the RIS. The RA problem for such a system is still in its infancy. The contributions

of this paper are summarized as follows.

- A two-tier RIS-assisted heterogeneous NOMA network is formulated, where a single-antenna SBS transmits wireless signals to multiple single-antenna small-cell users (SCUs) with the assistance of the RIS. For the single-user scenario, the rate maximization RA problem is formulated by jointly optimizing the transmit power of the SBS and the PS of the RIS under the SINR constraint of the macro-cell user (MCU) and the maximum transmit power constraint. A Dinkelbach-based iteration scheme is proposed to globally obtain the optimal solutions.

- For the multiuser scenario, the total rate maximization problem is formulated under the cross-tier interference constraint, the maximum transmit power, and the unit modulus constraint of the RIS. To deal with the non-convex problem, we propose an efficient RA scheme based on a semi-definite programming (SDP) approach and SCA to obtain the suboptimal solution by an alternating iteration manner.

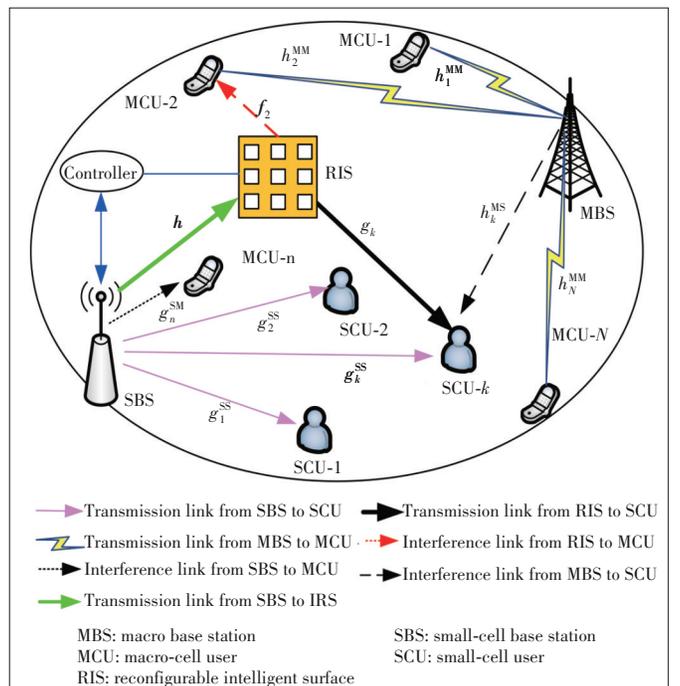
- Simulation results show that the proposed scheme has good convergence and transmission rates.

The rest of this paper is organized as follows. In Section 2, the system model and the single-user RA problem are introduced. Section 3 proposes a multi-user RA scheme for the multiuser scenario. Section 4 verifies the effectiveness of the proposed scheme. Section 5 concludes this paper.

## 2 System Model and Single-User RA Problem

### 2.1 System Model

A downlink two-tier RIS-assisted HetNet with NOMA is given in Fig. 1, where a RIS with  $M$  passive reflecting units is



▲ Figure 1. A downlink RIS-assisted heterogeneous non-orthogonal multiple access (NOMA) network

employed to enhance the transmission quality between the SBS and  $K$  NOMA SCUs. A macro base station (MBS) serves  $N$  MCUs. The SCUs can share the spectrum owned by MCUs via an underlying spectrum way. Since the RIS is developed in interior walls, where the reflected signals from the RIS caused by MCUs can be ignored. A separate wireless control link serves for information exchange between the RIS controller and the SBS, and other required information for implementing the transmit power<sup>[10]</sup>. Motivated by the works<sup>[5-7]</sup>, perfect CSI and continuous phase-shift coefficients are assumed here. For the single-user scenario, defining  $\mathbf{h} = [h_1, \dots, h_M]^H \in \mathbb{C}^{M \times 1}$ ,  $\mathbf{g} = [g_1, \dots, g_M]^H \in \mathbb{C}^{M \times 1}$ ,  $\mathbf{f} = [f_1, \dots, f_M]^H \in \mathbb{C}^{M \times 1}$  and  $\boldsymbol{\phi} = [\phi_1, \dots, \phi_M]^H$  as the channel gain from the SBS to the RIS, the channel gain from the RIS to the SCU, the channel gain from the RIS to the MCU, and the reflecting coefficient (RC) of the RIS, where each  $\phi_m = \beta_m e^{j\theta_m}$  comprises an amplitude coefficient  $\beta_m \in [0, 1]$ . Similar to Refs. [5-9], a continuous PS coefficient is assumed, e. g.,  $\theta_m \in [0, 2\pi]$  and  $|\phi_m| = 1$ .  $g^{\text{SM}}$  and  $g^{\text{SS}}$  are the channel gains from the SBS to the MCU and the SBS to the SCU. Define  $\boldsymbol{\Phi} = \text{diag}(\boldsymbol{\phi}) \in \mathbb{C}^{M \times M}$  as the RC matrix of the RIS. Since we aim to obtain the maximum designed signal, the amplitude coefficient is set as  $\beta_m = 1$  for simplicity<sup>[24]</sup>. Ignoring signals reflected by the RIS for two and more times<sup>[25]</sup>, the received SINRs at the SCU and the MCU are

$$\gamma^{\text{SCU}} = \frac{p |\mathbf{g}^H \boldsymbol{\Phi} \mathbf{h} + g^{\text{SS}}|^2}{Ph^{\text{MS}} + \sigma^2}, \quad (1)$$

$$\gamma^{\text{MCU}} = \frac{p |\mathbf{g}^H \boldsymbol{\Phi} \mathbf{h} + g^{\text{SS}}|^2}{Ph^{\text{MS}} + \sigma^2}, \quad (2)$$

where  $\sigma^2$  represents the noise power at the receiver.

## 2.2 Single-User RA Scheme

Motivated by the existing works in Refs. [5-10], the power allocation and PS ratio under the single-user scenario are jointly optimized by using an alternating iteration approach. However, the solution is just a suboptimal solution. In order to deal with this challenge, we try to get the optimal power allocation and PS solution under a single-user scenario.

Under this special case, we want to improve the transmission rate of the SCU by adjusting the transmit power and PS coefficients in a globally optimal solution way, while the QoS of the MCU is guaranteed. Therefore, the rate-maximization problem of the SCU is formulated as

$$\begin{aligned} & \max_{p, \phi_m} \log_2(1 + \gamma^{\text{SCU}}) \\ & \text{s.t. } C_1: \gamma^{\text{MCU}} \geq \gamma^{\text{min}}, \\ & C_2: |\phi_m| = 1, \forall m, \\ & C_3: 0 \leq p \leq p^{\text{max}}, \end{aligned} \quad (3)$$

where  $C_1$  denotes the QoS required by the MCU under the consideration of the interference power from the RIS and the SBS,  $\gamma^{\text{min}}$  represents the SINR threshold of the MCU.  $C_2$  accounts for the fact that each RIS reflecting element can only provide a phase shift, without amplifying the incoming signal.  $C_3$  denotes the transmit power range of the SBS, and  $p^{\text{max}}$  is the maximum transmit power at the SBS. Since we want to find the optimal  $p^*$  and  $\boldsymbol{\phi}^*$ , the maximum transmit power of the MBS is ignored.

Based on  $C_1$ , the upper bound of  $p$  is

$$p \leq \frac{Ph^{MM} - \gamma^{\text{min}} \sigma^2}{\gamma^{\text{min}} |g^{\text{SM}} + \mathbf{f}^H \boldsymbol{\Phi} \mathbf{h}|^2}. \quad (4)$$

Based on the monotonicity of  $p$  in the objective function in Problem (3), the optimal transmit power becomes

$$p^* = \min \left( p^{\text{max}}, \frac{Ph^{MM} - \gamma^{\text{min}} \sigma^2}{\gamma^{\text{min}} |g^{\text{SM}} + \mathbf{f}^H \boldsymbol{\Phi} \mathbf{h}|^2} \right) \quad (5)$$

To obtain the globally optimal solution, we substitute  $p^*$  into Problem (3), the PS coefficient optimization problem of Problem (3) can be reformulated as

$$\begin{aligned} & \max_{\phi_m} \log_2 \left( 1 + \frac{A |\mathbf{g}^H \boldsymbol{\Phi} \mathbf{h} + g^{\text{SS}}|^2}{|g^{\text{SM}} + \mathbf{f}^H \boldsymbol{\Phi} \mathbf{h}|^2} \right) \\ & \text{s.t. } C_2: |\phi_m| = 1, \forall m, \end{aligned} \quad (6)$$

where  $A = \frac{Ph^{MM} - \gamma^{\text{min}} \sigma^2}{\gamma^{\text{min}} |g^{\text{SM}} + \mathbf{f}^H \boldsymbol{\Phi} \mathbf{h}|^2}$  if  $p^* < p^{\text{max}}$ , otherwise  $A = p^{\text{max}}$ .

Since the data rate of Problem (6) is in proportion to its SINR, we have  $\max_{\phi_m} \log_2 \left\{ 1 + \text{SINR}(\phi_m) \right\} \Leftrightarrow \max_{\phi_m} \text{SINR}(\phi_m)$ . As a result, we can remove the logarithmic operation, Problem (6) becomes a nonlinear fractional programming problem<sup>[26]</sup>.

Based on Dinkelbach's method<sup>[26]</sup>, we have

$$\begin{aligned} & \max_{\phi_m} A |\mathbf{g}^H \boldsymbol{\Phi} \mathbf{h} + g^{\text{SS}}|^2 - \lambda |g^{\text{SM}} + \mathbf{f}^H \boldsymbol{\Phi} \mathbf{h}|^2 \\ & \text{s.t. } C_2: |\phi_m| = 1, \forall m. \end{aligned} \quad (7)$$

where  $\lambda$  is an auxiliary variable. Obviously, the objective function of Problem (7) is a strictly continuous and decreasing function with  $\lambda$ .

Theorem 1: For any  $\boldsymbol{\Phi}$  and  $\lambda$ ,  $Q(\lambda) = \max_{\phi_m} A |\mathbf{g}^H \boldsymbol{\Phi} \mathbf{h} + g^{\text{SS}}|^2 - \lambda |g^{\text{SM}} + \mathbf{f}^H \boldsymbol{\Phi} \mathbf{h}|^2$  is a decreasing function with  $\lambda$ , and  $Q(\lambda) \geq 0$ .

Proof: See Appendix A.

Since  $\boldsymbol{\phi} = [\phi_1, \dots, \phi_M]^H$ ,  $\bar{\mathbf{g}} = \text{diag}(\mathbf{g})\mathbf{h}$  and  $\bar{\mathbf{f}} = \text{diag}(\mathbf{f})\mathbf{h}$ , Problem (7) becomes

$$\begin{aligned}
& \max_{\phi_m} \phi^H (A\bar{g}\bar{g}^H - \lambda\bar{f}\bar{f}^H)\phi + A|g^{SS}|^2 - \lambda|g^{SM}|^2 + \\
& \text{Re}\{(Ag^{SS}\bar{g}^H - \lambda g^{SM}\bar{f}^H)\phi\} + \text{Re}\{\phi^H (Ag^{SS}\bar{g} - \lambda g^{SM}\bar{f})\} \\
& \text{s.t. } C_2: |\phi_m| = 1, \forall m,
\end{aligned} \tag{8}$$

where  $\text{Re}\{\cdot\}$  denotes the real part of a complex number.

Define  $\tilde{g} = A\bar{g}\bar{g}^H - \lambda\bar{f}\bar{f}^H$ ,  $\bar{h} = Ag^{SS}\bar{g} - \lambda g^{SM}\bar{f}$ , and  $\bar{A} = A|g^{SS}|^2 - \lambda|g^{SM}|^2$ , Problem (8) becomes

$$\begin{aligned}
& \max_{\phi_m} \phi^H \tilde{g}\phi + \text{Re}\{\phi^H \bar{h}\} + \text{Re}\{\bar{h}^H \phi\} + \bar{A} \\
& \text{s.t. } C_2: |\phi_m| = 1, \forall m.
\end{aligned} \tag{9}$$

Problem (9) is a non-convex quadratically constrained quadratic program (QCQP) problem<sup>[27]</sup>, which can be converted into a homogeneous QCQP problem. By introducing an auxiliary variable  $t$  ( $t \geq 0$ ), Problem (9) becomes

$$\begin{aligned}
& \max_{\phi_m} \bar{\phi}^H M \bar{\phi} + \bar{A} \\
& \text{s.t. } \tilde{C}_2: |\bar{\phi}_m| = 1, \forall m,
\end{aligned} \tag{10}$$

where

$$M = \begin{bmatrix} \tilde{g} & \bar{h} \\ \bar{h}^H & 0 \end{bmatrix} \text{ and } \bar{\phi} = \begin{bmatrix} \phi \\ t \end{bmatrix}, \tag{11}$$

and  $\bar{\phi} = [\bar{\phi}_1, \dots, \bar{\phi}_M] \in \mathbb{C}^{(M+1) \times 1}$ . But problem (9) is still non-convex. Note that  $\bar{\phi}^H M \bar{\phi} = \text{Trace}(M \bar{\phi} \bar{\phi}^H)$ . Define  $X = \bar{\phi} \bar{\phi}^H$ , which satisfies  $X \geq 0$  and  $\text{Rank}(X) = 1$ . Thus, Problem (9) becomes

$$\begin{aligned}
& \max_X \text{Trace}(MX) + \bar{A} \\
& \text{s.t. } \tilde{C}_3: X_m = 1, \forall m, \\
& C_4: X \geq 0, \text{Rank}(X) = 1.
\end{aligned} \tag{12}$$

By relaxing the rank-one constraint, Problem (12) is a convex SDP problem<sup>[26]</sup>, which can be efficiently solved by using the convex optimization tool, e. g., SeDuMi<sup>[28]</sup>. Generally, the solution  $X$  of Problem (12) does not satisfy its rank constraint<sup>[28]</sup>, namely,  $\text{Rank}(X) \neq 1$ , while the objective function of Problem (12) only serves an upper bound of it. To obtain a rank-one solution, the Gaussian randomization scheme can be used<sup>[29]</sup>. As a result, an iteration-based RA scheme is summarized in Algorithm 1. The convergence of Algorithm 1 is analyzed in Appendix B.

Moreover, the computational complexity of Algorithm 1 is analyzed as follows. The computational complexity is mainly decided by Dinkelbach's method. When the convergence precision  $\varepsilon$  and the maximum iteration number  $L_{\max}$  are determined, the computational complexity of the Dinkelbach-based

iterative algorithm is  $\mathcal{O}\left(\frac{\log(L_{\max})}{\varepsilon^2}\right)$ <sup>[30]</sup>. Besides, during each iteration, the complexity for obtaining  $\phi_m$  by using SeDuMi tool is  $\mathcal{O}\left(\sqrt{M+M^2} L \log\left(\frac{1}{l}\right)\right)$ , where  $L = M^5 + 3M^6 + M^7$ .  $l$  denotes the convergence precision of SeDuMi tool. As a result, the total complexity of Algorithm 1 is  $\mathcal{O}\left(\frac{\sqrt{M+M^2} L}{\varepsilon^2} \log\left(\frac{1}{l}\right) \log(L_{\max})\right)$ .

---

#### Algorithm 1. A Dinkelbach-based RA Scheme

---

- 1: Initialize the maximum number of iterations  $L_{\max}$ , and the maximum tolerance  $\varepsilon$ ;
  - 2: Set the auxiliary variable  $\lambda = 0$  and the iteration index  $l = 0$ ;
  - 3: **repeat** {Main loop}
  - 4: Solve the inner loop problem in Eq. (10) for a given  $\lambda$ ;
  - 5: Construct  $X$  by the Gaussian randomization scheme<sup>[13]</sup> to satisfy the rank condition; Obtain  $\phi_m$  by using SeDuMi tool<sup>[41]</sup>;
  - 6: **if**  $A|g^H \Phi h + g^{SS}|^2 - \lambda|f^H \Phi h + g^{SM}|^2 < \varepsilon$  **then**
  - 7:     Convergence = **true**
  - 8:     **return**  $\phi_m^*, \forall m$ , obtain  $p^*$  and  $\lambda^* = \frac{A|g^H \Phi h + g^{SS}|^2}{|f^H \Phi h + g^{SM}|^2}$ ;
  - 9: **else**
  - 10:     Set  $\lambda(l+1) = \frac{A|g^H \Phi(l)h + g^{SS}|^2}{|f^H \Phi(l)h + g^{SM}|^2}$  and  $l = l + 1$ ;
  - 11:     Convergence = **true**
  - 12: **end if**
  - 13: **until** Convergence = true or  $l = L_{\max}$
- 

### 3 Multi-User RA Scheme

Since the single-user scenario is a specific case for practical systems, it is only helpful to analyze system performance, which is too ideal for practical two-tier RIS-aided communication systems. In order to improve system capacity and support massive connectivity, in this section, we extend the special case of Problem (3) to the multi-user NOMA scenario. Assuming that there are  $K$  NOMA SCUs and  $N$  MCUs, the sets are denoted as  $\mathcal{K} = \{1, 2, \dots, K\}, \forall k \in \mathcal{K}$  and  $\mathcal{N} = \{1, 2, \dots, N\}, \forall n \in \mathcal{N}$ . The signal of the SBS is  $x = \sum_{k=1}^K \sqrt{p_k} x_k$ , where  $p_k$  and  $x_k$  denote the allocated power and the signal from the SBS to the  $k$ -th SCU, respectively. Therefore, the signal received at the  $k$ -th SCU is

$$y_k = g_k^{SS} x + g_k^H \Phi h x + h_k^{MS} \sqrt{P} s_n + n_k, \tag{13}$$

where  $g_k^{SS}$  denotes the channel response from the SBS to SCU  $k$ ,  $g_k$  denotes the channel vector from the RIS to SCU  $k$ ;  $h_k^{MS}$  denotes the channel coefficient from the MBS to the  $k$ -th SCU;  $P$  and  $s_n$  are the transmit power and signal from the MBS to the

$n$ -th MCU, respectively.  $n_k \sim \mathcal{CN}(0, \sigma_k^2)$  is the additive white Gaussian noise (AWGN) at the  $k$ -th SCU's receiver with zero mean and variance  $\sigma_k^2$ . The transmission signals of the SBS and the MBS satisfy  $E\left[|x_k|^2\right] = 1$  and  $E\left[|s_n|^2\right] = 1$ .

Assume that the channel coefficients are sorted in the descending order  $|g_1^{\text{SS}} + \mathbf{g}_1^H \mathbf{h}|^2 \geq |g_2^{\text{SS}} + \mathbf{g}_2^H \mathbf{h}|^2 \geq \dots \geq |g_K^{\text{SS}} + \mathbf{g}_K^H \mathbf{h}|^2$ . For any  $i < k$ , the SINR of SCU  $k$  is

$$\gamma_k^{\text{SCU}} = \frac{p_k |g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2}{|g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2 \sum_{i=1}^{k-1} p_i + |h_k^{\text{MS}}|^2 P + \sigma_k^2}, \quad (14)$$

where the first item in the denominator denotes the interference power from other SCUs, and the second item denotes the cross-tier interference from the MBS to the SCU  $k$ . Since the MBS is often far from SCUs, the interference power of the macrocell network can be omitted<sup>[31]</sup>.

Thus, the total sum rate for all SCUs is

$$R = \sum_{k=1}^K \log_2(1 + \gamma_k^{\text{SCU}}). \quad (15)$$

Since we only consider the transmit power optimization problem for SCUs due to the lower priority in spectrum usage, the joint optimization problem of the transmit power of SBS and the transmit power of MBS are beyond the scope of this paper. Thus, we assume that the MBS allocates the same transmit power  $P$  to each MCU. Thus, the received signal at the  $n$ -th MCU can be formulated as

$$y_n = \sqrt{P} h_n^{\text{MM}} s_n + \sum_{k=1}^K \sqrt{p_k} (g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}) x_k + z_n, \quad (16)$$

where  $z_n \sim \mathcal{CN}(0, \sigma_n^2)$  denotes the received AWGN at the  $n$ -th MCU receiver with zero mean and variance  $\sigma_n^2$ .

As a result, the SINR received at the  $n$ -th MCU is

$$\gamma_n^{\text{MCU}} = \frac{P |h_n^{\text{MM}}|^2}{\sum_{k=1}^K p_k |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2 + \sigma_n^2}, \quad (17)$$

where  $h_n^{\text{MM}}$  denotes the channel response from the MBS to the  $n$ -th MCU, and  $g_n^{\text{SM}}$  is the channel response from the SBS to the  $n$ -th MCU. The first item in the denominator denotes the cross-tier interference from the SBS to the  $n$ -th MCU.

Thus, the sum-rate maximization RA problem of the considered RIS-assisted heterogeneous NOMA network becomes

$$\begin{aligned} & \max_{p_k, \phi_m} \sum_{k=1}^K \log_2(1 + \gamma_k^{\text{SCU}}) \\ & \text{s.t. } C_2: |\phi_m| = 1, \forall m, \\ & C_5: \sum_{k=1}^K p_k \leq P^{\text{max}}, \\ & C_6: \gamma_n^{\text{MCU}} \geq \gamma_n^{\text{min}}, \end{aligned} \quad (18)$$

where  $\gamma_n^{\text{min}}$  denotes the minimum SINR required by the  $n$ -th MCU. However, it is difficult to obtain the optimal solution to Problem (18), since the objective function is non-concave with respect to either  $p_k$  or  $\phi_m$ . Moreover,  $C_6$  is non-convex due to the coupled  $p_k$  and  $\phi_m$ . To address this intractable problem, we propose a low-complexity and efficient scheme, which optimizes  $p_k$  and  $\phi_m$  by leveraging an alternating iteration approach and the logarithmic transformation method, respectively. The details are given in the following.

### 3.1 Optimizing $p_k$ with Given $\phi_m$

Based on  $C_6$ , we have

$$I_n^{\text{th}} = \frac{P |h_n^{\text{MM}}|^2}{\gamma_n^{\text{min}}} - \sigma_n^2 \geq \sum_{k=1}^K p_k |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2, \quad (19)$$

where  $I_n^{\text{th}}$  denotes the interference power limit of MCU  $n$ .

As a result, the power allocation subproblem of Problem (18) becomes

$$\begin{aligned} & \max_{p_k} \sum_{k=1}^K \log(1 + \gamma_k^{\text{SCU}}) \\ & \text{s.t. } C_5: \sum_{k=1}^K p_k \leq P^{\text{max}}, \\ & \bar{C}_6: \sum_{k=1}^K p_k |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2 \leq I_n^{\text{th}}. \end{aligned} \quad (20)$$

It is noted that  $C_5$  and  $\bar{C}_6$  are linear constraints, but the objective function is nonconvex.

Based on the successive convex approximation for low complexity (SCALE) algorithm in Ref. [32], for any  $z > 0$ , we have the following bound:

$$\log_2(1 + z) \geq \alpha \log_2 z + \beta, \quad (21)$$

where  $\alpha = \frac{z_0}{1 + z_0}$  and  $\beta = \log_2(1 + z_0) - \alpha \log_2 z_0$  are the auxiliary variables. The relationship is tight at  $z = z_0$ , and  $\alpha$  and  $\beta$  are the approximation constants.

As a result, we have the following sum rate

$$R = \sum_{k=1}^K \alpha_k \log_2 \gamma_k^{\text{SCU}} + \sum_{k=1}^K \beta_k, \quad (22)$$

which is a lower bound of the sum rate achieved by SCUs. Note that the relaxation is still a non-convex form due to the difference of convex (d.c.) structure.

Defining  $\bar{\sigma}_k = |h_k^{\text{MS}}|^2 P + \sigma_k^2$ , with the logarithmic transformation  $\bar{p}_k = \ln p_k, \forall k$ , we have the following convex optimization problem

$$\begin{aligned}
& \max_{\bar{p}_k} \sum_{k=1}^K \alpha_k \log_2(|g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2) - \sum_{k=1}^K \beta_k - \\
& \sum_{k=1}^K \alpha_k \log_2(\bar{\sigma}_k + |g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2 \sum_{i=1}^{k-1} e^{\bar{p}_i}) \\
& \text{s.t. } \bar{C}_5: \sum_{k=1}^K e^{\bar{p}_k} \leq p^{\max}, \\
& \bar{C}_6: \sum_{k=1}^K e^{\bar{p}_k} |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2 \leq I_n^{\text{th}}, \forall n.
\end{aligned} \quad (23)$$

Problem (23) is a standard concave maximization problem since the log-sum-exp is convex<sup>[33]</sup>. Considering the convexity of Problem (23), the closed-form solution of power allocation can be obtained by using the Lagrange dual theory. Thus, the Lagrange function of Problem (23) can be written as

$$\begin{aligned}
L(\bar{p}_k, \mu, \vartheta_n) &= \sum_{k=1}^K \alpha_k \bar{p}_k + \sum_{k=1}^K \alpha_k \log_2(|g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2) - \\
& \sum_{k=1}^K \beta_k - \sum_{k=1}^K \alpha_k \log_2\left(\bar{\sigma}_k + |g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2 \sum_{i=1}^{k-1} e^{\bar{p}_i}\right) + \\
& \mu \left(p^{\max} - \sum_{k=1}^K e^{\bar{p}_k}\right) + \sum_{n=1}^N \vartheta_n \left(I_n^{\text{th}} - \sum_{k=1}^K e^{\bar{p}_k} |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2\right),
\end{aligned} \quad (24)$$

where  $\mu$  and  $\vartheta_n$  are the non-negative Lagrange multipliers.

Based on Karush-Kuhn-Tucker (KKT) conditions, the transmit power is calculated by

$$p_k^* = \left[ \frac{\alpha_k}{\mu + \sum_{n=1}^N \vartheta_n |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2} \right]^+, \quad (25)$$

where  $[x]^+ = \max\{0, x\}$ .

Based on the sub-gradient updating method, Lagrange multipliers can be calculated by

$$\mu(l+1) = \left[ \mu(l) - d_1(l) \times \left(p^{\max} - \sum_{k=1}^K p_k(l)\right) \right]^+, \quad (26)$$

$$\begin{aligned}
\vartheta_n(l+1) &= \left[ \vartheta_n(l) - d_2(l) \times \left(I_n^{\text{th}} - \sum_{k=1}^K p_k(l) |g_n^{\text{SM}} + \right. \right. \\
& \left. \left. f_n^H \Phi \mathbf{h}|^2 \right) \right]^+,
\end{aligned} \quad (27)$$

where  $l$  is the iteration number.  $d_1(l) \geq 0$  and  $d_2(l) \geq 0$  denote the sufficiently small step sizes. When they satisfy the following relationship, we have

$$\sum_{l=1}^{\infty} d_i(l) = \infty, \lim_{l \rightarrow \infty} d_i(l) = 0, \forall i = \{1, 2\}. \quad (28)$$

The sub-gradient updating iteration algorithm can fast converge. The dual variables  $\mu(l+1)$  and  $\vartheta_n(l)$  will converge to the optimal values  $\mu^*$  and  $\vartheta_n^*$ ,  $\forall n$ ,  $\forall n$  as  $l \rightarrow \infty$ . Note that the duality gap for Problem (23) is zero and the solution is unique. As a result, the primal variable  $p_k^*(\mu(l), \vartheta_n(l))$  can reach its targeted value  $p_k^*$ ,  $\forall k$ .

### 3.2 Optimizing $\phi_m$ with Given $p_k^*$

Now, we perform the optimization over  $\phi_m$  with the fixed  $p_k^*$ . Then, Problem (23) is equivalent to

$$\begin{aligned}
& \max_{\phi_m} \sum_{k=1}^K \frac{\alpha_k p_k |g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2}{|g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2 \sum_{i=1}^{k-1} p_i + \bar{\sigma}_k} \\
& \text{s.t. } \bar{C}_2: |\phi_m| = 1, \forall m \\
& \bar{C}_6: \sum_{k=1}^K p_k |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2 \leq I_n^{\text{th}}.
\end{aligned} \quad (29)$$

Problem (29) is the sum of multiple-ratio fractional programming problems, and the non-convexity introduced by the fractional objective function and the non-convex unit-modulus constraint  $\bar{C}_2$ . We adopt a relaxation method to solve Problem (29) efficiently.

Defining  $\bar{\mathbf{F}} = \bar{\mathbf{f}} \bar{\mathbf{f}}^H$ , based on  $|x+y| \leq |x| + |y|$  and  $\bar{\mathbf{f}} = \text{diag}(\mathbf{f}_n) \mathbf{h}$ ,  $\bar{C}_6$  becomes

$$\begin{aligned}
& |g_n^{\text{SM}} + f_n^H \Phi \mathbf{h}|^2 \leq |g_n^{\text{SM}}|^2 + |\mathbf{f}_n^H \bar{\mathbf{f}}|^2 = |g_n^{\text{SM}}|^2 + \mathbf{f}_n^H \bar{\mathbf{F}} \mathbf{f}_n \leq \\
& I_n^{\text{th}} / \sum_{k=1}^K p_k.
\end{aligned} \quad (30)$$

Thus, we have

$$\mathbf{f}_n^H \bar{\mathbf{F}} \mathbf{f}_n \leq \bar{I}_n^{\text{th}}, \quad (31)$$

where  $\bar{I}_n^{\text{th}} = I_n^{\text{th}} / \sum_{k=1}^K p_k - |g_n^{\text{SM}}|^2$  is the equivalent interference level for the  $n$ -th MCU. Thus, Problem (29) becomes

$$\begin{aligned}
& \max_{\phi_m, \gamma_k} \sum_{k=1}^K \gamma_k \\
& \text{s.t. } \bar{C}_2: |\phi_m| = 1, \forall m, \\
& C_7: \frac{\alpha_k p_k |g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2}{|g_k^{\text{SS}} + \mathbf{g}_k^H \Phi \mathbf{h}|^2 \sum_{i=1}^{k-1} p_i + \bar{\sigma}_k} \geq \gamma_k, \forall k, \\
& C_8: \mathbf{f}_n^H \bar{\mathbf{F}} \mathbf{f}_n \leq \bar{I}_n^{\text{th}}, \forall n,
\end{aligned} \quad (32)$$

where  $\gamma_k \geq 0$  is an auxiliary variable.

Defining  $\bar{\mathbf{g}}_k = \text{diag}(\mathbf{g}_k) \mathbf{h}$ ,  $C_7$  can be rewritten as

$$\boldsymbol{\phi}^H \bar{\mathbf{g}}_k \bar{\mathbf{g}}_k^H \boldsymbol{\phi} + 2\text{Re}\{g_k^{SS,*} \bar{\mathbf{g}}_k^H \boldsymbol{\phi}\} + f_k \leq 0, \quad (33)$$

where  $(\cdot)^*$  denotes the conjugate of vector and  $f_k = \left| g_k^{SS} \right|^2 + \frac{y_k \bar{\sigma}_k}{y_k \sum_{i=1}^{k-1} p_i - \alpha_k p_k}$ .

In addition,  $C_2$  can be rewritten as

$$\boldsymbol{\phi}^H \mathbf{U} \boldsymbol{\phi} = 1, \quad (34)$$

where  $\mathbf{U} \in M \times M$  is a symmetric matrix with elements of zeros, except for  $u_{m,m} = 1$ .

Thus, we have

$$\begin{aligned} & \max_{\boldsymbol{\phi}, y_k} \sum_{k=1}^K y_k \\ & \text{s.t. } \bar{C}_4: \boldsymbol{\phi}^H \mathbf{U} \boldsymbol{\phi} = 1, \\ & \bar{C}_7: \bar{f}_k + f_k \leq 0, \\ & \bar{C}_8: \boldsymbol{\phi}^H \tilde{\mathbf{F}} \boldsymbol{\phi} \leq \bar{I}_n^{\text{th}}, \forall n, \end{aligned} \quad (35)$$

where  $\bar{f}_k = \boldsymbol{\phi}^H \bar{\mathbf{g}}_k \bar{\mathbf{g}}_k^H \boldsymbol{\phi} + 2\text{Re}\{g_k^{SS,*} \bar{\mathbf{g}}_k^H \boldsymbol{\phi}\}$ . Problem (35) is a convex optimization problem with the quadratic constraints.

As a result, the Lagrange function of Problem (35) can be written as

$$\begin{aligned} L(\boldsymbol{\phi}, y_k, \kappa, \rho_n, \boldsymbol{\varpi}_k) = & -\sum_{k=1}^K y_k + \kappa (\boldsymbol{\phi}^H \mathbf{U} \boldsymbol{\phi} - 1) + \sum_{k=1}^K \boldsymbol{\varpi}_k (\bar{f}_k + \\ & f_k) + \sum_{n=1}^N \rho_n (\boldsymbol{\phi}^H \tilde{\mathbf{F}} \boldsymbol{\phi} - \bar{I}_n^{\text{th}}), \end{aligned} \quad (36)$$

where  $\kappa, \boldsymbol{\varpi}_k$ , and  $\rho_n$  are the non-negative Lagrange multipliers.

Based on KKT conditions, the solution of  $y_k^*$  can be obtained as

$$y_k^* = \left[ \frac{\alpha_k p_k + \sigma \sqrt{\boldsymbol{\varpi}_k \alpha_k p_k}}{\sum_{i=1}^{k-1} p_i} \right]^+, \quad (37)$$

where the Lagrange multipliers are updated by

$$\boldsymbol{\varpi}_k(l+1) = \left[ \boldsymbol{\varpi}_k(l) + d_3(l) \times \{\bar{f}_k + f_k\} \right]^+, \quad (38)$$

$$\kappa(l+1) = \left[ \kappa(l) + d_4(l) \times (\boldsymbol{\phi}^H \mathbf{U} \boldsymbol{\phi} - 1) \right]^+, \quad (39)$$

$$\rho_n(l+1) = \left[ \rho_n(l) + d_5(l) \times (\boldsymbol{\phi}^H \tilde{\mathbf{F}} \boldsymbol{\phi} - \bar{I}_n^{\text{th}}) \right]^+, \quad (40)$$

where  $d_3(l), d_4(l)$  and  $d_5(l)$  are the step sizes. Based on the

same approach in Problem (28), the sub-gradient-based iterative method can guarantee the convergence of the proposed algorithm.

---

### Algorithm 2. An Iterative RA Scheme

---

**1: Input:**

$M, K, N, p^{\max}, I_n^{\text{th}}, \sigma_k^2, \mathbf{g}_k, \mathbf{h}, \mathbf{f}_n, \mathbf{g}_n^{\text{SM}}, \mathbf{g}_k^{\text{SS}}, h_n^{\text{MM}}, \varepsilon, L_l, L_j$ , and  $L_{\text{outer}}$ .

**2: Initialization:**  $\alpha_k, \beta_k, l = 0, p_k(0) = \frac{P^{\max}}{K}, \mu(0) > 0, \vartheta_n(0) >$

$0, \kappa(0) > 0, \rho_n(0) > 0, \boldsymbol{\varpi}_k(0) > 0, d_3(0), d_4(0), d_5(0)$ , and

$\boldsymbol{\Phi}(0) = \frac{\pi}{2} \mathbf{I}_M$ .

**3: while**  $l < L_{\text{max}}$  **do**

**4: Given**  $\boldsymbol{\Phi}$  **update**  $p_k, \forall k$ :

**5: for**  $i = 1:L_l$  **do**

6: Update transmit power  $p_k$  by Problem (26);

7: Update Lagrange multipliers  $\mu$  and  $\vartheta_n$  by Problems (27) and (28);

8: Update the auxiliary variable  $\alpha_k(i+1) = \frac{\gamma_k^{\text{SCU}}(i)}{1 + \gamma_k^{\text{SCU}}(i)}$  and  $\beta_k(i+1) = \log_2(1 + \gamma_k^{\text{SCU}}(i)) - \alpha(i) \log_2(\gamma_k^{\text{SCU}}(i))$ ;

**9: if**  $\| \mathbf{p}(i+1) - \mathbf{p}(i) \|^2 < \varepsilon$  **then**

10: Obtain the optimal transmit power  $\mathbf{p}^* = \mathbf{p}(i+1)$ ,

where  $\mathbf{p}(i+1) = [p_1(i+1), \dots, p_K(i+1)]^T$ ; break;

11: **else**

12:  $i = i + 1$ ;

13: **end if**

14: **end for**

15: Given  $\mathbf{p}^*, \forall k$  update  $\boldsymbol{\Phi}$ :

**16: for**  $j = 1:L_j$  **do**

17: Update the auxiliary variable  $y_k$  by Problem (38);

18: Update the Lagrange multipliers by Problems (39 - 41), respectively;

19: **Until**  $\| \boldsymbol{\gamma}(j+1) - \boldsymbol{\gamma}(j) \|^2 < \varepsilon$ , where  $\boldsymbol{\gamma}(j) =$

$[y_1(j), \dots, y_K(j)]^T$ ; Obtain  $y_k, \kappa, \rho_n$ , and  $\boldsymbol{\varpi}_k$ .

20: Obtain the RC matrix  $\boldsymbol{\Phi}$  by solving the SDP problem in problem (42).

21: **if**  $\| \boldsymbol{\Phi}(j+1) - \boldsymbol{\Phi}(j) \|^2 < \varepsilon$  **then**

22:  $\boldsymbol{\Phi}^* = \boldsymbol{\Phi}(j+1)$ ; break;

23: **else**

24:  $j = j + 1$ ;

25: **end if**

26: **end for**

27: **until** Convergence = true or  $l = L_{\text{outer}}$ .

28: **end while**

29: **Output:**  $\mathbf{p}^*$  and  $\boldsymbol{\Phi}$

---

In order to solve the PS coefficient, based on a Schur complement, the dual problem of Problem (35) becomes

$$\begin{aligned} & \max \gamma \\ & \text{s.t. } \gamma \geq 0, \\ & \tilde{\mathbf{M}} \geq 0, \end{aligned} \quad (41)$$

where  $\tilde{\mathbf{M}} = \begin{bmatrix} B_1 & B_2 \\ B_2^T & B_3 \end{bmatrix}$ ,  $B_1 = \kappa \mathbf{U} + \sum_{k=1}^K \varpi_k \bar{\mathbf{g}} \bar{\mathbf{g}}^H + \sum_{n=1}^N \rho_n \tilde{\mathbf{F}}$ ,  $B_2 = \sum_{k=1}^K \varpi_k \text{Re}\{g_k^{\text{SS},*} \bar{\mathbf{g}}_k^H\}$ , and  $B_3 = \sum_{k=1}^K \varpi_k f_k - \sum_{k=1}^K y_k - \kappa - \sum_{n=1}^N \rho_n \bar{I}_n^{\text{th}} - \gamma$ , and  $\gamma$  is an auxiliary variable. Problem (41) is a convex SDP problem and the Slater's constraint qualification is satisfied<sup>[28]</sup>. Particularly, an iterative-based RA scheme for the multiuser RIS-assisted heterogeneous NOMA network is summarized in Algorithm 2.

### 3.3 Computational Complexity

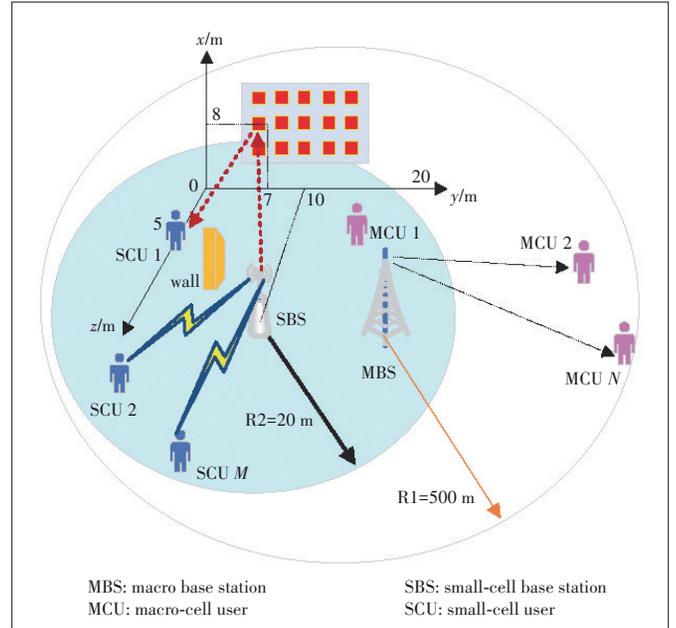
For the multi-user system, the complexity of the algorithm mainly depends on the number of users and the optimization approach. Regarding to Algorithm 2, it involves the inner and outer iterations. The inner layer is used to obtain variables  $\mathbf{p}$  and  $\Phi$  respectively, while the outer layer is taken for alternating iterations. We denote the maximum alternating iteration number as  $L_{\text{outer}}$ , and the iteration numbers of obtaining  $\mathbf{p}$  and  $\Phi$  as  $L_I$  and  $L_J$  respectively. There exists a polynomial-time complexity  $(K(N+1))$  to solve  $\mathbf{p}$  by using sub-gradient updating methods. Therefore, its computational complexity is  $(K(N+1)L_I)$ . Similarly, the computational complexity of the SDP problem to solve  $\phi_m$  is  $(\sqrt{M} + 2L \log(1/w))^{\text{[28]}}$ , where  $L = M^3 + M^6 + M^3(M+1)^2(M^3 + M + 1) + M^9$  and  $w$  is the convergence precision of the SDP problem. Based on the above discussion, the total complexity of Algorithm 2 is  $((K(N+1)L_I \sqrt{M} + 2L \log(1/w)L_J)L_{\text{outer}})$ .

## 4 Simulation Results

In this section, simulation results are given to demonstrate the effectiveness of the proposed RA scheme by comparing it with the rate-maximization based RA scheme without RIS<sup>[33]</sup> which is defined as "the traditional RA scheme without RIS". A spectrum-sharing small cell is randomly distributed in the coverage area of the macrocell. SCUs are uniformly distributed in the coverage area of their associated SBS. The coverage radii of the macrocell and small cell are 500 m and 20 m, respectively. The distance-dependent path loss model for the large-scale fading is given by  $\frac{1}{(d_m r_k)^\alpha}$ <sup>[13]</sup>, where  $d_m$  is the distance between the SBS to the reflecting unit  $m$  of RIS,  $r_k$  is the distance between the RIS and the SCU  $k$ , and  $\alpha$  denotes the path loss exponent. The distance between the SBS and the RIS is 10 m. The distance between RIS and the SCU is 5 m. The small-scale fading is considered as Rayleigh fading channel<sup>[13]</sup>. The stopping criterion for convergence is  $\varepsilon = 10^{-6}$ . The noise power is  $\sigma^2 = -100$  dBm, and  $\alpha = 3$ . The simulation setup is shown in Fig. 2. The detailed parameter settings are shown in Table 1.

### 4.1 Single-User Case

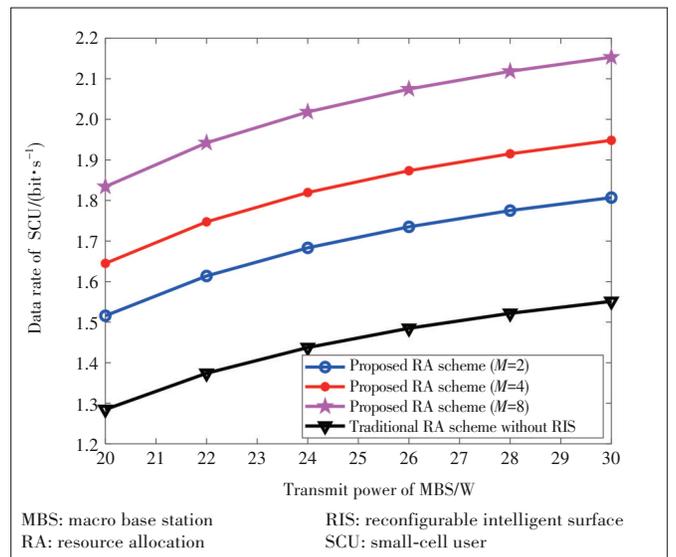
Fig. 3 depicts the data rate of the SCU versus the transmit power at the MBS. The target SINR of MCU is  $\gamma^{\text{min}} = 2$  dB. From this figure, the data rate of the SCU improves quickly as the increasing transmit power  $P$ . The reason is that the large



▲ Figure 2. Simulation setup

▼ Table 1. System parameters

Parameters	Values	Parameters	Values
$M$	$4^{\text{[1]}}$	$K$	$3^{\text{[1]}}$
$N$	$2^{\text{[29]}}$	$P$	20 dBm <sup>[29]</sup>
$I_n^{\text{th}}$	$10^{-3\text{[21]}}$	$p^{\text{max}}$	20 mW <sup>[21]</sup>
$\gamma^{\text{min}}$	2 dB <sup>[29]</sup>	$r_k$	5 m <sup>[6]</sup>
$d_m$	10 m <sup>[6]</sup>	$\varepsilon$	$10^{-6\text{[21]}}$
$\alpha$	$3^{\text{[6]}}$	$\sigma^2$	-100 dBm <sup>[6]</sup>



▲ Figure 3. Data rate of SCU versus transmit power of the MBS  $P$

transmit power of the MBS increases the available transmit power from the SBS to the SCU according to the constraint  $C_1$ . Besides, this figure also shows that the data rate under the proposed RA scheme is larger than that of the RA scheme without RIS, which indicates that the RIS can improve the rate performance in a passive reflecting way. Moreover, the data rate increases heavily with the increasing number of reflecting elements at the RIS, which indicates that a massive number of passive reflecting elements brings better performance improvement.

Fig. 4 shows the data rate of the SCU versus the minimum SINR requirement of the MCU (e.g.,  $\gamma^{\min}$ ). The transmit power of the MBS is assumed to be  $P = 20$  dBm. With the increasing  $\gamma^{\min}$ , the data rate of the SCU decreases a lot. The reason is that, under the fixed transmit power at the MBS, the allowed transmit power from the SBS to the SCU becomes smaller to avoid less interference power to the MCU. It is noted that as for the achieved transmission performance, i.e., the achieved data rate, the proposed RA scheme is better than that of the RA scheme without RIS.

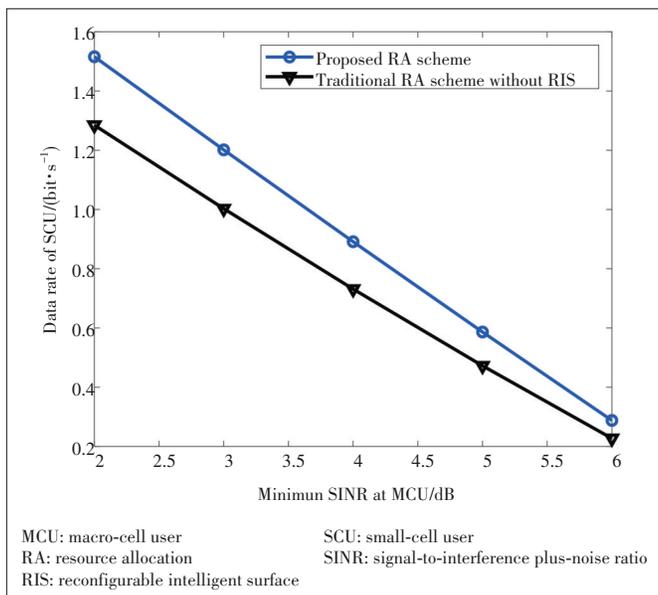
Fig. 5 gives the allocated power to the SCU versus the number of reflecting elements  $M$  under different transmit power of the MBS  $P$ . It is observed that the required transmit power of the proposed RA scheme is around 40% lower than that of the RA scheme without RIS under different transmit power  $P$ . This demonstrates that the energy consumption can be reduced a lot when the RIS is adopted. Moreover, the received transmit power at the SCU decreases with the increasing number of reflecting elements  $M$ . Because larger reflecting elements provide more multipath interference to the MCU. The available transmit power of the SBS decreases for guaranteeing the QoS of the MCU (namely, cross-tier interference constraint). When the transmit power at the MBS is increased, the available transmit power at the SBS is increased accordingly, since large transmit power at

the MBS allows more transmit power to improve the communication quality of SCU via constraint (4).

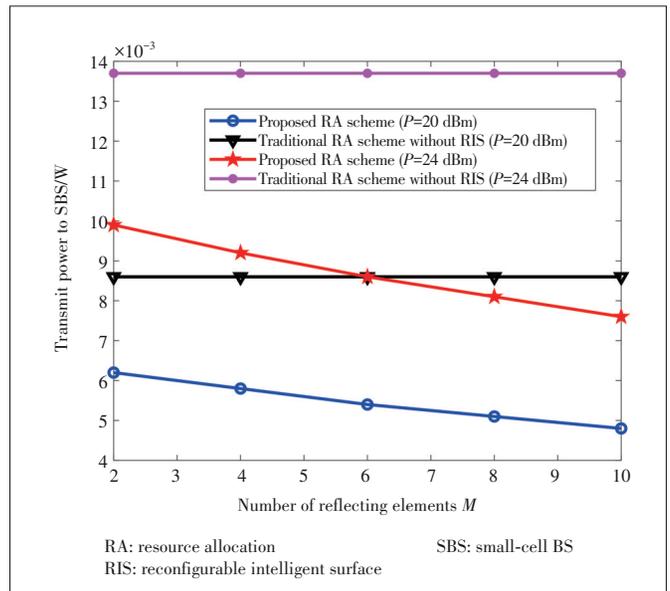
### 4.2 Multi-User Case

The numbers of SCUs and MCUs are  $K=3$  and  $N=2$ , respectively. The reflecting elements are  $M=4$ . Assume that each MCU receiver has the same interference power limit. The maximum transmit power of the SBS is  $p^{\max} = 20$  mW, and the interference power limit is  $I_{th}^n = 10^{-3}$  mW<sup>[33]</sup>.

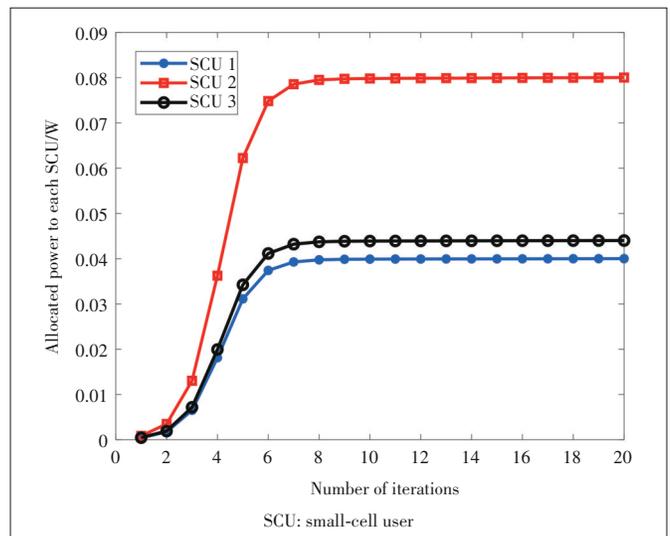
Fig. 6 shows the convergence performance of the proposed RA scheme. It is observed that the allocated power to each SCU can quickly reach the equilibrium points within eight iterations, which demonstrates that the proposed RA scheme has a good convergence.



▲ Figure 4. Data rate of SCU versus SINR threshold of MCU  $\gamma^{\min}$



▲ Figure 5. Allocated power from SBS to small-cell user (SCU) versus the number of reflecting elements  $M$



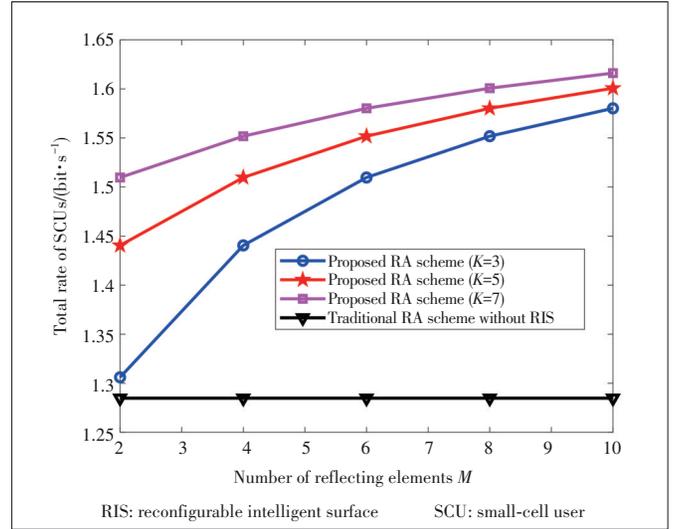
▲ Figure 6. Convergence of proposed resource allocation (RA) scheme

Fig. 7 shows the sum rate of all SCUs versus the number of reflecting elements with different SCUs. It is observed that the total rate achieved by SCUs under the proposed RA scheme increases with the increasing number of reflecting elements. Since the RIS can provide more reflecting signals to strengthen the desired signals to SCUs. However, the total rate of SCUs under the traditional RA scheme always keeps stable with the increasing number of reflecting elements  $M$  since no RIS is deployed in this scheme. Besides, the total rate of SCUs with a large  $K$  is better than that with a small  $K$ . The gap of the total rate of SCUs under a small  $m$  is larger than that under a large  $M$ . Because the interference power among different SCUs is small due to less transmission path, the total rate is decided by the number of users. Thus, there is a big performance gap in the region with small reflecting elements. However, due to the effect of the cross-tier interference constraint, the overall performance of SCUs is limited. It is impossible to unrestrictedly improve the performance with the increasing number of reflecting elements  $M$ .

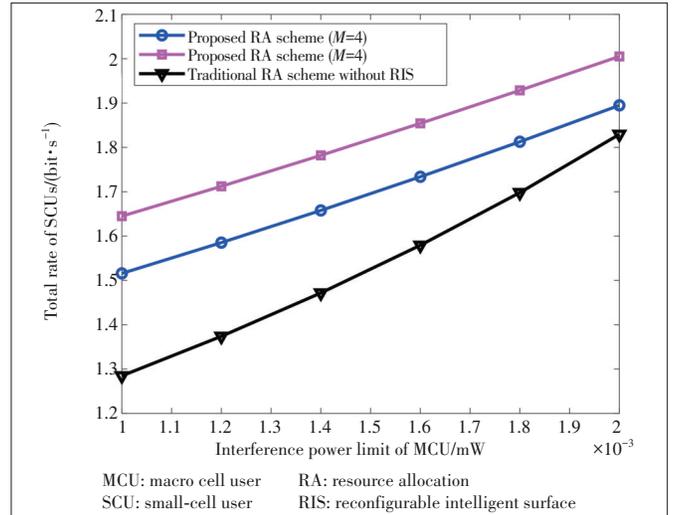
Fig. 8 shows the total rate of SCUs versus the different interference power limit of MCU  $I_n^{th}$ . The total rate of SCUs increases with the increasing  $I_n^{th}$ . The bigger  $I_n^{th}$  means that the MCU can tolerate more harmful interference power from SCUs and RIS. Correspondingly, the SBS allocates more transmit power to each SCU for pursuing a higher data rate. Moreover, the proposed RA scheme with a larger  $M$  achieves better performance than the other two cases. The reason is that the RIS can effectively enhance system performance by providing multiple flexible transmission paths.

## 5 Conclusions

In this paper, we have maximized the total rate of SCUs by jointly optimizing the transmit power at the SBS and PS coefficients at the RIS in RIS-assisted heterogeneous NOMA networks. Specifically, we firstly propose an optimal RA scheme to maximize the transmission rate of the SCU in a single-user scenario, subject to the SINR constraint of MCU, the maximum transmit power constraint of SBS, and the PS coefficient constraint. Then, we design a suboptimal RA scheme to maximize the total rate of SCUs for the multi-user scenario, subject to the sum transmit power constraint of the SBS, the cross-tier interference constraint of each MCU, and the unit modulus constraint of the RIS. The alternating iteration approach and the SCA approach are used to solve the originally non-convex optimization problem. Simulation results show that the proposed RA scheme has a good converge and transmission rate compared with the traditional RA scheme without RIS. In our future work, we will study the joint optimization of beamforming and PS coefficients under the discrete PS constraint and imperfect CSI in a multi-antenna heterogeneous NOMA network with RIS.



▲ Figure 7. Total rate of SCUs versus the number of reflecting elements  $M$



▲ Figure 8. Total rate of SCUs versus the interference power limit of the MCU  $I_n^{th}$

## Appendix A: Proof of Theorem 1

Based on Problem (7), an equivalent objective function is given by

$$Q(\lambda) = \max_{\phi_n} A \left| \mathbf{g}^H \Phi \mathbf{h} + \mathbf{g}^{SS} \right|^2 - \lambda \left| \mathbf{g}^{SM} + \mathbf{f}^H \Phi \mathbf{h} \right|^2. \quad (\text{A.1})$$

Assume there are two optimal auxiliary variables  $\lambda_1$  and  $\lambda_2$  ( $\lambda_1 > \lambda_2$ ) for the optimal phase shift policies  $\Phi_1$  and  $\Phi_2$ , respectively, then

$$\begin{aligned} Q(\lambda_2) &= A \left| \mathbf{g}^H \Phi_2 \mathbf{h} + \mathbf{g}^{SS} \right|^2 - \lambda_2 \left| \mathbf{g}^{SM} + \mathbf{f}^H \Phi_2 \mathbf{h} \right|^2 > \\ &A \left| \mathbf{g}^H \Phi_1 \mathbf{h} + \mathbf{g}^{SS} \right|^2 - \lambda_2 \left| \mathbf{g}^{SM} + \mathbf{f}^H \Phi_1 \mathbf{h} \right|^2 > \\ &A \left| \mathbf{g}^H \Phi_1 \mathbf{h} + \mathbf{g}^{SS} \right|^2 - \lambda_1 \left| \mathbf{g}^{SM} + \mathbf{f}^H \Phi_1 \mathbf{h} \right|^2 = \\ &Q(\lambda_1). \end{aligned} \quad (\text{A.2})$$

Therefore,  $Q(\lambda)$  is a strictly decreasing function with the variable  $\lambda$ . Define  $\Phi'$  as any solution for Problem (7) and

$$\lambda' = \frac{A |g^H \Phi' \mathbf{h} + g^{SS}|^2}{|g^{SM} + f^H \Phi' \mathbf{h}|^2}, \text{ we have}$$

$$Q(\lambda') = \max_{\phi_m} A |g^H \Phi \mathbf{h} + g^{SS}|^2 - \lambda' |g^{SM} + f^H \Phi \mathbf{h}|^2 \geq 0$$

$$A |g^H \Phi' \mathbf{h} + g^{SS}|^2 - \lambda' |g^{SM} + f^H \Phi' \mathbf{h}|^2 = 0. \quad (\text{A.3})$$

As a result, we have  $Q(\cdot) \geq 0$ . The proof is complete.

## Appendix B: Proof of Convergence

According to the algorithm procedure of Algorithm 1, the convergence is determined by the Dinkelbach-based approach. To show the convergence of Dinkelbach's method, we define  $(\lambda(l), \phi_m(l))$  as the optimal solution in the  $l$ -th iteration,  $\lambda(l) \neq \lambda^*$  and  $\lambda(l+1) \neq \lambda^*$  as the EE at the  $l$ -th and  $(l+1)$ -th iteration, respectively. We have the following proposition.

**Proposition:** If  $(\lambda', \phi'_m)$  is an arbitrary feasible solution of Problem (7), we have

$$\lambda' = \frac{A |g^H \Phi' \mathbf{h} + g^{SS}|^2}{|g^{SM} + f^H \Phi' \mathbf{h}|^2}, \quad (\text{B.1})$$

then  $F(\lambda', \Phi) \geq 0$ .

**Proof:** We first have

$$F(\lambda', \Phi) = \max_{\phi_m} \left\{ A |g^H \Phi \mathbf{h} + g^{SS}|^2 - \lambda' |g^{SM} + f^H \Phi \mathbf{h}|^2 \right\} \geq 0$$

$$A |g^H \Phi' \mathbf{h} + g^{SS}|^2 - \lambda' |g^{SM} + f^H \Phi' \mathbf{h}|^2 = 0. \quad (\text{B.2})$$

Based on the Proposition, we have  $F(\lambda', \Phi(l)) > 0$  and  $F(\lambda^{l+1}, \Phi(l+1)) > 0$  because of  $\lambda(l) \neq \lambda^*$  and  $\lambda(l+1) \neq \lambda^*$ . As a result,  $F(\lambda', \Phi(l)) > 0$  can be rewritten as

$$F(\lambda(l), \Phi(l)) = A |g^H \Phi(l) \mathbf{h} + g^{SS}|^2 - \lambda(l) |g^{SM} + f^H \Phi(l) \mathbf{h}|^2 =$$

$$|g^{SM} + f^H \Phi(l) \mathbf{h}|^2 \{ \lambda(l+1) - \lambda(l) \} > 0 \Rightarrow$$

$$\lambda(l+1) > \lambda(l). \quad (\text{B.3})$$

Since  $\lambda(l+1) > \lambda(l)$ , while  $F(\lambda, \Phi)$  is a strictly decreasing function in  $\lambda$  according to Theorem 1, we can show that with the increase of iterations,  $F(\lambda(l), \Phi(l))$  can gradually approach to zero, namely  $\lim_{l \rightarrow \infty} F(\lambda(l), \Phi(l)) = 0$ . Accordingly, we have  $\lambda(l+1) = \lambda(l)$ ,  $l \rightarrow \infty$ . Thus, the Dinkelbach-based approach can guarantee the convergence of the algorithm. The proof is complete.

## References

- [1] XU Y, GAO Z, WANG Z, et al. RIS-enhanced WPCNs: joint radio resource allocation and passive beamforming optimization [J]. IEEE Transactions on vehicular technology, 2021, 70(8): 7980 - 7991
- [2] HUANG C W, HU S, ALEXANDROPOULOS G C, et al. Holographic MIMO surfaces for 6G wireless networks: opportunities, challenges, and trends [J]. IEEE wireless communications, 2020, 27(5): 118 - 125. DOI: 10.1109/MWC.001.1900534transactions
- [3] HUANG C, MO R, YUEN C. Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning [J]. IEEE journal on selected areas in communications, 2020, 38(8):1839 - 1850. DOI: 10.1109/JSAC.2020.3000835
- [4] GAO Z, XU Y, WANG Q, et al. Outage-constrained energy efficiency maximization for RIS-assisted WPCNs [J]. IEEE communications letters, 2021, 25(10): 3370 - 3374
- [5] YU X, XU D, SCHÖBER R. Miso wireless communication systems via intelligent reflecting surfaces [C]//2019 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2019: 735 - 740
- [6] GUO H, LIANG Y C, CHEN J, et al. Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks [J]. IEEE transactions on wireless communications, 2020, 19(5):3064 - 3076
- [7] HUANG C, ZAPPONE A, ALEXANDROPOULOS G C, et al. Reconfigurable intelligent surfaces for energy efficiency in wireless communication [J]. IEEE transactions on wireless communications, 2019, 18(8):4157 - 4170
- [8] JIANG W, ZHANG Y, WU J, et al. Intelligent reflecting surface assisted secure wireless communications with multiple-transmit and multiple-receive antennas [J]. IEEE access, 2020, 8: 86659 - 86673. DOI: 10.1109/ACCESS.2020.2992613
- [9] SOUTO V D P, SOUZA R D, UCHÔA-FILHO B F, et al. Beamforming optimization for intelligent reflecting surfaces without CSI [J]. IEEE Wireless Communications Letters, 2020, 9(9):1476 - 1480
- [10] PAN C H, REN H, WANG K Z, et al. Multicell MIMO communications relying on intelligent reflecting surfaces [J]. IEEE transactions on wireless communications, 2020, 19(8): 5218 - 5233. DOI: 10.1109/TWC.2020.2990766
- [11] CHEN W J, MA X Y, LI Z X, et al. Sum-rate maximization for intelligent reflecting surface based terahertz communication systems [C]//Proceedings of 2019 IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops). IEEE, 2019: 153 - 157. DOI: 10.1109/ICCCChinaW.2019.8849960
- [12] WU Q Q, ZHANG R. Weighted sum power maximization for intelligent reflecting surface aided SWIPT [J]. IEEE wireless communications letters, 2020, 9(5): 586 - 590. DOI: 10.1109/LWC.2019.2961656
- [13] DING Z G, SCHÖBER R, POOR H V. On the impact of phase shifting designs on IRS-NOMA [J]. IEEE wireless communications letters, 2020, 9(10): 1596 - 1600. DOI: 10.1109/LWC.2020.2991116
- [14] WANG H, LIU C, SHI Z, et al. On power minimization for IRS-aided downlink NOMA systems [J]. IEEE wireless communications letters, 2020, 9(11): 1808 - 1811. DOI: 10.1109/LWC.2020.2999097
- [15] FU M, ZHOU Y, SHI Y M, et al. Reconfigurable intelligent surface empowered downlink non-orthogonal multiple access [J]. IEEE transactions on communications, 2021, 69(6): 3802 - 3817. DOI: 10.1109/TCOMM.2021.3066587
- [16] WANG W, LIU Y W, LUO Z Q, et al. Toward cross-layer design for non-orthogonal multiple access: A quality-of-experience perspective [J]. IEEE wireless communications, 2018, 25(2): 118 - 124. DOI: 10.1109/MWC.2018.1700081
- [17] HOU T W, LIU Y W, SONG Z Y, et al. Reconfigurable intelligent surface aided NOMA networks [J]. IEEE journal on selected areas in communications, 2020, 38(11): 2575 - 2588. DOI: 10.1109/JSAC.2020.3007039
- [18] BJÖRNSSON E, ÖZDOĞAN Ö, LARSSON E G. Intelligent reflecting surface versus decode-and-forward: how large surfaces are needed to beat relaying? [J]. IEEE wireless communications letters, 2020, 9(2): 244 - 248. DOI: 10.1109/LWC.2019.2950624
- [19] CANBILEN A E, BASAR E, IKKI S S. Reconfigurable intelligent surface-assisted space shift keying [J]. IEEE wireless communications letters, 2020, 9(9): 1495 - 1499. DOI: 10.1109/LWC.2020.2994930
- [20] WEI L, HUANG C W, ALEXANDROPOULOS G C, et al. Channel estimation

- for RIS-empowered multi-user MISO wireless communications [J]. *IEEE transactions on communications*, 2021, 69(6): 4144 – 4157. DOI: 10.1109/TCOMM.2021.3063236
- [21] XU Y J, LI G Q, YANG Y, et al. Robust resource allocation and power splitting in SWIPT enabled heterogeneous networks: A robust minimax approach [J]. *IEEE Internet of Things journal*, 2019, 6(6): 10799 – 10811. DOI: 10.1109/JIOT.2019.2941897
- [22] DAMNJANOVIC A, MONTOJO J, WEI Y B, et al. A survey on 3GPP heterogeneous networks [J]. *IEEE wireless communications*, 2011, 18(3): 10 – 21. DOI: 10.1109/MWC.2011.5876496
- [23] XU Y J, GUI G, GACANIN H, et al. A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges [J]. *IEEE communications surveys & tutorials*, 2021, 23(2): 668 – 695. DOI: 10.1109/COMST.2021.3059896
- [24] DINKELBACH W. On nonlinear fractional programming [J]. *Management science*, 1967, 13(7): 492 – 498. DOI: 10.1287/mnsc.13.7.492
- [25] BOYD S, VANDENBERGHE L. *Convex Optimization* [M]. Cambridge: Cambridge University Press, 2004. DOI: 10.1017/cbo9780511804441
- [26] STURM J F. Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones [J]. *Optimization methods and software*, 1999, 11(1 – 4): 625 – 653. DOI: 10.1080/10556789908805766
- [27] CROUZEIX J P, FERLAND J A. Algorithms for generalized fractional programming [J]. *Mathematical programming*, 1991, 52(1 – 3): 191 – 207. DOI: 10.1007/BF01582887
- [28] ZHANG H J, HUANG S T, JIANG C X, et al. Energy efficient user association and power allocation in millimeter-wave-based ultra-dense networks with energy harvesting base stations [J]. *IEEE journal on selected areas in communications*, 2017, 35(9): 1936 – 1947. DOI: 10.1109/JSAC.2017.2720898
- [29] ZHANG H J, JIANG C X, BEAULIEU N C, et al. Resource allocation in spectrum-sharing OFDMA femtocells with heterogeneous services [J]. *IEEE transactions on communications*, 2014, 62(7): 2366 – 2377. DOI: 10.1109/TCOMM.2014.2328574
- [30] PAPANDRIOPOULOS J, EVANS J S. SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks [J]. *IEEE transactions on information theory*, 2009, 55(8): 3711 – 3724. DOI: 10.1109/TIT.2009.2023751
- [31] PALOMAR D P, CHIANG M. A tutorial on decomposition methods for network utility maximization [J]. *IEEE journal on selected areas in communications*, 2006, 24(8): 1439 – 1451. DOI: 10.1109/JSAC.2006.879350
- [32] ZHU Z Y, WANG N, HAO W M, et al. Robust beamforming designs in secure MIMO SWIPT IoT networks with a nonlinear channel model [J]. *IEEE Internet of Things journal*, 2021, 8(3): 1702 – 1715. DOI: 10.1109/JIOT.2020.3014933
- [33] OGUEJIOFOR O S, ZHANG L X, NAWAZ N. UE-centric clustering and resource allocation for practical two-tier heterogeneous cellular networks [J]. *IET communications*, 2018, 12(18): 2384 – 2392. DOI: 10.1049/iet-com.2018.5422

### Biographies

**XU Yongjun** received his Ph.D. degree (Hons.) in communication and information system from Jilin University, China in 2015. He received the Outstanding Doctoral Thesis of Jilin Province in 2016. He is currently an associate professor with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, China. From 2018 to 2019, he was a visiting scholar with Utah State University, USA. He has authored or coauthored more than 100 papers for renowned journals, including *IEEE COMMUN SURV/TVT/TCOM/THITCCN/TGCN/CL/WCL*, etc. His recent interests include heterogeneous networks, resource allocation, intelligent reflecting surface, energy harvesting, backscatter communications, mobile edge computing, and UAV communi-

tions. He serves as an associate editor of *EURASIP Journal on Wireless Communications and Networking*, an editor of *Physical Communications*, and an academic editor of *Digital Communications and Networks*. He serves as a reviewer for *IEEE TWC/TSP/ITCOM/TVT/ICL*, etc. He is a senior member of IEEE.

**YANG Zhaohui** received his Ph.D. degree in communication and information system with the National Mobile Communications Research Laboratory, Southeast University, China in May 2018. From May 2018 to October 2020, he was a postdoctoral research associate with the Center for Telecommunications Research, Department of Informatics, King's College London, UK. He is currently a visiting associate professor with College of Information Science and Electronic Engineering Zhejiang Key Lab of Information Processing Communication and Networking, Zhejiang University, China and also a research fellow with the Department of Electronic and Electrical Engineering, University College London, UK. He is an associate editor for the *IEEE Communications Letters*, *IET Communications* and *EURASIP Journal on Wireless Communications and Networking*. He was an exemplary reviewer for *IEEE Transactions on Communications* in 2019 and 2020.

**HUANG Chongwen** (chongwenhuang@zju.edu.cn) obtained his B.Sc. degree in 2010 from Nankai University, China, and the M.Sc. degree from the University of Electronic Science and Technology of China in 2013, and Ph.D. degree from Singapore University of Technology and Design (SUTD). From Oct. 2019 to Sept. 2020, he is a Postdoc in his mother University SUTD. Since Sept. 2020, he joined Zhejiang University as a tenure-track young professor. Dr. HUANG is the recipient of IEEE Marconi Prize Paper Award in Wireless Communications, and IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2021. He has served as an editor of *IEEE Communications Letter*, *Elsevier Signal Processing*, *EURASIP Journal on Wireless Communications and Networking* and *Physical Communication* since 2021. His main research interests focus on holographic MIMO surface/reconfigurable intelligent surface, B5G/6G wireless communication, mmWave/THz communications, deep learning technologies for wireless communications, etc.

**YUEN Chau** received his B.Eng. and Ph.D. degrees from Nanyang Technological University (NTU), Singapore in 2000 and 2004, respectively. He was a Post-Doctoral Fellow with Lucent Technologies Bell Labs, USA in 2005, and a visiting assistant professor with The Hong Kong Polytechnic University, China in 2008. From 2006 to 2010, he was with the Institute for Infocomm Research (I2R), Singapore, where he was involved in an industrial project on developing an 802.11n Wireless LAN system and participated actively in 3Gpp LTE and LTE-Advanced (LTE-A) standardization. Since 2010, he has been with the Singapore University of Technology and Design. He was a recipient of the Lee Kuan Yew Gold Medal, the Institution of Electrical Engineers Book Prize, the Institute of Engineering of Singapore Gold Medal, the Merck Sharp and Dohme Gold Medal, and twice a recipient of the Hewlett Packard Prize. He is a fellow of IEEE.

**GUI Guan** received his doctoral degree in information and communication engineering from University of Electronic Science and Technology of China in 2012. Since 2015, he has been a professor with Nanjing University of Posts and Telecommunications (NJUPT), China. His recent research interests include artificial intelligence, deep learning, non-orthogonal multiple access, wireless power transfer, and physical layer security. He received the IEEE Communications Society Heinrich Hertz Award in 2021 and Highly Cited Chinese Researchers by Elsevier in 2020. In addition, he served as the IEEE VTS Ad Hoc Committee Member in AI Wireless, Executive Chair of VTC 2021-Fall, Vice Chair of WCNC 2021, TPC Chair of PHM 2021, Symposium Chair of WCSP 2021, and General Co-Chair of Mobimedia 2020.



# Markovian Cascaded Channel Estimation for RIS Aided Massive MIMO Using 1-Bit ADCs and Oversampling

SHAO Zhichao, YAN Wenjing, YUAN Xiaojun

(National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China)

DOI: 10.12142/ZTECOM.202201007

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220305.1545.002.html>, published online March 8, 2022

Manuscript received: 2021-12-27

**Abstract:** A reconfigurable intelligent surface (RIS) aided massive multiple-input multiple-output (MIMO) system is considered, where the base station employs a large antenna array with low-cost and low-power 1-bit analog-to-digital converters (ADCs). To compensate for the performance loss caused by the coarse quantization, oversampling is applied at the receiver. The main challenge for the acquisition of cascaded channel state information in such a system is to handle the distortion caused by the 1-bit quantization and the sample correlation caused by oversampling. In this work, Bussgang decomposition is applied to deal with the coarse quantization, and a Markov chain is developed to characterize the banded structure of the oversampling filter. An approximate message-passing based algorithm is proposed for the estimation of the cascaded channels. Simulation results demonstrate that our proposed 1-bit systems with oversampling can approach the 2-bit systems in terms of the mean square error performance while the former consumes much less power at the receiver.

**Keywords:** massive MIMO; reconfigurable intelligent surface; channel estimation; 1-bit ADCs; oversampling

**Citation** (IEEE Format): Z. C. Shao, W. J. Yan, and X. J. Yuan, "Markovian cascaded channel estimation for RIS aided massive MIMO using 1-bit ADCs and oversampling," *ZTE Communications*, vol. 20, no. 1, pp. 48 – 56, Mar. 2022. doi: 10.12142/ZTECOM.202201007.

## 1 Introduction

Massive multiple-input multiple-output (MIMO) has been identified as a key technology for future communication systems<sup>[1]</sup>. In fact, large spatial degrees of freedoms (DoFs) can increase spectral efficiency without requiring extra spectral resources. Recently, reconfigurable intelligent surfaces (RISs) have been proposed as a cost-effective technology for tuning the wireless propagation channel among transceivers<sup>[2]</sup>. A RIS consists of a large number of meta-atoms that can be controlled by the software to modify their phase shifts, so that incident electromagnetic waves can be mostly reflected to the desired receiver, which makes the wireless transmission more energy-efficient. The combination of RIS and massive MIMO is treated as one of the promising technologies for the sixth-generation wireless communication systems<sup>[3]</sup>.

The channel estimation method for RIS aided massive MIMO systems is a serious challenge, since there exist two cascaded channels, namely, the channel between the users and the RIS and the channel between the RIS and the base station (BS), to be estimated. The acquisition of channel state information (CSI) has been recently studied in Refs. [4 – 7]. TAHA et al.<sup>[4]</sup> considered a RIS architecture which is a mixture of active and passive elements. This method facilitates the channel estimation but increases the hardware cost and

energy consumption. In Ref. [5], a non-iterative two-stage channel estimation framework for passive RIS aided millimeter-wave MIMO systems was proposed, where every stage is formulated as a multi-dimensional direction-of-arrival estimation problem. Similarly, the authors in Ref. [6] have proposed a two-stage channel estimation algorithm, namely, sparse matrix factorization and matrix completion, to exploit the rank-deficient structure of the channel. In Ref. [7], the cascade channel estimation is converted into a sparse signal recovery problem by utilizing the properties of Katri-Rao and Kronecker products.

The receiver design in massive MIMO systems, however, becomes more challenging since the power consumption increases rapidly as the number of antennas grows. Among all the components in the radio frequency (RF) chain, a large portion of the total power consumption lies in the analog-to-digital converters (ADCs), whose power consumption grows exponentially with the number of quantization bits<sup>[9]</sup>. The deployment of current high-resolution (8 – 12 bits) ADCs is a critical bottleneck for the practical use of large-scale MIMO. To alleviate this issue, the use of low-resolution ADCs (1 – 4 bits) can largely reduce the power consumption and is more suitable for the deployment of large-scale MIMO systems.

In this paper, we consider the extreme case of 1-bit resolu-

tion, where the in-phase and quadrature components of the received samples are separately quantized to 1 bit. This solution is particularly attractive to massive MIMO systems, since each of the RF chains only contains simple limiting amplifiers (LAs) without the automatic gain control (AGC). This hardware change can largely reduce both the power consumption and the hardware cost at the BS. Prior works on 1-bit massive MIMO have analyzed the sum rate<sup>[10]</sup>, channel estimation<sup>[11]</sup>, and signal detection<sup>[12]</sup>. Moreover, oversampling is applied to further compensate for the performance loss caused by the coarse quantization<sup>[13-14]</sup>. Furthermore, the distortion caused by 1-bit quantization and the sample correlation caused by oversampling make the cascaded channel estimation problem even more challenging.

In this paper, we develop an approximate message passing (AMP) based algorithm to solve the considered cascaded channel estimation problem, where the received signal is sampled at a rate beyond Nyquist sampling and quantized to 1-bit. Bussgang decomposition is applied to deal with the coarse quantization and a Markov chain is developed to characterize the correlation of adjacent oversampled samples. The corresponding factor graph is presented and the AMP algorithm is derived. Unlike prior works on AMP-based cascaded channel estimation<sup>[7-8]</sup>, this work considers the statistical characteristic of 1-bit quantization and uses the oversampling technique to increase the estimation accuracy. Simulations show that our proposed algorithm outperforms the method in Refs. [7-8] and can even approach the 2-bit Nyquist-sampled systems in terms of the normalized mean square error (NMSE) while the former consumes less power at the receiver.

## 2 System Model and Problem Statement

In this work, a single-cell uplink RIS aided multi-user 1-bit massive MIMO system with  $N_t$  single-antenna users, a RIS with  $L$  passive reflecting elements, and a BS with  $N_r$  receive antennas are considered, where  $N_r \gg N_t$ . The system model is depicted in Fig. 1<sup>1</sup>, where  $p(t)$  is the pulse shaping filter for

transmission and  $m(t)$  is the matched filter for detection. The received data signal at the  $n_r$ -th receive antenna  $y_{n_r}^d(t)$  is

$$y_{n_r}^d(t) = m(t) * \sum_{l=1}^L h_{n_r,l}(t) * \left( s_l(t) \left( g_{l,n_t}(t) * p(t) * x_{n_t}(t) \right) \right), \quad (1)$$

where  $x_{n_t}(t)$  is the transmitted signal from the  $n_t$ -th user;  $s_l(t)$  represents the  $l$ -th reflecting element at the RIS;  $g_{l,n_t}(t)$  and  $h_{n_r,l}(t)$  are the channel impulse responses from the user  $n_t$  to the  $l$ -th reflecting element and from the  $l$ -th reflecting element to the  $n_r$ -th receive antenna, respectively;  $*$  denotes the operation of convolution.

Flat fading channels are considered in this work, i. e.,  $g_{n_t,l}(t)$  and  $h_{n_r,l}(t)$  can be written as

$$g_{l,n_t}(t) = g_{l,n_t} \delta(t) \quad \text{and} \quad h_{n_r,l}(t) = h_{n_r,l} \delta(t), \quad (2)$$

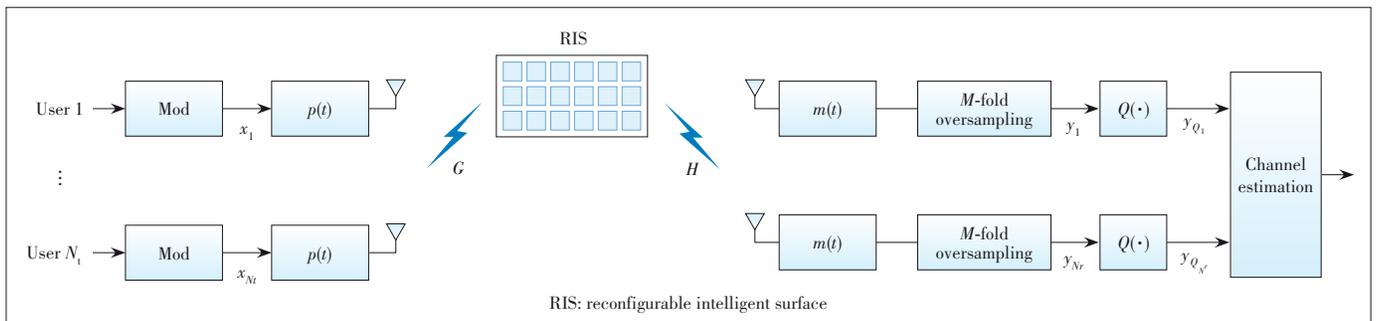
where  $g_{n_t,l}$  and  $h_{n_r,l}$  are the corresponding channel gains and  $\delta(t)$  is the Dirac delta function. Consider a transmission block with length  $N$ :

$$x_{n_t}(t) = \sum_{i=0}^{N-1} x_{n_t,i} \delta(t - iT_s) \quad \text{and} \quad s_l(t) = \sum_{i=0}^{N-1} s_{l,i} e^{j\theta_{l,i}} \delta_{l,iT_s}, \quad (3)$$

where  $T_s$  is the symbol duration;  $x_{n_t,i}$  is the transmitted symbol at the time instant  $iT_s$ ;  $s_{l,i} \in \{0,1\}$  is the on/off state<sup>2</sup> and  $\theta_{l,i} \in (0, 2\pi]$  is the phase shift of the  $l$ -th reflecting element of RIS at the time instant  $iT_s$ ;  $\delta_{l,iT_s}$  is the Kronecker delta function, where  $\delta_{l,iT_s} = 1$  for  $t = t'$  and  $\delta_{l,iT_s} = 0$  otherwise. Eq. (1) can be simplified as:

$$y_{n_r}^d(t) = \sum_{l=1}^L h_{n_r,l} g_{l,n_t} z(t) * \left( s_l(t) x_{n_t}(t) \right), \quad (4)$$

where  $z(t) = p(t) * m(t)$ . In oversampled systems, Eq. (4) can be discretized as:



▲ Figure 1. System model of RIS aided multi-user 1-bit massive MIMO system with oversampling at the receiver

1. Note that the channel matrix in the direct link can be estimated by turning off the RIS<sup>[13]</sup>. Therefore, the direct-link channel estimation is omitted throughout the paper.

2. In this paper, we assume that the change of the reflection coefficients is synchronized with the transmitted signal. Moreover, the response time of the PIN diode at each reflecting element is assumed to be small enough, so that the duration of each reflection coefficient is the same as the symbol duration. With the above ideal assumptions, no extra harmonics are generated from the surface. The non-ideal case that extra harmonics are generated is beyond the scope of this paper.

$$y_{n_r}^d \left[ \frac{i}{M} \right] = \sum_{l=1}^L h_{n_r,l} g_{l,n_r} \sum_{k=-M}^M z \left[ \frac{k}{M} \right] s_l \left[ \frac{i-k}{M} \right] x_{n_r} \left[ \frac{i-k}{M} \right],$$

$$0 \leq i \leq MN - 1, \quad (5)$$

where  $M$  is the oversampling factor. Since there are no data symbols at time instants  $\frac{i}{M}T_s$  ( $i \neq 0, M, 2M, \dots$ ), each pair of transmitted data symbols at adjacent time instants are interpolated with  $M - 1$  zeros. Let  $\mathbf{x}_{n_r} \triangleq [x_{n_r,0}, \dots, x_{n_r,N-1}]^T \in \mathbb{C}^{N \times 1}$ ,  $\mathbf{s}_l \triangleq [s_{l,0} e^{j\theta_{l,0}}, \dots, s_{l,N-1} e^{j\theta_{l,N-1}}]^T \in \mathbb{C}^{N \times 1}$  and  $y_{n_r}^d \left[ \frac{i}{M} \right] \triangleq y_{n_r}^d \left[ \frac{i}{M} \right]$ , and we rewrite Eq. (5) in a matrix form as:

$$\mathbf{y}_{n_r}^d = \begin{bmatrix} y_{n_r,0}^d & y_{n_r, \frac{1}{M}}^d & \dots & y_{n_r, \frac{MN-1}{M}}^d \end{bmatrix}^T = \sum_{l=1}^L h_{n_r,l} g_{l,n_r} \mathbf{Z} (\mathbf{I}_N \otimes \mathbf{u}) \text{diag}(\mathbf{s}_l) \mathbf{x}_{n_r}, \quad (6)$$

where  $\mathbf{Z} \in \mathbb{R}^{MN \times MN}$  is the Toeplitz matrix with the form as

$$\mathbf{Z} = \begin{bmatrix} z[0] & z\left[\frac{1}{M}\right] & \dots & z[1] & 0 & 0 & \dots & 0 & 0 \\ z\left[\frac{1}{M}\right] & z[0] & \dots & z\left[\frac{M-1}{M}\right] & z[1] & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & z[-1] & \dots & z\left[\frac{1}{M}\right] & z[0] \end{bmatrix}. \quad (7)$$

In Eq. (6),  $\mathbf{u} = [0, \dots, 0, 1]^T \in \mathbb{R}^{M \times 1}$  is the zero-inserting vector and  $\mathbf{I}_N$  denotes the  $N \times N$  identity matrix. Furthermore,  $\otimes$  represents the Kronecker product and  $\text{diag}(\mathbf{a})$  is a diagonal matrix with the diagonal specified by  $\mathbf{a}$ . In particular,  $M = 1$  refers to the case of Nyquist sampling rate.

Similar to Eq. (6), the received oversampled noise samples at the  $n_r$ -th receive antenna  $\mathbf{y}_{n_r}^n \in \mathbb{C}^{MN \times 1}$  is

$$\mathbf{y}_{n_r}^n = \mathbf{F} (\mathbf{I}_N \otimes \mathbf{u}) \mathbf{w}_{n_r}, \quad (8)$$

where  $\mathbf{F} \in \mathbb{R}^{MN \times MN}$  is the Toeplitz matrix constituted by  $m \left[ \frac{k}{M} \right]$  with the form similar to Eq. (7), and  $\mathbf{w}_{n_r} \sim \mathcal{CN}(0_{N \times 1}, \sigma^2 \mathbf{I}_N)$  represents the complex Gaussian random variables with zero mean and variance  $\sigma^2$ .

Combining Eqs. (6) and (8), the received oversampled samples at the  $n_r$ -th receive antenna  $\mathbf{y}_{n_r} \in \mathbb{C}^{MN \times 1}$  are

$$\mathbf{y}_{n_r} = \mathbf{y}_{n_r}^d + \mathbf{y}_{n_r}^n = \mathbf{Z} (\mathbf{I}_N \otimes \mathbf{u}) \sum_{l=1}^L h_{n_r,l} g_{l,n_r} \text{diag}(\mathbf{s}_l) \mathbf{x}_{n_r} + \mathbf{F} (\mathbf{I}_N \otimes \mathbf{u}) \mathbf{w}_{n_r}. \quad (9)$$

Defining  $\mathbf{Z}'' \triangleq \mathbf{Z} (\mathbf{I}_N \otimes \mathbf{u})$ ,  $\mathbf{F}'' \triangleq \mathbf{F} (\mathbf{I}_N \otimes \mathbf{u})$ , and stacking the  $n_r$ -th received oversampled samples on top of the previous receive antenna, Eq. (9) is extended to

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^T & \dots & \mathbf{y}_{N_r}^T \end{bmatrix}^T = \mathbf{Z}' \text{vec}\{\mathbf{A}^T\} + \mathbf{F}' \text{vec}\{\mathbf{W}^T\}, \quad (10)$$

where  $\mathbf{Z}' \triangleq \mathbf{I}_{N_r} \otimes \mathbf{Z}''$ ,  $\mathbf{F}' \triangleq \mathbf{I}_{N_r} \otimes \mathbf{F}''$  and  $\mathbf{W} = [\mathbf{w}_1^T; \dots; \mathbf{w}_{N_r}^T] \in \mathbb{C}^{N_r \times N}$ . Consider all the users  $\mathbf{A} \triangleq \mathbf{H} (\mathbf{S} \odot (\mathbf{G}\mathbf{X})) \in \mathbb{C}^{N_r \times N}$ , where  $\mathbf{G} \in \mathbb{C}^{L \times N_r}$  and  $\mathbf{H} \in \mathbb{C}^{N_r \times L}$  are the channel matrices from the users to the RIS and from the RIS to the BS, respectively;  $\mathbf{X} = [\mathbf{x}_1^T; \dots; \mathbf{x}_{N_r}^T] \in \mathbb{C}^{N_r \times N}$  and  $\mathbf{S} = [\mathbf{s}_1^T; \dots; \mathbf{s}_L^T] \in \mathbb{C}^{L \times N}$ .  $\text{vec}\{\mathbf{A}\}$  denotes the operation of vectorization by stacking the columns of  $\mathbf{A}$  sequentially on top of one another.

Let  $\mathcal{Q}(\cdot)$  represent the 1-bit quantization function. The resulting quantized signal  $\mathbf{y}_Q$  is given by

$$\mathbf{y}_Q = \mathcal{Q}(\mathbf{y}) = \frac{1}{\sqrt{2}} \left( \text{sign}(\mathbf{y}^{\Re}) + j \text{sign}(\mathbf{y}^{\Im}) \right), \quad (11)$$

where  $\text{sign}(\cdot)$  is the signum function, and  $j = \sqrt{-1}$ .  $(\cdot)^{\Re}$  and  $(\cdot)^{\Im}$  denote the real and imaginary parts, respectively.

The problem to be solved is estimating the channel parameters in  $\mathbf{G}$  and  $\mathbf{H}$ . The minimum mean square error (MMSE) estimators of the channel matrices are given by

$$\min_{\hat{\mathbf{G}}} E \left\{ \left\| \mathbf{G} - \hat{\mathbf{G}} \right\|_F^2 \right\} \quad \text{and} \quad \min_{\hat{\mathbf{H}}} E \left\{ \left\| \mathbf{H} - \hat{\mathbf{H}} \right\|_F^2 \right\}. \quad (12)$$

Based on the Bayes' rule, the closed-form solutions of Eq. (12) are given by  $E\{\mathbf{G}|\mathbf{y}_Q\}$  and  $E\{\mathbf{H}|\mathbf{y}_Q\}$ . For avoiding the high computational complexity of calculating  $p(\mathbf{G}|\mathbf{y}_Q)$  and  $p(\mathbf{H}|\mathbf{y}_Q)$  in the marginalization of  $p(\mathbf{G}, \mathbf{H}|\mathbf{y}_Q)$ , a practical message-passing based algorithm is presented in the next section.

## 3 Markovian Cascaded Channel Estimation

### 3.1 Markov Chain for Oversampling

Due to the banded structure of the matrix  $\mathbf{Z}$  in Eq. (7), we develop a Markov chain, where every two adjacent columns in  $\mathbf{A}$  are combined together for defining one state  $\mathbf{b}_n$  as

$$\mathbf{b}_1 = \text{vec} \left\{ \begin{bmatrix} 0_{N_r \times 1} & \mathbf{a}_1 \end{bmatrix}^T \right\}, \dots, \mathbf{b}_N = \text{vec} \left\{ \begin{bmatrix} \mathbf{a}_{N-1} & \mathbf{a}_N \end{bmatrix}^T \right\}, \quad (13)$$

and each state is with the mean and covariance:

$$\begin{aligned}\hat{\mathbf{b}}_n &= \text{vec}\left\{\left[\hat{\mathbf{a}}_{n-1}, \hat{\mathbf{a}}_n\right]^T\right\} \text{ and} \\ \mathbf{V}_n^b &= \text{diag}\left(\text{vec}\left\{\left[\mathbf{v}_{n-1}^a, \mathbf{v}_n^a\right]^T\right\}\right),\end{aligned}\quad (14)$$

with  $\mathbf{a}_n \in \mathbb{C}^{N_r \times 1}$  being the  $n$ -th column of  $\mathbf{A}$ . The transition function from the current state to the next is

$$\mathbf{b}_{n+1} = \mathbf{T}\mathbf{b}_n + \text{vec}\left\{\left[0_{N_r \times 1}, \mathbf{a}_{n+1}\right]^T\right\}, \quad (15)$$

where  $\mathbf{T} = \mathbf{I}_{N_r} \otimes [0, 1; 0, 0]$  and the conditional probability is given by

$$\begin{aligned}p(\mathbf{b}_{n+1} | \mathbf{b}_n, \mathbf{a}_{n+1}) &= \\ \mathcal{CN}(\mathbf{b}_{n+1}; \mathbf{T}\hat{\mathbf{b}}_n + \text{vec}\left\{\left[0_{N_r \times 1}, \hat{\mathbf{a}}_{n+1}\right]^T\right\}, \\ \text{diag}\left(\text{vec}\left\{\left[0_{N_r \times 1}, \mathbf{v}_{n+1}^a\right]^T\right\}\right).\end{aligned}\quad (16)$$

The system model of Eq. (11) can be decomposed by using the Bussgang theorem<sup>[15]</sup>:

$$\mathbf{y}_\mathcal{Q} = \mathbf{K}\mathbf{Z}'\text{vec}\{\mathbf{A}^T\} + \mathbf{w}' \quad \text{with} \quad \mathbf{K} = \sqrt{\frac{2}{\pi}} \text{diag}(\mathbf{C}_y)^{\frac{1}{2}}, \quad (17)$$

where  $\mathbf{w}' \triangleq \mathbf{K}\mathbf{F}'\text{vec}\{\mathbf{W}\} + \mathbf{n}_q$  is the equivalent noise containing the filtered noise and the quantization noise, assumed to follow the Gaussian distribution with zero mean and covariance  $\mathbf{V}^{\mathbf{w}'} = \sigma_n^2 \mathbf{K}\mathbf{F}\mathbf{F}'\mathbf{K}' + \mathbf{C}_{n_q}$ . And  $\text{diag}(\mathbf{C}_y)$  is the diagonal matrix with the diagonal specified by the diagonal of  $\mathbf{C}_y$ . The received quantized signal  $\mathbf{y}_{\mathcal{Q}_n}$  at the state  $\mathbf{b}_n$  is then calculated as

$$\mathbf{y}_{\mathcal{Q}_n} = \mathbf{D}\text{vec}\left\{\left[\mathbf{a}_{n-1}, \mathbf{a}_n\right]^T\right\} + \mathbf{w}''_n = \mathbf{D}\mathbf{b}_n + \mathbf{w}''_n, \quad (18)$$

where  $\mathbf{D} \triangleq (\mathbf{I}_{N_r} \otimes [0_{M \times M} \quad \mathbf{I}_M])\mathbf{K}_n\mathbf{Z}'_n$ ;  $\mathbf{K}_n \in \mathbb{C}^{2N_r M \times 2N_r M}$ ,  $\mathbf{Z}'_n \in \mathbb{C}^{2N_r M \times 2N_r}$  and  $\mathbf{w}''_n \in \mathbb{C}^{2N_r M \times 1}$  represent the corresponding submatrices of  $\mathbf{K}$ ,  $\mathbf{Z}'$  and  $\mathbf{w}'$ , respectively. The prior probability  $p(\mathbf{y}_{\mathcal{Q}_n} | \mathbf{b}_n)$  is given by

$$p(\mathbf{y}_{\mathcal{Q}_n} | \mathbf{b}_n) = \mathcal{CN}(\mathbf{y}_{\mathcal{Q}_n}; \mathbf{D}\hat{\mathbf{b}}_n, \mathbf{V}_n^{\mathbf{w}''}), \quad (19)$$

where  $\mathbf{V}_n^{\mathbf{w}''}$  is the covariance of  $\mathbf{w}''_n$  in Eq. (18).

### 3.2 Factor Graph Representation

With Eqs. (16) and (19), the posterior probability  $p(\mathbf{b}, \mathbf{A} | \mathbf{y}_\mathcal{Q})$  is calculated as

$$p(\mathbf{b}, \mathbf{A} | \mathbf{y}_\mathcal{Q}) = \frac{1}{p(\mathbf{y}_\mathcal{Q})} \prod_{n=1}^N p(\mathbf{y}_{\mathcal{Q}_n} | \mathbf{b}_n) p(\mathbf{b}_n | \mathbf{b}_{n-1}, \mathbf{a}_n) \quad (20)$$

Defining  $\mathbf{C} \triangleq (\mathbf{S} \odot \mathbf{G}\mathbf{X}) \in \mathbb{C}^{L \times N}$  in Eq. (10), the joint posterior probability can be further factored as

$$\begin{aligned}p(\mathbf{b}, \mathbf{A}, \mathbf{G}, \mathbf{H}, \mathbf{C} | \mathbf{y}_\mathcal{Q}) &= \\ \frac{1}{p(\mathbf{y}_\mathcal{Q})} p(\mathbf{y}_\mathcal{Q} | \mathbf{b}) p(\mathbf{b} | \mathbf{A}) p(\mathbf{A} | \mathbf{H}, \mathbf{C}) p(\mathbf{C} | \mathbf{G}) p(\mathbf{H}) p(\mathbf{G}) &= \\ \frac{1}{p(\mathbf{y}_\mathcal{Q})} \left( \prod_{n=1}^N p(\mathbf{y}_{\mathcal{Q}_n} | \mathbf{b}_n) p(\mathbf{b}_n | \mathbf{b}_{n-1}, \mathbf{a}_n) \right) \left( \prod_{n_r=1}^{N_r} \prod_{n_l=1}^{N_l} p(a_{n_r, n_l} | \mathbf{h}_{n_r}, \mathbf{c}_{n_l}) \right) \\ \left( \prod_{l=1}^L \prod_{n=1}^N p(c_{l, n} | \mathbf{g}_l) \right) \left( \prod_{n_r=1}^{N_r} \prod_{l=1}^L p(h_{n_r, l}) \right) \left( \prod_{l=1}^L \prod_{n_r=1}^{N_r} p(\mathbf{g}_{l, n_r}) \right),\end{aligned}\quad (21)$$

where the first bracket is from the Markov chain and the rests belong to the bilinear structure described in Ref. [13]. Some involved probabilities are

$$\begin{aligned}p(a_{n_r, n_l} | \mathbf{h}_{n_r}, \mathbf{c}_{n_l}) &= \delta(a_{n_r, n_l} - \mathbf{h}_{n_r}^T \mathbf{c}_{n_l}) \quad \text{and} \\ p(c_{l, n} | \mathbf{g}_l) &= \delta(c_{l, n} - s_{l, n} \mathbf{g}_l^T \mathbf{x}_n),\end{aligned}\quad (22)$$

where  $\mathbf{h}_{n_r} \in \mathbb{C}^{L \times 1}$  and  $\mathbf{g}_l \in \mathbb{C}^{N_r \times 1}$  are the  $n_r$ -th and  $l$ -th row of  $\mathbf{H}$  and  $\mathbf{G}$ , respectively;  $\mathbf{c}_{n_l} \in \mathbb{C}^{L \times 1}$  is the  $n$ -th column of  $\mathbf{C}$ . The second formula in Eq. (22) comes from the definition of  $\mathbf{C}$ , where  $c_{l, n}$  and  $s_{l, n}$  are the  $(l, n)$ -th element of  $\mathbf{C}$  and  $\mathbf{S}$ , respectively, and  $\mathbf{x}_n$  is the  $n$ -th column of  $\mathbf{X}$ . The factor graph representation of Eq. (21) is shown in Fig. 2, where the hollow circles and the solid squares represent the variable nodes and the factor nodes, respectively. The message passing in the first part will be described in the next subsection, while the illustration of the second part can be found in Ref. [8].

### 3.3 Message Passing in Markov Chain

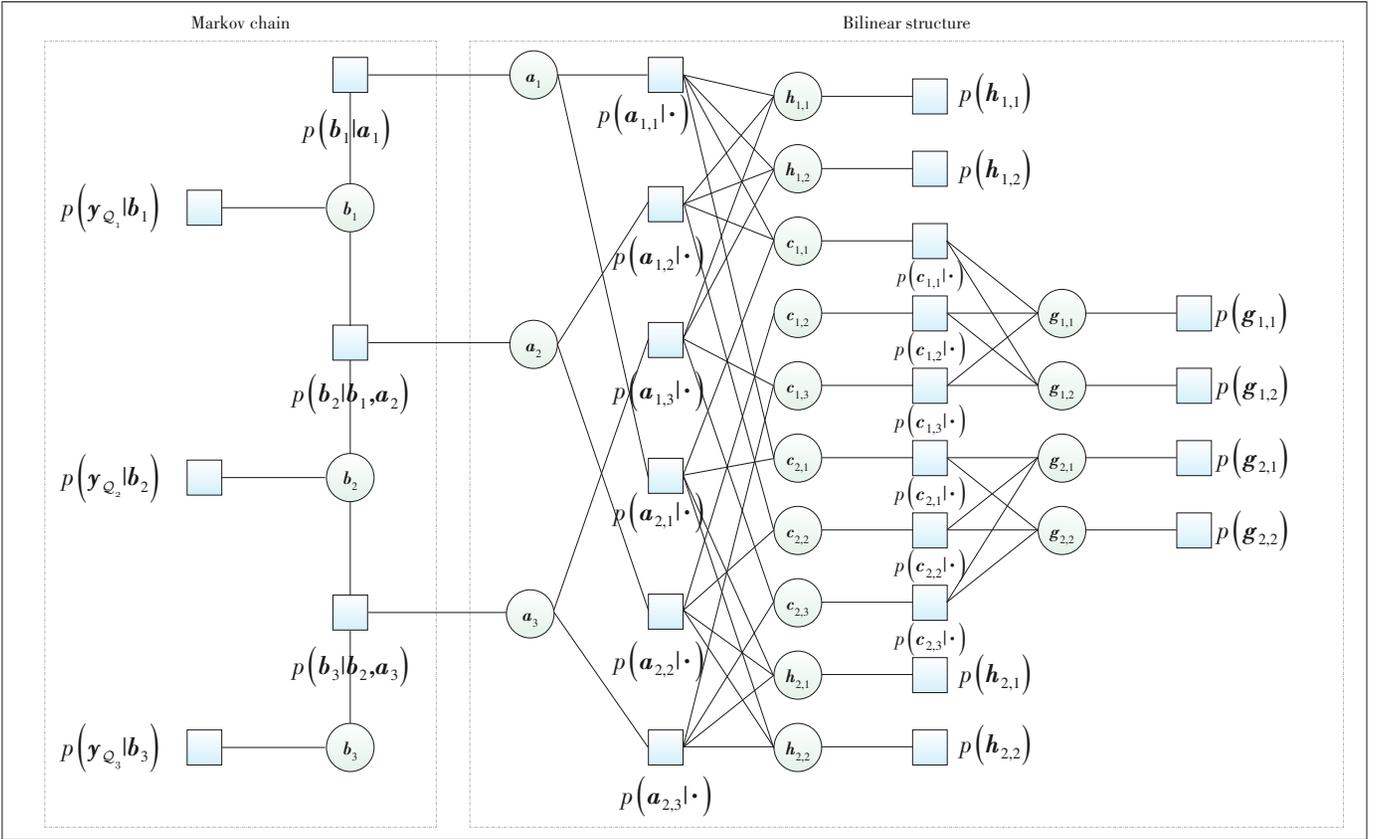
In the following, the approximate message passing in the Markov chain is derived according to the sum-product rule. The linear operator  $\mathbf{K}$  described in Eq. (17) involves the covariance of unquantized signal  $\mathbf{y}$  in Eq. (10), calculated as

$$\mathbf{C}_y = \mathbf{Z}' \text{diag}\left(\text{vec}\{\mathbf{V}^{\mathbf{A}^T}\}\right) \mathbf{Z}'^H + \sigma^2 \mathbf{F}' \mathbf{F}'^H. \quad (23)$$

During the message updates, the covariance of  $\mathbf{A}$  at the  $i$ -th iteration is  $\mathbf{V}^{\mathbf{A}}(i)$ , and Eq. (23) can be rewritten as

$$\mathbf{C}_y(i) = \mathbf{Z}' \text{diag}\left(\text{vec}\{\mathbf{V}^{\mathbf{A}^T}(i)\}\right) \mathbf{Z}'^H + \sigma^2 \mathbf{F}' \mathbf{F}'^H. \quad (24)$$

The covariance of the quantization noise at the  $i$ -th iteration is



▲ Figure 2. Factor graph representation of Eq. (21) for an example with  $M = N_r = N_t = L = 2$  and  $N = 3$

$$\mathbf{C}_{n_q}(i) = \mathbf{C}_{Q_i}(i) - \mathbf{K}(i) \mathbf{C}_y(i) \mathbf{K}(i)^H, \quad (25)$$

where  $\mathbf{C}_{Q_i}(i)$  is calculated from Ref. [16] as

$$\mathbf{C}_{Q_i}(i) = \frac{2}{\pi} (\arcsin(\mathbf{K}(i) \mathbf{C}_y^{\Re}(i) \mathbf{K}(i)^H) + j \arcsin(\mathbf{K}(i) \mathbf{C}_y^{\Im}(i) \mathbf{K}(i)^H)). \quad (26)$$

### 3.3.1 Downward AMP

With Eq. (19), the message  $m_{b_n \rightarrow p(b_{n+1}|b_n, a_{n+1})}$  at the  $i$ -th iteration can be approximated as Gaussian distribution with the following mean and variance:

$$\hat{\mathbf{b}}_n^{Down}(i) = \widehat{\mathbf{b}}_n^{Down}(i) + \mathbf{V}'_n^{Down}(i) \mathbf{D}^H \mathbf{R} (\mathbf{y}_{Q_n} - \mathbf{D} \widehat{\mathbf{b}}_n^{Down}(i)), \quad (27)$$

$$\mathbf{V}_n^{Down}(i) = \mathbf{V}'_n^{Down}(i) - \mathbf{V}'_n^{Down}(i) \mathbf{D}^H \mathbf{R} \mathbf{D} \mathbf{V}'_n^{Down}(i), \quad (28)$$

where  $\mathbf{R} \triangleq (\mathbf{V}_n^{w*} + \mathbf{D} \mathbf{V}'_n^{Down}(i) \mathbf{D}^H)^{-1}$ . Moreover, with Eq. (15), the message from the factor node  $p(b_{n+1}|b_n, a_{n+1})$  to the variable node  $b_{n+1}$  at the  $i$ -th iteration is given by

$$m_{p(b_{n+1}|b_n, a_{n+1}) \rightarrow b_{n+1}} = \mathcal{CN}(b_{n+1}; \widehat{\mathbf{b}}_{n+1}^{Down}(i), \mathbf{V}'_{n+1}^{Down}(i)), \quad (29)$$

$$\widehat{\mathbf{b}}_{n+1}^{Down}(i) = \mathbf{T} \widehat{\mathbf{b}}_n^{Down}(i) + \text{vec} \left\{ \left[ \mathbf{0}_{N_r \times 1}, \widehat{\mathbf{a}}_{n+1}(i) \right]^T \right\}, \quad (30)$$

$$\mathbf{V}'_{n+1}^{Down}(i) = \mathbf{T} \mathbf{V}'_n^{Down}(i) \mathbf{T}^H + \text{diag} \left\{ \left[ \mathbf{0}_{N_r \times 1}, \mathbf{v}_{n+1}^a(i) \right]^T \right\}. \quad (31)$$

Especially, when  $n = 1$ , Eqs. (30) and (31) are reduced to

$$\widehat{\mathbf{b}}_1^{Down}(i) = \hat{\mathbf{b}}_1(i) \quad \text{and} \quad \mathbf{V}'_1^{Down}(i) = \mathbf{V}_1^b(i). \quad (32)$$

### 3.3.2 Upward AMP

Similar to Eqs. (27) and (28), the message  $m_{b_n \rightarrow p(b_n|b_{n-1}, a_n)}$  at the  $i$ -th iteration is approximated as Gaussian distribution with the following mean and variance

$$\hat{\mathbf{b}}_n^{Up}(i) = \widehat{\mathbf{b}}_n^{Up}(i) + \mathbf{V}'_n^{Up}(i) \mathbf{D}^H \mathbf{R} (\mathbf{y}_{Q_n} - \mathbf{D} \widehat{\mathbf{b}}_n^{Up}(i)), \quad (33)$$

$$\mathbf{V}_n^{Up}(i) = \mathbf{V}'_n^{Up}(i) - \mathbf{V}'_n^{Up}(i) \mathbf{D}^H \mathbf{R} \mathbf{D} \mathbf{V}'_n^{Up}(i), \quad (34)$$

where  $\mathbf{R} \triangleq (\mathbf{V}_n^{w*} + \mathbf{D} \mathbf{V}'_n^{Up}(i) \mathbf{D}^H)^{-1}$ . Moreover, with the inverse of Eq. (15), the message from the factor node  $p(b_n|b_{n-1}, a_n)$  to the variable node  $b_{n-1}$  at the  $i$ -th iteration is given by

$$m_{p(b_n|b_{n-1}, a_n) \rightarrow b_{n-1}} = \mathcal{CN}(b_{n-1}; \widehat{\mathbf{b}}_{n-1}^{Up}(i), \mathbf{V}_{n-1}^{Up}(i)), \quad (35)$$

where

$$\widehat{\mathbf{b}}_{n-1}^{Up}(i) = \mathbf{T}^\dagger \widehat{\mathbf{b}}_n^{Up}(i) + \text{vec} \left\{ \left[ \widehat{\mathbf{a}}_{n-2}(i), \mathbf{0}_{N_r \times 1} \right]^T \right\}, \quad (36)$$

$$\mathbf{V}_{n-1}^{Up}(i) = \mathbf{T}^\dagger \mathbf{V}_n^{Up}(i) (\mathbf{T}^H)^\dagger + \text{diag} \left\{ \text{vec} \left\{ \left[ \mathbf{v}_{n-2}^a(i), \mathbf{0}_{N_r \times 1} \right]^T \right\} \right\}, \quad (37)$$

where  $(\cdot)^\dagger$  denotes the operation of pseudo-inverse. Especially, when  $n = N$ , Eqs. (36) and (37) are reduced to

$$\widehat{\mathbf{b}}_N^{Up}(i) = \widehat{\mathbf{b}}_N(i) \quad \text{and} \quad \mathbf{V}_N^{Up}(i) = \mathbf{V}_N^b(i). \quad (38)$$

### 3.3.3 Backward AMP

Combining with  $m_{b_{n-1} \rightarrow p(b_n|b_{n-1}, a_n)}$  and  $m_{b_n \rightarrow p(b_n|b_{n-1}, a_n)}$ , the message from the factor node  $p(b_n|b_{n-1}, a_n)$  back to the variable node  $a_n$  is

$$m_{p(b_n|b_{n-1}, a_n) \rightarrow a_n} = \int_{b_{n-1}} \mathcal{CN}(b_n; \mathbf{T} \widehat{\mathbf{b}}_{n-1}^{Down}(i), \mathbf{T} \mathbf{V}_{n-1}^{Down}(i) \mathbf{T}^H) \mathcal{CN}(b_n; \widehat{\mathbf{b}}_n^{Up}(i), \mathbf{V}_n^{Up}(i)), \quad (39)$$

where the multiplication of two Gaussian functions is another Gaussian function<sup>3</sup>.

## 4 Further Discussions

The overall Markovian cascaded channel estimation algorithm is presented in Algorithm 1. In lines 3 – 6, the estimates of the each mean and variance in the matrix product  $\mathbf{A} = \mathbf{H}\mathbf{C}$  are calculated. The message passing in the Markov chain is illustrated in lines 7 – 9. Subsequently, the residual  $q_{n,r,n}$  and the inverse residual variance  $u_{n,r,n}$  are calculated in lines 10 – 11. In lines 12 – 13, these residual terms are used to compute  $v_{n,r,l}^h$  and  $\widehat{h}_{n,r,l}^h$ , which can be interpreted as an observation of  $\widehat{h}_{n,r,l}^h$  under an AWGN channel with zero mean and a variance of  $v_{n,r,l}^h$ . The posterior means and variances of each elements in  $\mathbf{H}$  are estimated in lines 14 and 15. Similarly, the posterior means and variances of each elements in the auxiliary matrix  $\mathbf{C}$  are estimated in lines 16 – 21; the same is performed for  $\mathbf{G}$  in lines 22 – 27.

In this paper, a damping method is applied to improve the robustness of the proposed algorithm. Specifically, in each iteration the posterior means and variances of  $\mathbf{H}$  (lines 14 – 15),

$\mathbf{C}$  (lines 20 – 21) and  $\mathbf{G}$  (lines 26 – 27) are updated by using a linear combination of the current and previous updates. For example, the updates of the posterior mean  $\widehat{h}_{n,r,l}^h$  and variance  $v_{n,r,l}^h$  in lines 14 – 15 of Algorithm 1 are replaced by

$$\widehat{h}_{n,r,l}^h(i+1) = (1 - \beta) \widehat{h}_{n,r,l}^h(i) + \beta \widehat{h}_{n,r,l}^h(i+1), \quad (40)$$

$$v_{n,r,l}^h(i+1) = (1 - \beta) v_{n,r,l}^h(i) + \beta v_{n,r,l}^h(i+1), \quad (41)$$

where  $\beta \in [0, 1]$  is the damping factor. In our work,  $\beta$  is chosen within  $[0.2, 0.4]$ .

We now give a brief discussion on the computational complexity of the proposed algorithm. Note that the total algorithm is separated into two parts, the Markov chain and the bilinear structure. We thus sketch the respective complexity as follows. First, the complexity of the Markov chain is dominated by basic matrix multiplications in Eqs. (27), (28), (33), (34) and (39), requiring  $\mathcal{O}((MN_r)^3)$  flops per iteration. Second, the complexity of the bilinear structure is dominated by the calculations in Lines 3 – 5, 10 – 11, 12 – 13, 16 – 17 and 22 – 25, requiring  $\mathcal{O}((N_r + N_t)LN)$  flops per iteration. Finally, the complexity of the proposed algorithm is  $I_{\max} \mathcal{O}((MN_r)^3 + (N_r + N_t)LN)$ , where  $I_{\max}$  is the maximum number of the iterations.

### Algorithm 1. Markovian cascaded channel estimation

**Input:**  $\mathbf{y}, \mathbf{x}, \mathbf{S}, \sigma^2$ , prior distributions  $p(\mathbf{G}), p(\mathbf{H})$

**Output:**  $\widehat{\mathbf{C}}, \widehat{\mathbf{H}}$

1: **Initialization:**  $\forall n, l, n: \widehat{h}_{n,r,l}^h(1), v_{l,n}^c(1), v_{n,r,l}^h(1), \widehat{c}_{l,n}(1),$

$$\eta_{l,n}(0) = 0, u_{n,r,n}(0) = 0;$$

2: **for**  $i=1, \dots, I_{\max}$  **do**

3:  $\forall n, r, n: v_{n,r,n}^p(i) = \sum_l \left| \widehat{h}_{n,r,l}^h(i) \right|^2 v_{l,n}^c(i) + v_{n,r,l}^h(i) \left| \widehat{c}_{l,n}(i) \right|^2$

4:  $\forall n, r, n: \widehat{p}_{n,r,n}^p(i) = \sum_l \widehat{h}_{n,r,l}^h(i) \widehat{c}_{l,n}(i)$

5:  $\forall n, r, n: v_{n,r,n}^p(i) = v_{n,r,n}^p(i) + \sum_l v_{n,r,l}^h(i) v_{l,n}^c(i)$

6:  $\forall n, r, n: \widehat{p}_{n,r,n}^p(i) = \widehat{p}_{n,r,n}^p(i) - u_{n,r,n}(i-1) v_{n,r,n}^p(i)$

7: Calculate the adaptive linear operator and quantization noise via Eqs. (24) and (25) using  $v_{n,r,n}^p(i)$  and  $\widehat{p}_{n,r,n}^p(i)$

8: Calculate the downward and upward messages at each check node in the Markov chain via Eqs. (27), (28), (33), and (34)

9: Calculate the backward messages via Eq. (39) and ob-

3.  $\mathcal{CN}(\mathbf{x}; \mathbf{m}_1, \mathbf{V}_1) \mathcal{CN}(\mathbf{x}; \mathbf{m}_2, \mathbf{V}_2) \propto \mathcal{CN}(\mathbf{x}; \mathbf{m}, \mathbf{V})$ , where  $\mathbf{V} = (\mathbf{V}_1^{-1} + \mathbf{V}_2^{-1})^{-1}$  and  $\mathbf{m} = \mathbf{V}(\mathbf{V}_1^{-1} \mathbf{m}_1 + \mathbf{V}_2^{-1} \mathbf{m}_2)$ .

tain  $v_{n_r,n}^a(i)$  and  $\hat{a}_{n_r,n}(i)$

$$10: \quad \forall n_r, n: q_{n_r,n}(i) = (1 - \frac{v_{n_r,n}^a(i)}{v_{n_r,n}^p(i)}) / v_{n_r,n}^p(i)$$

$$11: \quad \forall n_r, n: u_{n_r,n}(i) = (\hat{a}_{n_r,n}(i) - \hat{p}_{n_r,n}(i)) / v_{n_r,n}^p(i)$$

**Update H:**

$$12: \quad \forall n_r, l: v_{n_r,l}^{h'}(i) = 1 / (\sum_n q_{n_r,n}(i) |\hat{c}_{l,n}(i)|^2)$$

$$13: \quad \forall n_r, l: \hat{h}'_{n_r,l}(i) = \hat{h}_{n_r,l}(i) (1 - v_{n_r,l}^{h'}(i) \sum_n q_{n_r,n}(i) v_{l,n}^c(i)) + v_{n_r,l}^{h'}(i) \sum_n u_{n_r,n}(i) \hat{c}_{l,n}^*(i)$$

$$14: \quad \forall n_r, l: \hat{h}'_{n_r,l}(i+1) = E \{ h_{n_r,l} | \hat{h}'_{n_r,l}(i), v_{n_r,l}^{h'}(i) \}$$

$$15: \quad \forall n_r, l: v_{n_r,l}^{h'}(i+1) = Var \{ h_{n_r,l} | \hat{h}'_{n_r,l}(i), v_{n_r,l}^{h'}(i) \}$$

**Update C:**

$$16: \quad \forall l, n: v_{l,n}^{c'}(i) = s_{l,n} / (\sum_{n_r} q_{n_r,n}(i) |\hat{h}_{n_r,l}(i)|^2)$$

$$17: \quad \forall l, n: \hat{c}'_{l,n}(i) = s_{l,n} (\hat{c}_{l,n}(i) (1 - v_{l,n}^{c'}(i) \sum_{n_r} q_{n_r,n}(i) v_{n_r,l}^h(i)) + v_{l,n}^{c'}(i) \sum_{n_r} u_{n_r,n}(i) \hat{h}_{n_r,l}^*(i))$$

$$18: \quad \forall l, n: v_{l,n}^{c''}(i) = \sum_{n_1} v_{l,n_1}^g(i) |x_{n_1,n}|^2$$

$$19: \quad \forall l, n: \hat{c}''_{l,n}(i) = \sum_{n_1} \hat{g}_{l,n_1}(i) x_{n_1,n} - \eta_{l,n}(i-1) \sum_{n_1} v_{l,n_1}^g(i) |x_{n_1,n}|^2$$

$$20: \quad \forall l, n: \hat{c}_{l,n}(i) = E \{ c_{l,n} | \hat{c}'_{l,n}(i), v_{l,n}^{c'}(i), \hat{c}''_{l,n}(i), v_{l,n}^{c''}(i) \}$$

$$21: \quad \forall l, n: v_{l,n}^c(i) = Var \{ c_{l,n} | \hat{c}'_{l,n}(i), v_{l,n}^{c'}(i), \hat{c}''_{l,n}(i), v_{l,n}^{c''}(i) \}$$

**Update G:**

$$22: \quad \forall l, n: \eta_{l,n}(i) = s_{l,n} (\hat{c}_{l,n}(i) - \hat{c}''_{l,n}(i)) / v_{l,n}^{c''}(i)$$

$$23: \quad \forall l, n: v_{l,n}^\eta(i) = s_{l,n} (1 - v_{l,n}^c(i) / v_{l,n}^{c''}(i)) / v_{l,n}^{c''}(i)$$

$$24: \quad \forall l, n_i: v_{l,n_i}^g(i) = 1 / (\sum_n v_{l,n}^\eta(i) |x_{n,n}|^2)$$

$$25: \quad \forall l, n_i: \hat{g}'_{l,n_i}(i) = \hat{g}_{l,n_i}(i) + v_{l,n_i}^g(i) \sum_n \eta_{l,n}(i) x_{n,n}^*$$

$$26: \quad \forall l, n_i: \hat{g}_{l,n_i}(i+1) = E \{ g_{l,n_i} | \hat{g}'_{l,n_i}(i), v_{l,n_i}^g(i) \}$$

$$27: \quad \forall l, n_i: v_{l,n_i}^g(i+1) = Var \{ g_{l,n_i} | \hat{g}'_{l,n_i}(i), v_{l,n_i}^g(i) \}$$

28: **if** stopping criterion is met **then**

29: **break**

30: **end if**

31: **end for**

(RRC) filters with a roll-off factor of 0.8. The channel is assumed to experience Rayleigh block fading. The simulation results presented here are obtained by averaging over 100 independent realizations of the channel matrices, noise and pilots.

The signal-to-noise ratio (SNR) is defined as  $10 \log \left( \frac{\rho L N_1}{\sigma^2} \right)$ ,

where  $\rho$  is the sparsity level of  $\mathbf{S}^4$  and is set as 0.3. The pilot length is 200. For the correlated channel, the channel matrices are modeled as

$$\mathbf{H} = \mathbf{R}_r^{\frac{1}{2}} \mathbf{H}' \mathbf{R}_r^{\frac{1}{2}} \quad \text{and} \quad \mathbf{G} = \mathbf{G}' \mathbf{R}_d^{\frac{1}{2}}, \quad (42)$$

where the elements of  $\mathbf{H}'$  and  $\mathbf{G}'$  are independent and identically distributed (i. i. d.) complex Gaussian random variables with zero mean and unit variance. The matrix  $\mathbf{R}_r$  denotes the receive correlation matrix with the following form:

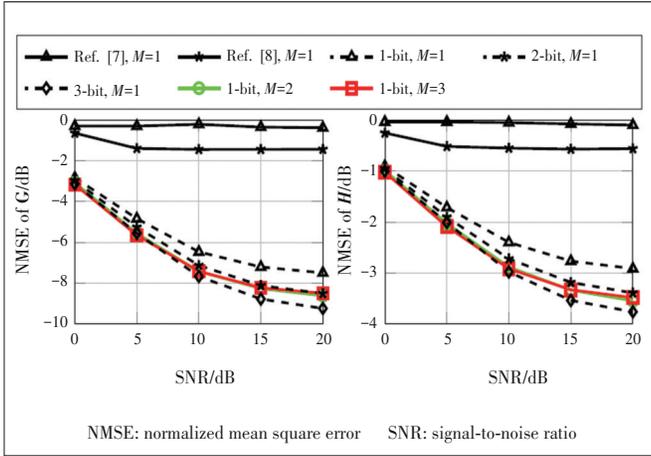
$$\mathbf{R}_r = \begin{bmatrix} 1 & \rho & \cdots & \rho^{(N_r-1)} \\ \rho^* & 1 & \cdots & \rho^{(N_r-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{*(N_r-1)} & \rho^{*(N_r-2)} & \cdots & 1 \end{bmatrix}, \quad (43)$$

where  $\rho$  is the correlation index of neighboring antennas at the BS and set as  $0.4 + 0.3j$ . ( $|\rho| = 0$  represents an uncorrelated scenario and  $|\rho| = 1$  implies a fully correlated scenario.) The matrices  $\mathbf{R}_r$  and  $\mathbf{R}_d$  have the same form as Eq. (43), where  $\rho$  is set as  $0.2 + 0.5j$  and  $0.1 + 0.2j$  at the RIS, respectively.

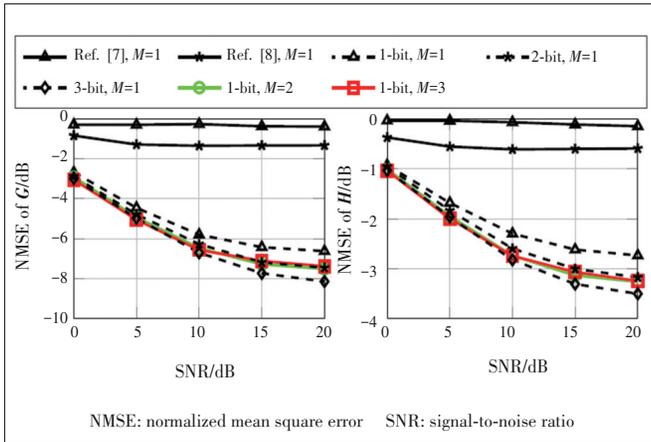
Figs. 3 and 4 show the NMSEs of  $\mathbf{G}$  and  $\mathbf{H}$  under uncorrelated and correlated channels, respectively. For the multi-bit systems sampling at the Nyquist rate, the calculation of the posterior probability  $p(\mathbf{A} | \mathbf{y}_q)$  is referred to as Eqs. (23) and (24) in Ref. [17]. From the figures, it can be seen that our proposed 1-bit multi-fold oversampled systems outperform 1-bit systems sampling at the Nyquist rate, and can even approach the performance of 2-bit systems sampling at the Nyquist rate. Another observation is that after 2-fold oversampling, the performance of 1-bit systems goes into the saturation field, and no further gain can be achieved. This reveals the performance limit of the proposed method. Furthermore, we have also compared the performance of Refs. [7] and [8] as references, where the resolutions of the ADCs are changed to 1-bit. From the results, the performance of the references is worse than that of our proposed method, since they do not consider the impact of 1-bit quantization and the advantages of oversampling.

The advantage of 1-bit ADCs is that they do not require au-

4. In simulations, the phases of the RIS are set to zeros, and the matrix  $\mathbf{S}$  is generated as a 0-1 random matrix.



▲ Figure 3. NMSE of  $G$  and  $H$  comparisons between the multi-bit systems with Nyquist rate ( $M=1$ ) and the 1-bit systems with multi-fold oversampling under uncorrelated channel



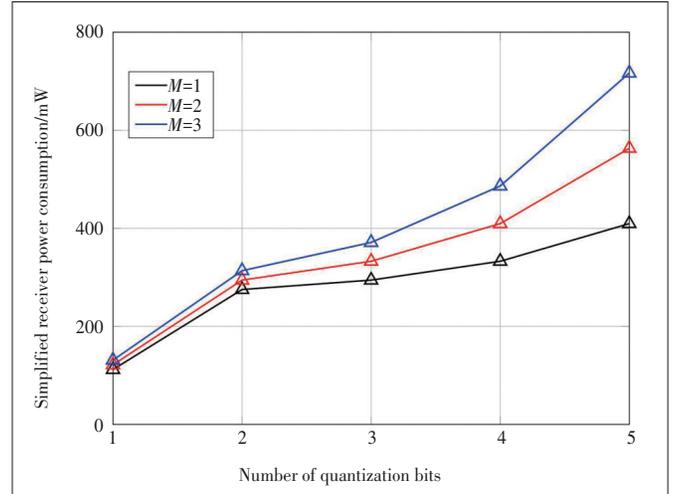
▲ Figure 4. NMSE of  $G$  and  $H$  comparisons between the multi-bit systems with Nyquist rate ( $M=1$ ) and the 1-bit systems with multi-fold oversampling under correlated channel

tomatic gain control (AGC), and can be replaced by simple LAs. Fig. 5 shows the simplified<sup>5</sup> receiver power consumption as a function of the quantization bits and the oversampling factor  $M$ , which is calculated as

$$P_{\text{simplified}} = 2N_r (cP_{\text{AGC}} + (1-c)P_{\text{LA}} + \text{FOM} \times Mf_{\text{Nyquist}} 2^q), \quad (44)$$

where  $P_{\text{AGC}}$  and  $P_{\text{LA}}$  denote the power consumption of AGC and LA, respectively;  $q$  is the quantization bits and  $f_{\text{Nyquist}}$  is the Nyquist-sampling rate;  $c$  is chosen as 0 for 1-bit system and 1 for systems with multi-bit. Numerical parameters<sup>[18]</sup> are  $P_{\text{AGC}} = 2$  mW,  $P_{\text{LA}} = 0.8$  mW,  $f_{\text{Nyquist}} = 2.5$  GHz and figures-of-merit (FOM) = 15 fJ. From the results, it can be seen that 1-bit systems consume much less power than multi-bit systems with either the Nyquist rate or the oversampling rate.

5. The power consumption of shared components in different systems is neglected.



▲ Figure 5. Simplified power consumption at the receiver as a function of the number of quantization bits

## 6 Conclusions

In this work, we propose RIS aided 1-bit massive MIMO systems with oversampling at the receiver. The aim of oversampling is to compensate for the performance loss due to the coarse quantization. A Markovian cascaded channel estimation algorithm is developed for such systems. Simulation results have shown good performance gains of the proposed oversampled system, which can achieve the same performance of the corresponding 2-bit system sampling at the Nyquist rate while consuming less power at the receiver.

## References

- [1] LARSSON E G, EDFORS O, TUFVESSON F, et al. Massive MIMO for next generation wireless systems [J]. IEEE communications magazine, 2014, 52(2): 186 - 195. DOI: 10.1109/MCOM.2014.6736761
- [2] LIASKOS C, NIE S, TSIOLIARIDOU A, et al. A new wireless communication paradigm through software-controlled metasurfaces [J]. IEEE communications magazine, 2018, 56(9): 162 - 169. DOI: 10.1109/MCOM.2018.1700659
- [3] JAMALI V, TULINO A M, FISCHER G, et al. Intelligent surface-aided transmitter architectures for millimeter-wave ultra massive MIMO systems [J]. IEEE open journal of the communications society, 2020, 2: 144 - 167. DOI: 10.1109/OJCOMS.2020.3048063
- [4] TAHA A, ALRABEIAH M, ALKHATEEB A. Enabling large intelligent surfaces with compressive sensing and deep learning [J]. IEEE access, 2021, 9: 44304 - 44321. DOI: 10.1109/ACCESS.2021.3064073
- [5] ARDAH K, GHEREKHLOO S, DE ALMEIDA A L F, et al. TRICE: A channel estimation framework for RIS-aided millimeter-wave MIMO systems [J]. IEEE signal processing letters, 2021, 28: 513 - 517. DOI: 10.1109/

LSP.2021.3059363

- [6] HE Z Q, YUAN X J. Cascaded channel estimation for large intelligent metasurface assisted massive MIMO [J]. *IEEE wireless communications letters*, 2020, 9 (2): 210 – 214. DOI: 10.1109/LWC.2019.2948632
- [7] WANG P L, FANG J, DUAN H P, et al. Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems [J]. *IEEE signal processing letters*, 2020, 27: 905 – 909. DOI: 10.1109/LSP.2020.2998357
- [8] HE Z Q, LIU H, YUAN X J, et al. Semi-blind cascaded channel estimation for reconfigurable intelligent surface aided massive MIMO [EB/OL]. [2021-11-15]. <https://arxiv.org/abs/2101.07315>
- [9] WALDEN R H. Analog-to-digital converter survey and analysis [J]. *IEEE journal on selected areas in communications*, 1999, 17(4): 539 – 550. DOI: 10.1109/49.761034
- [10] MO J H, HEATH R W. Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information [J]. *IEEE transactions on signal processing*, 2015, 63(20): 5498 – 5512. DOI: 10.1109/TSP.2015.2455527
- [11] LI Y Z, TAO C, SECO-GRANADOS G, et al. Channel estimation and performance analysis of one-bit massive MIMO systems [J]. *IEEE transactions on signal processing*, 2017, 65(15): 4075 – 4089. DOI: 10.1109/TSP.2017.2706179
- [12] SHAO Z C, DE LAMARE R C, LANDAU L T N. Iterative detection and decoding for large-scale multiple-antenna systems with 1-bit ADCs [J]. *IEEE wireless communications letters*, 2018, 7(3): 476 – 479. DOI: 10.1109/LWC.2017.2787159
- [13] SHAO Z C, LANDAU L T N, DE LAMARE R C. Channel estimation for large-scale multiple-antenna systems using 1-bit ADCs and oversampling [J]. *IEEE access*, 2020, 8: 85243 – 85256. DOI: 10.1109/ACCESS.2020.2992246
- [14] SHAO Z C, LANDAU L T N, DE LAMARE R C. Dynamic oversampling for 1-bit ADCs in large-scale multiple-antenna systems [J]. *IEEE transactions on communications*, 2021, 69(5): 3423 – 3435. DOI: 10.1109/TCOMM.2021.3059303
- [15] BUSSGANG J J. Crosscorrelation functions of amplitude-distorted Gaussian signals [R]. Cambridge, USA: Massachusetts Institute of Technology, 1952
- [16] JACOVITTI G, NERI A. Estimation of the autocorrelation function of complex Gaussian stationary processes by amplitude clipped signals [J]. *IEEE transactions on information theory*, 1994, 40(1): 239 – 245. DOI: 10.1109/18.272490
- [17] WEN C K, WANG C J, JIN S, et al. Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs [J]. *IEEE transactions on signal processing*, 2016, 64(10): 2541 – 2556. DOI: 10.1109/TSP.2015.2508786
- [18] ROTH K, NOSSEK J A. Achievable rate and energy efficiency of hybrid and digital beamforming receivers with low resolution ADC [J]. *IEEE journal on selected areas in communications*, 2017, 35(9): 2056 – 2068. DOI: 10.1109/JSAC.2017.2720398

## Biographies

**SHAO Zhichao** received the B.S. degree in information engineering from Xidian University, China in 2012, the M.S. degree in electrical engineering from Technische Universität Dresden, Germany in 2016, and the Ph.D. degree in electrical engineering from the Pontifical Catholic University of Rio de Janeiro, Brazil in 2020. Currently, he is a post-doctoral researcher with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China (UESTC). His research interests lie in communications and signal processing.

**YAN Wenjing** received the B.S. degree from Chongqing University, China in 2018, and the M.S. degree from University of Electronic Science and Technology of China (UESTC) in 2021. She is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology (HKUST), China. Her contribution to the work in this paper was done when she was with UESTC.

**YUAN Xiaojun** (xjyuan@uestc.edu.cn) received the Ph.D. degree in electrical engineering from the City University of Hong Kong, China in 2009. From 2009 to 2011, he was a research fellow with the Department of Electronic Engineering, City University of Hong Kong. He was also a visiting scholar with the Department of Electrical Engineering, University of Hawaii at Manoa, USA in spring and summer 2009, as well as in the same period of 2010. From 2011 to 2014, he was a research assistant professor with the Institute of Network Coding, The Chinese University of Hong Kong, China. From 2014 to 2017, he was an assistant professor with the School of Information Science and Technology, Shanghai Tech University, China. He is currently a State-Specially-Recruited Professor with the Center for Intelligent Networking and Communications, University of Electronic Science and Technology of China. His research interests include signal processing, machine learning, and wireless communications, including but not limited to multi-antenna and cooperative communications, sparse and structured signal recovery, Bayesian approximate inference, and network coding. He has published more than 180 peer-reviewed research articles in the leading international journals and conferences in the related areas. He has served on a number of technical programs for international conferences. He was a co-recipient of the Best Paper Award of IEEE International Conference on Communications (ICC) 2014 and a co-recipient of the Best Journal Paper Award of IEEE Technical Committee on Green Communications and Computing (TCGCC) 2017. He has been an editor of *IEEE Transactions on Communications* since 2017 and *IEEE Transactions on Wireless Communications* since 2018.



# RIS: Spatial-Wideband Effect Analysis and Off-Grid Channel Estimation

JIAN Mengnan<sup>1,2</sup>, ZHANG Nan<sup>1,2</sup>, CHEN Yijian<sup>1,2</sup>

(1. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;

2. Algorithm Department, Wireless Product R&D Institute, ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202201008

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220303.1411.006.html>, published online March 6, 2022

Manuscript received: 2022-01-24

**Abstract:** As a critical candidate technology for 5G-advanced and 6G, reconfigurable intelligent surfaces (RIS) have received extensive attention from academia and industry. RIS has the promising features of passiveness, reconfigurable ability, and low cost. RIS channel estimation faces the challenges of high matrix dimension, passive estimation, and spatial-wideband effect. In this article, we analyze the impact of the spatial-wideband effect on the RIS channel to account for the propagation delay across RIS elements and estimate sparse channel parameters such as angle and gain through a super-resolution compressive sensing (CS) algorithm. The simulation results explore the influence of the spatial-wideband effect on the RIS channel and verify the effectiveness of the proposed algorithm.

**Keywords:** RIS; channel estimation; spatial-wideband effect; compressed sensing

**Citation** (IEEE Format): M. N. Jian, N. Zhang, and Y. J. Chen, "RIS: spatial-wideband effect analysis and off-grid channel estimation," *ZTE Communications*, vol. 20, no. 1, pp. 57 – 62, Mar. 2022. doi: 10.12142/ZTECOM.202201008.

## 1 Introduction

With the acceleration of 5G commercialization, research on 5G-advanced and 6G communication systems is in full swing. The scarcity of wireless spectrum resource has become severe<sup>[1]</sup>, and the increasing network coverage, cost and energy efficiency requirements become a pain point restricting the mobile communication industry. New expediting study for innovative solutions is in urgent need. The reconfigurable intelligent surface (RIS), a cross-fusion of information metamaterials and mobile communications<sup>[2]</sup>, is a new and revolutionary wireless communication enhancement technology. Its basic principle is to construct an intelligent and controllable environment by manipulating the electromagnetic property of metamaterials through digital programming with low-cost hardware architecture and tremendous flexibility.

It is necessary to obtain accurate channel state information (CSI) to support the promising capabilities of the RIS, such as expanding coverage, increasing transmission freedom, and conducting environmental perception and positioning. Different from the massive multiple input multiple output (MIMO), the RIS changes the electromagnetic wave propagation with segmented channels<sup>[3]</sup>. As RIS elements are passive and massive, it becomes challenging to estimate the channel.

Many channel estimation algorithms have been designed for RIS-assisted communications. In Ref. [4], a channel estimation method based on the ON-OFF approach is introduced, splitting the RIS into multiple elements and performing channel estimation separately for each element. In Ref. [5], the

high-frequency scenario is considered, the low-rank property of the RIS channel is exploited, and a matrix filling problem is designed for channel estimation. The hybrid RIS architecture with sensing and signal processing ability is considered in Ref. [6], and the sparseness is utilized to obtain high-resolution channel parameters. To make full use of the collected data, a potential solution is to use the artificial intelligence (AI) method for channel estimation<sup>[7]</sup>. In addition, the position information is helpful for designing a low-complexity channel estimation method to obtain critical information such as the direction of arrival (DOA).

Moreover, as a RIS is usually large, new channel characteristics appear and affect channel estimation accuracy. New algorithms and protocols need to be designed to carry out practical channel estimation and avoid complicated onboard signal processing operations. Most existing works are based on the channel model directly extended from the traditional MIMO channel model<sup>[8]</sup>, and the propagation delay is usually neglected. Different RIS elements receive signals with different path delays when a far-field signal impinges on the RIS. The delay between different elements can reach the same order of magnitude as the symbol interval, and different RIS elements will receive different amplitudes and phases of the identical symbol or even distinct symbols for the same sampling period. This phenomenon is called spatial-wideband effect. The frequency-wideband effect caused by multipath delay spread is considered in recent studies, while the spatial-wideband effect is always ignored. As a result, the beam generated for dif-

ferent frequencies cannot point toward the same direction, and reciprocally, the estimated incident angles are inaccurate if only the frequency-wideband effect is considered.

In Ref. [9], the influence of the spatial-wideband effect on the massive MIMO channel is discussed, and the transceiver is redesigned for orthogonal frequency-division multiplexing (OFDM) systems. In Ref. [10], the discrete-time channel for the hybrid mmWave massive MIMO system with the spatial-wideband effect is modeled and the wideband channel is estimated through the Newtonized fully corrective forward greedy selection-cross validation-based algorithm. The authors of Ref. [11] develop a spatial-wideband channel model for lens antenna arrays and design a fast angle estimation algorithm to establish point-to-point mappings between the spatial-frequency patterns and the strongest discrete Fourier transform beam containing the incident angle. Nevertheless, few work has considered the spatial-wideband channel model for RIS, and high-resolution RIS channel estimation approach is lacking.

In this paper, we analyze the frequency-dependent spatial-wideband effect for RIS-assisted communications and present a high-resolution channel estimation algorithm based on off-grid sparse Bayesian learning (SBL) approach. Simulation results are provided to verify the correctness of the spatial-wideband effect and the effectiveness of the proposed compressive sensing (CS) algorithm. Moreover, the challenges and promising directions of RIS future research are discussed.

## 2 Spatial-Wideband Channel Analysis

The array steering vector commonly used in MIMO is,

$$\mathbf{a}(\varphi) = \left[ 1, e^{j2\pi d \sin \varphi / \lambda}, \dots, e^{j2\pi(N-1)d \sin \varphi / \lambda} \right], \quad (1)$$

where  $d$  is the antenna spacing,  $\varphi$  is the angle of arrival and  $\lambda$  is the wavelength. When the number of array elements is enormous and the transmission bandwidth is wide, for example, in a typical RIS-assisted communication, new affecting factors need to be considered.

In this paper, the millimeter frequency band is considered, the RIS is deployed in fixed locations such as building walls without mobility, and the BS is deployed at a high location with almost no obstacles around. The line-of-sight (LoS) path transmission between BS and RIS is usually stable and dominates the energy of the channel. In this way, the LoS assumption is adopted for the channel between BS and RIS. For RIS-assisted communications, the LoS channel can be obtained with prior knowledge of the location information and simple RIS signaling. The BS-RIS channel parameters can be conveyed to user equipment (UE) through the control channel for RIS-UE channel estimation. Multiple RIS-UE channel estimations are performed within a single BS-RIS channel estimation period as the BS-RIS channel changes slowly.

The RIS-UE channel with spatial-wideband effect is analyzed. For multipath propagation, the passband signal re-

ceived by the  $(m, n)$ -th element of RIS is

$$\tilde{y}_{m,n}(t) = \left( \sum_{l=0}^{L-1} \beta_l e^{j2\pi f_c (t - \tau_{l,m,n})} \delta(t - \tau_{l,m,n}) \right) \times s(t) + w_{m,n}(t), \quad (2)$$

where  $L$  is the path number,  $f_c$  is the carrier frequency,  $\beta_l$  is the  $l$ -th path gain, and  $\tau_{l,m,n}$  is the delay of the  $l$ -th path received by the  $(m, n)$ -th RIS element. Since RIS is a two-dimensional planar structure, we have

$$\tau_{l,m,n} = \tau_l + m \frac{\psi^1}{f_c} + n \frac{\psi^2}{f_c}, \psi^1 = \frac{d \sin \varphi \cos \phi}{\lambda_c}, \psi^2 = \frac{d \cos \varphi}{\lambda_c}, \quad (3)$$

where  $\tau_l$  denotes the delay of path  $l$  for the first RIS element, and  $\varphi \in [0, 2\pi)$  and  $\phi \in [0, \pi)$  are the azimuth and elevation angles.

As a result, the received baseband signal by the  $(m, n)$ -th RIS element in the time domain is

$$y_{m,n}(t) = \left( \sum_{l=0}^{L-1} \bar{\beta}_l e^{-j2\pi f_c \left( m \frac{\psi_l^1}{f_c} + n \frac{\psi_l^2}{f_c} \right)} \delta \left( t - \tau_l - m \frac{\psi_l^1}{f_c} - n \frac{\psi_l^2}{f_c} \right) \right) \times s(t) + w_{m,n}(t), \quad (4)$$

where  $\bar{\beta}_l = \beta_l e^{-j2\pi f_c \tau_l}$  is the equivalent path gain for path  $l$ . Correspondingly, the received signal in the frequency domain is

$$y_{m,n}(f) = \left( \sum_{l=0}^{L-1} \bar{\beta}_l e^{-j2\pi f_c \left( m \frac{\psi_l^1}{f_c} + n \frac{\psi_l^2}{f_c} \right)} e^{-j2\pi f \left( m \frac{\psi_l^1}{f_c} + n \frac{\psi_l^2}{f_c} \right)} \right) s(f) + w_{m,n}(f) = \left( \sum_{l=0}^{L-1} \bar{\beta}_l e^{-j2\pi (m\psi_l^1 + n\psi_l^2) \left( 1 + \frac{f}{f_c} \right)} \right) s(f) + w_{m,n}(f). \quad (5)$$

The frequency domain signals received on all RIS elements can be written in a matrix form as

$$\mathbf{Y}(f) = \left( \sum_{l=0}^{L-1} \bar{\beta}_l \mathbf{A}_R \left( \left( 1 + \frac{f}{f_c} \right) \boldsymbol{\psi}_l^1, \left( 1 + \frac{f}{f_c} \right) \boldsymbol{\psi}_l^2 \right) \right) s(f) + \mathbf{W}(f), \quad (6)$$

where  $\mathbf{A}_R$  is the uniform rectangular spatial-domain steering matrix with  $\mathbf{A}_R(m, n) = e^{-j2\pi (m\psi_l^1 + n\psi_l^2) \left( 1 + \frac{f}{f_c} \right)}$ . Vectorization is performed, resulting in

$$\check{\mathbf{y}}(f) = \left( \sum_{l=0}^{L-1} \bar{\beta}_l \mathbf{a}_R \left( \left( 1 + \frac{f}{f_c} \right) \boldsymbol{\psi}_l^1, \left( 1 + \frac{f}{f_c} \right) \boldsymbol{\psi}_l^2 \right) \right) s(f) + \mathbf{w}(f) = \mathbf{g}(f) s(f) + \mathbf{w}(f). \quad (7)$$

$N$ -point OFDM modulation is adopted to overcome frequency selective fading, and the received signal can be arranged as

$$\mathbf{G} = [\mathbf{g}(0) \quad \mathbf{g}(\Delta f) \quad \dots \quad \mathbf{g}((N-1)\Delta f)] = \sum_{l=0}^{L-1} \bar{\beta}_l \mathbf{a}_R \left( \left(1 + \frac{f}{f_c}\right) \psi_l^1, \left(1 + \frac{f}{f_c}\right) \psi_l^2 \right) \mathbf{b}^T(\tau_l), \quad (8)$$

$$\text{where } \mathbf{b}(\tau) = \left[ e^{-j2\pi f \tau_l} \quad \dots \quad e^{-j2\pi(f + (N-1)\Delta f)\tau_l} \right].$$

For the spatial-narrowband case, the frequency-selective effect induced by the multipath delay spread is considered, while the spatial-selective effect is ignored. The approximation  $s\left(t - \tau_l - m \frac{\psi_l^1}{f_c} - n \frac{\psi_l^2}{f_c}\right) \approx s(t - \tau_l)$  holds and the propagation delay across the RIS can be neglected. The spatial-narrowband channel can be represented by

$$\mathbf{G}_{\text{Narrow}} = \sum_{l=0}^{L-1} \bar{\beta}_l \mathbf{a}_R(\psi_l^1, \psi_l^2) \mathbf{b}^T(\tau_l). \quad (9)$$

It can be observed that, for spatial-narrowband signals, a fixed beam can be generated with no frequency dependence. However, the generated beam has slight variations of the antenna radiation pattern correlated with the frequency for spatial-wideband signals. This is the main difference between spatial-wideband and spatial-narrowband channels.

When a RIS is large enough and the signal bandwidth is wide enough, the physical propagation delay is non-negligible. The spatial-wideband effect brings challenging channel estimation difficulties and seriously affects system performance. Moreover, the inherent passive characteristics of the RIS pose the same phase shift for all frequencies. Thus, RIS elements cannot adapt to different received signal phases, and the spatial-wideband effect cannot be compensated. It is necessary to carefully design channel estimation and beamforming schemes for RIS with the spatial-wideband effect.

Through the above analysis, we have obtained the expression of the RIS-UE channel considering the spatial-wideband effect. The spatial-wideband OFDM channel can be represented by parameters including the angle, delay, and path gain. Once these parameters are obtained, the overall channel can be obtained easily.

For simplification, the received OFDM signal can be written as

$$\mathbf{y} = \sum_{l=0}^{L-1} \bar{\beta}_l \text{vec} \left[ \mathbf{a}_R \left( \left(1 + \frac{f}{f_c}\right) \psi_l^1, \left(1 + \frac{f}{f_c}\right) \psi_l^2 \right) \mathbf{b}^T(\tau_l) \right] \mathbf{s} + \mathbf{n} = \sum_{l=0}^{L-1} \bar{\beta}_l \mathbf{c}(\psi_l^1, \psi_l^2, \tau_l) \mathbf{s} + \mathbf{n}, \quad (10)$$

$$\text{where } \mathbf{c}(\psi_l^1, \psi_l^2, \tau_l) = \text{vec} \left[ \mathbf{a}_R \left( \left(1 + \frac{f}{f_c}\right) \psi_l^1, \left(1 + \frac{f}{f_c}\right) \psi_l^2 \right) \mathbf{b}^T(\tau_l) \right].$$

Moreover, to estimate the RIS-UE channel from the received OFDM signal, a simple idea is to select part of the received signal to recover those parameters, avoiding the high dimension disadvantage.

### 3 Off-Grid CS Based Channel Estimation

To estimate the RIS channel with severe spatial-wideband effect, we design an off-grid SBL algorithm to estimate the channel parameters  $\hat{\phi}$ ,  $\hat{\varphi}$ ,  $\hat{\tau}$ , and  $\hat{\beta}$ . It should be noted that, according to the obtained parameters, we can not only recover the pilot channel but also recover the channel on all subcarriers and extrapolate it to the channel of the complete RIS.

On-grid CS algorithms such as orthogonal matching pursuit (OMP)<sup>[12]</sup> are hard to achieve satisfying performance with low complexity as the number of channel parameters is enormous. The SBL algorithm is adopted to achieve super-resolution parameter estimation performance. A parameterized dictionary with a varying grid is utilized to overcome the grid mismatch problem. According to Ref. [13], uniform distribution is adopted as a priori distribution for angle and delay parameters. In this way, the channel estimation problem can be transformed into maximizing the posterior probability of multi-dimensional variables of angle, delay, and channel gain jointly.

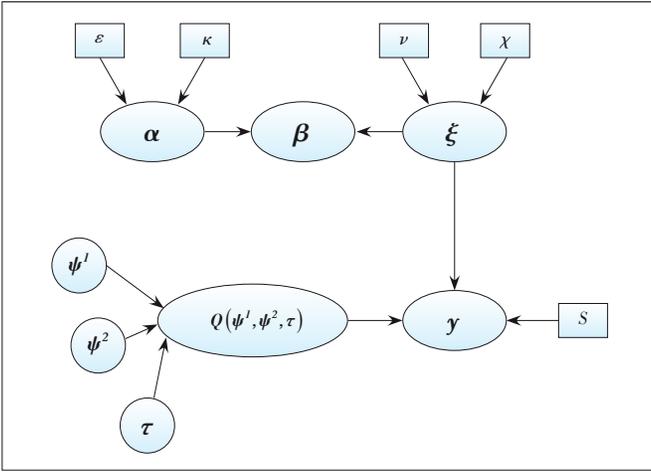
Define the dictionary as  $\mathbf{Q}(\boldsymbol{\psi}^1, \boldsymbol{\psi}^2, \boldsymbol{\tau}) = \left[ \mathbf{c}(\psi_1^1, \psi_1^2, \tau_1) \quad \dots \quad \mathbf{c}(\psi_D^1, \psi_D^2, \tau_D) \right]$ , where  $D$  is the path number and the parameters of each column are variable. As the path number is unknown,  $D$  is set to be a relatively large value initially and adjusted by deleting or adding a new column  $\mathbf{c}(\psi_l^1, \psi_l^2, \tau_l)$ .

The posterior probability distribution function (PDF) of the received spatial-wideband OFDM signal is

$$p(\mathbf{y} | \boldsymbol{\psi}^1, \boldsymbol{\psi}^2, \boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\xi}) = CN\left((\mathbf{S}^T \otimes \mathbf{I}) \times \mathbf{Q}(\boldsymbol{\psi}^1, \boldsymbol{\psi}^2, \boldsymbol{\tau}) \boldsymbol{\beta}, \boldsymbol{\xi}^{-1} \mathbf{I}\right). \quad (11)$$

The Bayesian network of the RIS-UE channel model is shown in Fig. 1. The prior distribution assumption for noise inverse variance is gamma distribution. The channel gain obeys the two-layer prior model, where the first layer is the zero-mean Gaussian distribution w. r. t. hyperparameter vector  $\boldsymbol{\alpha}$  and the second layer follows gamma distribution  $\Gamma(\cdot | \boldsymbol{\varepsilon}, \boldsymbol{\kappa})$  that connects  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  as  $p(\boldsymbol{\beta} | \boldsymbol{\alpha}, \boldsymbol{\xi}) = CN(\mathbf{0}, \boldsymbol{\xi} \mathbf{A}), p(\boldsymbol{\alpha}_i) = \Gamma(\boldsymbol{\alpha}_i | \boldsymbol{\varepsilon}, \boldsymbol{\kappa})$  and  $\mathbf{A} = \text{diag}(\boldsymbol{\alpha})$ . Moreover, those angle-related parameters and delay parameters are assumed to be uniformly distributed.

The type II sparse Bayesian learning estimator<sup>[13]</sup> is adopted and the associated problem becomes the search for



▲ Figure 1. Bayesian network for reconfigurable intelligent surface (RIS)-user equipment (UE) channel

$\{\hat{\psi}^1, \hat{\psi}^2, \hat{\tau}, \hat{\alpha}, \hat{\xi}\}$  utilizing the maximum a posteriori (MAP) criterion. The channel parameters can be inferred from

$$\begin{aligned} \{\hat{\psi}^1, \hat{\psi}^2, \hat{\tau}, \hat{\alpha}, \hat{\xi}\} &= \arg \max \ln p(\psi^1, \psi^2, \tau, \alpha, \xi | y) \\ &\propto \arg \max \ln p(y | \psi^1, \psi^2, \tau, \alpha, \xi) p(\psi^1) p(\psi^2) p(\tau) p(\alpha) p(\xi). \end{aligned} \quad (12)$$

After calculation, we find that the marginal likelihood obeys Gaussian distribution as

$$p(y | \psi^1, \psi^2, \tau, \alpha, \xi) = CN(\mathbf{0}, \xi^{-1} \mathbf{C}), \quad (13)$$

where  $\mathbf{C} = \mathbf{C}_{-i} + \alpha_i \mathbf{c}(\psi_i^1, \psi_i^2, \tau_i) \mathbf{c}^H(\psi_i^1, \psi_i^2, \tau_i)$  and  $(\cdot)_{-i}$  represent a new matrix with the component related to the  $i$ -th path removed. Defining  $L(\psi^1, \psi^2, \tau, \alpha, \xi) = \ln p(y | \psi^1, \psi^2, \tau, \alpha, \xi)$ , the MAP criterion is equivalent to maximize  $L$ .

It can be inferred from Eq. (13) that  $L$  can be split into one part that is related to the  $i$ -th path and the other parts related to the other paths. An iterative approach can be utilized to find the desired solution for each path separately. The detail can be summarized as:

- 1) The noise variance is obtained by setting the partial derivative of  $L$  w.r.t.  $\xi$  to be zero.
- 2) For path  $i$ , the hyperparameter  $\alpha_i$  can be obtained by choosing a value to make the partial derivative w.r.t.  $\alpha_i$  to be zero.
- 3) The estimated  $\psi^1, \psi^2$  and  $\tau$  are obtained separately through the gradient descent or Newton method.

Although the globally optimal condition cannot be guaranteed, the proposed off-grid SBL algorithm can achieve a continuous convergence of the objective function and obtain a locally optimal solution. As the proposed algorithm processes continuous angle and delay directly, it can effectively overcome the grid mis-

match problem and achieve super-resolution estimation accuracy.

## 4 Simulation Results

In this section, the channel estimation performance is analyzed. The RIS adopted is composed of  $16 \times 16$  elements, and  $4 \times 4$  elements are selected for channel estimation. The number of OFDM subcarriers is 64, and 8 or 16 subcarriers are selected for channel estimation. The central frequency is 60 GHz and the bandwidth is 1 GHz.

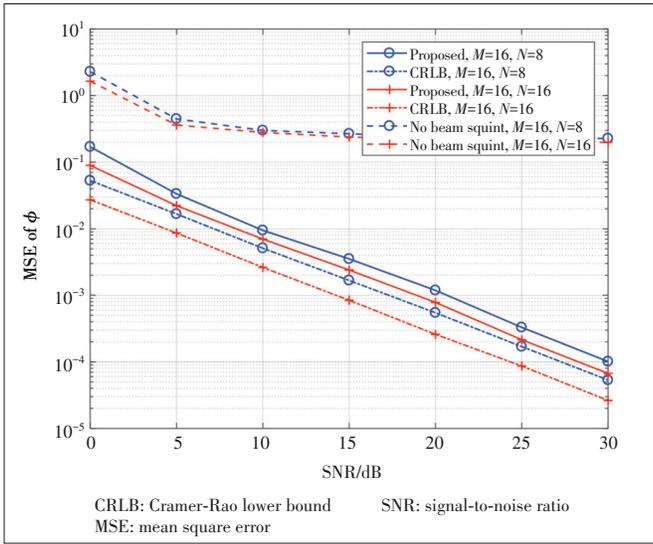
In Fig. 2, we simulate the estimated mean square error (MSE) of  $\phi$ , which is defined as  $MSE_\phi = |\phi - \hat{\phi}|^2$  and  $\phi$  is measured in angles. Firstly, the angle estimation performance of the one that does not consider spatial-wideband effect is abysmal. It does not decrease with the signal-to-noise ratio (SNR), remaining a value larger than 0.1 even for noiseless cases. This proves the existence of the spatial-wideband effect in addition to the frequency-wideband effect, verifying the correctness of the proposed RIS-assisted channel model. Secondly, it can be found that the proposed off-grid SBL channel estimation algorithm achieves a better angular MSE performance that decreases approximately linearly with SNR. Moreover, the proposed algorithm approaches the Cramer-Rao lower bound (CRLB). The superiority of the proposed method is verified, although the global optimal cannot be guaranteed. Increasing the number of subcarriers used for channel estimation can improve angle estimation accuracy. The estimation of  $\phi$  and  $\tau$  have similar results, and the detail is omitted due to space limitations.

In Fig. 3, the normalized MSE of the RIS-UE channel for all RIS elements, which is defined as  $MSE_H = |\mathbf{H} - \hat{\mathbf{H}}|^2 / |\mathbf{H}|^2$ , is analyzed. The MSE performance of those ignoring the spatial-wideband effect is worse than the proposed algorithm and does not improve with higher SNR. In addition, the MSE of the proposed algorithm decreases continuously with SNR, and the error floor that exists for the on-grid OMP algorithm can be overcome. Though very fine grids make the channel estimation performance of OMP better, it becomes numerically unstable and computationally unacceptable and can hardly beat the proposed off-grid one.

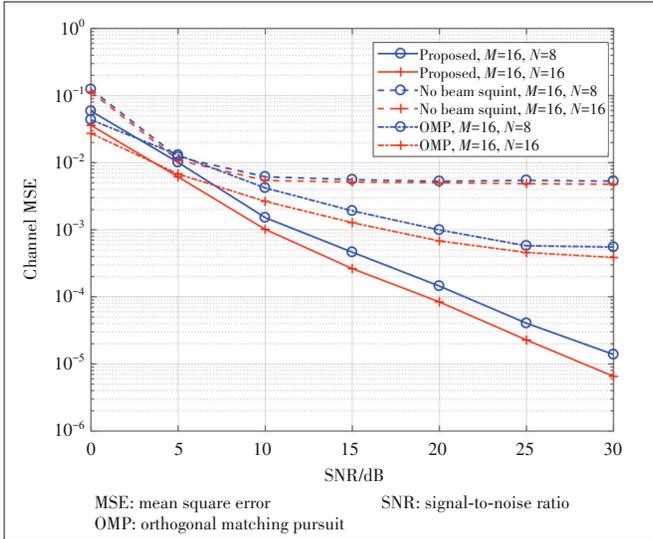
## 5 Foresight of Challenges

### 5.1 Channel Modeling

Most of the existing channel modeling methods follow the wireless channel models of 3GPP and ITU. However, the new challenges brought by the novel characteristics of RIS are ignored, and the lack of RIS wireless channel measurement becomes a tremendous challenge for the current stage. It is urgent to develop a thoroughgoing RIS channel model considering various fading factors. In addition, it is necessary to reasonably model hardware non-ideal factors such as the mutual coupling between sub-wavelength RIS elements and verify it



▲ Figure 2. MSE of angle parameter  $\phi$



▲ Figure 3. Channel estimation MSE versus SNR

through practical channel measurement.

The electromagnetic mutual coupling means that when multiple antennas are close to each other, the electric field of one antenna will affect the current distribution of other antennas and generate a voltage at its terminals, thereby affecting the performance of the antenna array. To realize the intelligent control of the environment, a RIS is usually composed of tens of thousands of electromagnetic elements, with the element interval being less than half a wavelength. The mutual coupling within RIS is severe, and the radiation pattern is affected dramatically. If no compensation is performed, the system performance will be decreased. The commonly used electromagnetic coupling analysis methods are the method of moment approach<sup>[14]</sup> and the mutual coupling model based on S parameters<sup>[15]</sup>. However, the hardware implementation of the RIS is diverse, including structures based on PIN tubes, varactor diodes, liquid crystals, etc. How to reasonably model the imped-

ance of RIS needs in-depth research.

## 5.2 Near-Field Analysis

In MIMO communications, it is assumed that the signals received by the antenna array are parallel. However, this approximation only holds when the receiving array is in the far-field of the transmitting array. The case is different for a RIS as the near-field condition occurs more frequently. The transmission characteristics of spherical waves need to be considered, and near-field research needs to be promoted urgently.

According to different transmission distances, the receiving area is divided into the reactive near-field, near-field and far-field, and the boundaries are  $0.6\sqrt{D_a^3/\lambda}$  and  $2D_a^2/\lambda$  respectively, with  $D_a$  being the array aperture. As the array aperture for the RIS is usually large, the signals received by different RIS elements have different transmission distances and different effective receiving areas, and polarization mismatch errors are inevitable<sup>[16]</sup>. These factors cause significant differences in the signal received by different RIS elements, making the RIS near-field analysis more complicated than MIMO. The performance analysis and design optimization of near-field communications are helpful for understanding the characteristics and potential advantages of RIS-assisted communications.

## 5.3 Multi-User Transmission

A significant feature of RIS is that the electromagnetic element can only produce a certain phase adjustment at a particular moment<sup>[17]</sup>, and it is difficult for RIS to achieve ideal multi-beams for multi-user transmission. One solution is maximizing a criterion and achieving a tradeoff by optimizing the RIS phase shift matrix. However, this solution is complicated and generally difficult to generate beams with precise directivity, restricting its value in actual products. Other solutions include RIS segmentation service for multi-users, matching BS-RIS and RIS-UE beam pairs to obtain suitable beams supporting multi-users, etc. However, the inherent performance loss due to the amplitude restriction of RIS element is difficult to compensate. Different subcarriers incident into the same RIS will cause different RIS electromagnetic responses, and different beams will be generated due to the spatial-wideband effect. Reasonable use of these properties may bring new inspiration to multi-user RIS-assisted transmission.

After the introduction of RIS, the communication system will face many challenges such as a more complex propagation environment, an increase in factors affecting channel capacity, etc. RIS needs to perform different functions according to different requirements, such as beamforming, energy focusing and near-field broadcasting<sup>[18]</sup>. Studying the maximum capacity, energy efficiency, or optimization of other indicators under various schemes will be a hot spot for future research.

## 5.4 Standardization

Standardization is an essential connection from technology

to application. Although the research on RIS is still in the development stage, it is indispensable to study the impact of RIS standardization. At present, there are various channel estimation and beamforming approaches in academia, while the most basic question considered by the industry is how to standardize and apply them.

As a simple version of RIS is hoped to be deployed in 5G-advanced network, the standardization of RIS is just around the corner. The points to be standardized for RIS include the use cases and requirements, RIS channel modeling, air interface design, and the RIS network architecture. The standardization mainly includes near-field and far-field channel modeling, time slot structure design, channel and signaling design, channel measurement design, feedback design, and related protocol process design. The possible air interface designs include reference signaling, channel feedback signaling, codebook, control signaling, etc.

## 6 Conclusions

The spatial-wideband effect is inevitable in the RIS channel, and the communication quality will be affected seriously if the spatial-wideband effect is not handled well. To deal with this problem, we reanalyze the wideband RIS channel representation and design an off-grid SBL method to estimate the sparse channel parameters. The estimation of the channel parameters through the activation of some RIS elements is realized and the complete channel is restored successfully. The simulation results prove the effectiveness of the off-grid SBL algorithm. Finally, the challenges to be dealt with for RIS-assisted communications are discussed.

## References

- [1] STRINATI E C, BARBAROSSA S, GONZALEZ-JIMENEZ J L, et al. 6G: the next frontier: from holographic messaging to artificial intelligence using subterahertz and visible light communication [J]. *IEEE vehicular technology magazine*, 2019, 14(3): 42 – 50. DOI: 10.1109/MVT.2019.2921162
- [2] CUI T J. Information metamaterial and metasurface-from concept to system [C]//43rd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz). IEEE, 2018: 1 – 3. DOI: 10.1109/IRMMW-THz.2018.8510031
- [3] NGUYEN C L, GEORGIOUS O, GRADONI G, et al. Wireless fingerprinting localization in smart environments using reconfigurable intelligent surfaces [J]. *IEEE access*, 2021, 9: 135526 – 135541. DOI: 10.1109/ACCESS.2021.3115596
- [4] MISHRA D, JOHANSSON H. Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer [C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 4659 – 4663. DOI: 10.1109/ICASSP.2019.8683663
- [5] HE Z Q, YYAN X. Cascaded channel estimation for large intelligent metasurface assisted massive MIMO [J]. *IEEE wireless communications letters*, 2019, 9(2): 210 – 214. DOI: 10.1109/LWC.2019.2948632
- [6] JIAN M, ZHAO Y. A modified off-grid SBL channel estimation and transmission strategy for RIS-assisted wireless communication systems [C]//2020 International Wireless Communications and Mobile Computing (IWCMC). IEEE, 2020: 1848 – 1853. DOI: 10.1109/IWCMC48107.2020.9148537
- [7] LIU C, LIU X, NG D W K, et al. Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications [J]. *IEEE transactions on wireless communications*, 2021. DOI: 10.1109/TWC.2021.3100148
- [8] TAHA A, ALRABEIAH M, ALKHATEEB A. Enabling large intelligent surfaces with compressive sensing and deep learning [J]. *IEEE access*, 2021, 9: 44304 – 44321. DOI: 10.1109/ACCESS.2021.3064073
- [9] WANG B, JIAN M, GAO F, et al. Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems [J]. *IEEE transactions on signal processing*, 2019, 67(23): 5893 – 5908. DOI: 10.1109/TSP.2019.2949502
- [10] KIM I, CHOI J. Spatial wideband channel estimation for mmWave massive MIMO systems with hybrid architectures and low-resolution ADCs [J]. *IEEE transactions on wireless communications*, 2021, 20(6): 4016 – 4029. DOI: 10.1109/TWC.2021.3054998
- [11] WU K, NI W, SU T, et al. Exploiting spatial-wideband effect for fast AoA estimation at lens antenna array [J]. *IEEE journal of selected topics in signal processing*, 2019, 13(5): 902 – 917. DOI: 10.1109/JSTSP.2019.2937691
- [12] KULKARNI A, MOHSENIN T. Low overhead architectures for OMP compressive sensing reconstruction algorithm [J]. *IEEE transactions on circuits and systems I: regular papers*, 2017, 64(6): 1468 – 1480. DOI: 10.1109/TCSI.2017.2648854
- [13] JIAN M, GAO F, TIAN Z, et al. Angle-domain aided UL/DL channel estimation for wideband mmWave massive MIMO systems with beam squint [J]. *IEEE transactions on wireless communications*, 2019, 18(7): 3515 – 3527. DOI: 10.1109/TWC.2019.2915072
- [14] GANESH M, SUBHASHINI K R, JAGAN B. A method of moments based algorithm to synthesize antenna arrays [C]//2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2017: 1091 – 1095. DOI: 10.1109/WiSPNET.2017.8299931
- [15] AZHAR A A, ABD RASHIDN E, MAHMOOD M K A, et al. Effects of mutual coupling and array dissimilarity on angle estimation in MIMO radar [C]//2019 IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE). IEEE, 2019: 1 – 6. DOI: 10.1109/APACE47377.2019.9020903
- [16] BJORMSON E, SANGUONETTI L. Power scaling laws and near-field behaviors of massive MIMO and intelligent reflecting surfaces [J]. *IEEE open journal of the communications society*, 2020, 1: 1306 – 1324. DOI: 10.1109/OJCOMS.2020.3020925
- [17] KAMMOUN A, CHAABAN A, DEBBAH M, et al. Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems [J]. *IEEE transactions on wireless communications*, 2020, 19(12): 7748 – 7764. DOI: 10.1109/TWC.2020.2986438
- [18] PAN C, REN H, WANG K, et al. Reconfigurable intelligent surfaces for 6G systems: principles, applications, and research directions [J]. *IEEE communications magazine*, 2021, 59(6): 14 – 20. DOI: 10.1109/MCOM.001.2001076

## Biographies

**JIAN Mengnan** (jian.mengnan@zte.com.cn) received the B.E. degree in information engineering from Beijing Institute of Technology, China in 2016, and the M.S. degree from Tsinghua University, China in 2019. She is currently an engineer with ZTE Corporation. Her research interests include reconfigurable intelligent surfaces and orbital angular momentum.

**ZHANG Nan** received the bachelor degree in communication engineering and the master degree in integrated circuit engineering from Tongji University, China in July 2012 and March 2015, respectively. He is now a senior engineer with ZTE Corporation and works on the standardization of LTE and NR systems. His current research interests include channel modeling, MIMO, NOMA techniques, satellite/ATG communications and reconfigurable intelligent surfaces.

**CHEN Yijian** received the B.S. degree from Central South University, China in 2006. He is currently a senior engineer with ZTE Corporation. His current research interests include massive MIMO, coordinated multi-point transmission, high-frequency communications, and channel modeling.

# Dual-Polarized RIS-Based STBC Transmission with Polarization Coupling Analysis



ZHOU Mingyong<sup>1</sup>, CHEN Xiangyu<sup>1</sup>, TANG Wankai<sup>1</sup>,  
KE Jun Chen<sup>2</sup>, JIN Shi<sup>1</sup>, CHENG Qiang<sup>2</sup>, CUI Tie Jun<sup>2</sup>

(1. National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China;

2. State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China)

DOI: 10.12142/ZTECOM.202201009

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220302.1148.002.html>,

published online March 3, 2022

Manuscript received: 2021-05-12

**Abstract:** The rapid development of the reconfigurable intelligent surface (RIS) technology has given rise to a new paradigm of wireless transmitters. At present, most research works on RIS-based transmitters focus on single-polarized RISs. In this paper, we propose a dual-polarized RIS-based transmitter, which realizes 4-transmit space-time block coding (STBC) transmission by properly partitioning RIS's unit cells and utilizing the degree of freedom of polarization. The proposed scheme is evaluated through a prototype system that utilizes a fabricated dual-polarized phase-adjustable RIS. In particular, the polarization coupling phenomenon in each unit cell of the employed dual-polarized RIS is modeled and analyzed. The experimental results are in good agreement with the theoretical modeling and analysis results, and an initial research effort is made on characterizing the polarization coupling property in the dual-polarized RIS.

**Keywords:** reconfigurable intelligent surface; space-time coding; dual polarization; coupling modeling; prototyping

**Citation** (IEEE Format): M. Y. Zhou, X. Y. Chen, W. K. Tang, et al., "Dual-polarized RIS-based STBC transmission with polarization coupling analysis," *ZTE Communications*, vol. 20, no. 1, pp. 63 - 75, Mar. 2022. doi: 10.12142/ZTECOM.202201009.

## 1 Introduction

With the commercialization of the fifth-generation (5G) mobile communication networks, the spectrum resources of millimeter wave band are gradually developed and utilized, and base stations (BSs) also begin to assemble massive multiple-input multiple-output (MIMO) hardware and corresponding algorithms to provide better communication services<sup>[1-2]</sup>. However, the antenna array equipped with massive MIMO needs numerous radio frequency (RF) chains, which leads to high hardware costs, high energy consumption, high system complexity, and severe heat dissipation issues. These problems will become more serious in the sixth-generation (6G) mobile communication networks. The emergence of reconfigurable intelligent surfaces (RISs) in recent years may bring new solutions to these challenges<sup>[3-6]</sup>.

The concept of digital coding and programmable metamaterials was first proposed in 2014<sup>[7]</sup>. When applied to the field of wireless communications, the reconfigurable two-dimensional

metasurfaces are usually called RISs, which are composed of many sub-wavelength unit cells<sup>[8-9]</sup>. Each unit cell of a RIS can be configured to change its response to electromagnetic (EM) waves, so as to adjust and control EM wave parameters, such as the amplitude or phase, during the interaction between the unit cell and EM waves<sup>[10]</sup>. A RIS is usually controlled by a field programmable gate array (FPGA) for real-time configuration to realize the regulation and control of EM waves.

The research on RISs can be categorized into two major application directions: the RIS-based transmitter and the RIS-based relay. The RIS-based transmitter focuses on using RIS to regulate EM wave parameters, thus modulating information on EM waves. The RIS-based relay focuses on improving the channel quality through RISs, thus enhancing the signal strength and coverage performance<sup>[11-12]</sup>.

There have been some prior research works on RIS-based transmitters. The authors of Ref. [13] proposed a binary frequency shift keying (BFSK) wireless communication system based on RIS, and the authors of Refs. [14 - 16] realized several RIS-based phase shift keying (PSK) transmission prototypes. A RIS-based space-time transmission system was proposed and implemented in Ref. [17]. However, previous research on RIS-based transmitters was based on single-polarized RISs, and the potential of RISs in the degree of freedom of polarization was not fully exploited. Recently, some fabricated RISs have been able to realize EM control in different polarization directions<sup>[18]</sup>. The authors of Ref. [19] realized

Corresponding author: JIN Shi

This work was supported by the National Key Research and Development Program of China (2018YFA0701602, 2017YFA0700201, 2017YFA0700202, 2017YFA0700203), the National Natural Science Foundation of China (NSFC) under Grants 61941104, 61921004, 61531011, the Fundamental Research Funds for the Central Universities under Grant 2242022R10062, and the Key Research and Development Program of Shandong Province under Grant 2020CXGC010108.

MIMO transmission based on a dual-polarized RIS, and the authors of Ref. [20] proposed a communication system assisted by a dual-polarized RIS.

In this paper, a 4-transmit space-time block coding (STBC) transmission system based on a dual-polarized RIS is proposed and validated by experiments. In this system, a 4-transmit space-time transmitter is realized by utilizing the degree of freedom of the RIS's polarization. In addition, the polarization coupling phenomenon in the unit cell of the employed dual-polarized RIS is modeled and analyzed.

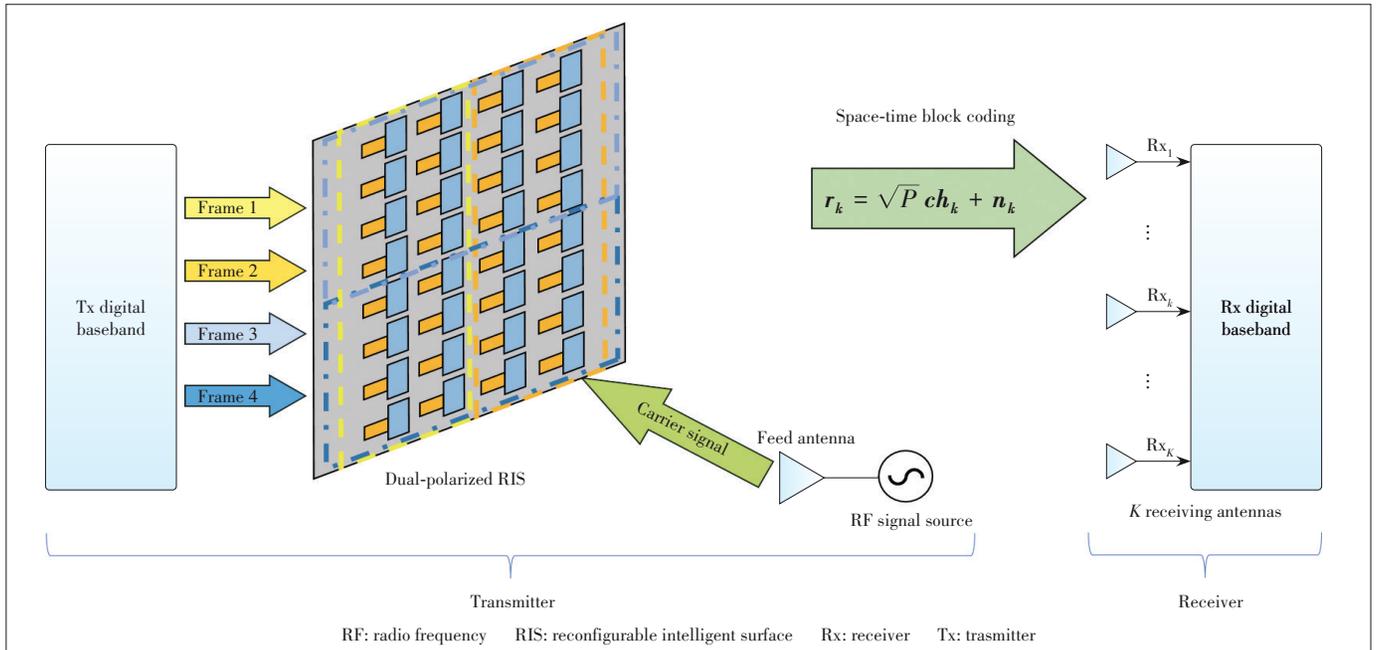
The rest of the paper is organized as follows. Section 2 introduces the system model of the proposed 4-transmit RIS-based STBC transmission system. Section 3 gives the design details of the proposed system and Section 4 analyzes the bit error rate (BER) performance by considering polarization coupling interference in each unit cell. The prototype system and experimental measurement results are shown in Section 5. Section 6 summarizes the paper.

## 2 System Model

Fig. 1 shows the diagram of the 4-transmit space-time transmission system based on a dual-polarized RIS. The transmitter consists of a digital baseband module, a dual-polarized phase-adjustable RIS and a radio frequency (RF) signal source. The receiver has  $K$  receiving antennas and a baseband processing module. The dual-polarized RIS has  $N$  rows and  $M$  columns of unit cells, and each unit cell can realize the phase regulation in horizontal polarization and vertical polarization, respectively.  $M$  and  $N$  are assumed to be even. The single tone carrier signal incident on the dual-polarized RIS is emitted by a feed antenna.

The incident signal power of each unit cell is  $p$  in both horizontal and vertical polarization. Let  $p$  represent the polarization direction,  $p = 1$  means vertical polarization, and  $p = 0$  means horizontal polarization. The unit cell in row  $n$  and column  $m$  is denoted as  $U_{n,m}$ .  $A_{p,n,m}$  and  $\varphi_{p,n,m}$  represent the reflection amplitude and reflection phase shift of  $U_{n,m}$  in polarization  $p$ , and the channel from  $U_{n,m}$  to the  $k$ -th receiving antenna in polarization  $p$  is denoted as  $h_k^{p,n,m}$ . Then the signal received from  $U_{n,m}$  by the  $k$ -th receiving antenna in polarization  $p$  is  $y_k^{p,n,m} = h_k^{p,n,m} \sqrt{P} A_{p,n,m} e^{j\varphi_{p,n,m}}$ . It can be noted that  $x_{p,n,m} = A_{p,n,m} e^{j\varphi_{p,n,m}}$  represents the data symbol modulated by  $U_{n,m}$  onto the single-tone carrier in polarization  $p$  and we have  $y_k^{p,n,m} = h_k^{p,n,m} \sqrt{P} x_{p,n,m}$ .

The  $n \times m$  unit cells of the RIS can be divided into left and right parts in horizontal polarization, and upper and lower parts in vertical polarization. As shown in Fig. 1, the horizontal left part is marked as the yellow part, the horizontal right part is marked as the orange part, the vertical upper part is marked as the blue part, and the vertical lower part is marked as the indigo part. Then the wireless channel from the yellow part to the  $k$ -th receiving antenna is denoted as  $h_{k,1} = \sum_{m=1}^{\frac{M}{2}} \sum_{n=1}^N h_k^{0,n,m}$ , and accordingly,  $h_{k,2} = \sum_{m=\frac{M}{2}+1}^M \sum_{n=1}^N h_k^{0,n,m}$ ,  $h_{k,3} = \sum_{m=1}^{\frac{M}{2}} \sum_{n=\frac{N}{2}+1}^N h_k^{1,n,m}$ , and  $h_{k,4} = \sum_{m=\frac{M}{2}+1}^M \sum_{n=\frac{N}{2}+1}^N h_k^{1,n,m}$  stand for the wireless channel between the orange part and the  $k$ -th receiving antenna, the wireless channel between the blue part and the  $k$ -th receiving antenna, and the wireless channel between the indigo part and the  $k$ -th receiving antenna, respectively. Therefore, the



▲ Figure 1. Proposed 4-transmit space-time wireless communication system based on a dual-polarized RIS

channel parameters from the 4-transmit space-time dual-polarized RIS-based transmitter to the receiver can be written as

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{K,1} \\ h_{1,2} & \cdots & h_{K,2} \\ h_{1,3} & \cdots & h_{K,3} \\ h_{1,4} & \cdots & h_{K,4} \end{bmatrix} \quad (1)$$

The STBC matrix is designed as<sup>[21]</sup>:

$$\mathbf{c} = \begin{bmatrix} c_1 & c_2 & c_3 & 0 \\ -c_2^* & c_1^* & 0 & c_3 \\ -c_3^* & 0 & c_1^* & -c_2 \\ 0 & -c_3^* & c_2^* & c_1 \end{bmatrix} \quad (2)$$

where  $c_1$ ,  $c_2$ , and  $c_3$  are three transmitted baseband symbols. The coding matrix encodes three data streams into four data frames and transmits them in four time slots as a period. The received signals of the  $k$ -th receiving antenna in the corresponding time slot of coding are

$$\begin{aligned} r_1^k &= \sqrt{P} c_1 h_{k,1} + \sqrt{P} c_2 h_{k,2} + \sqrt{P} c_3 h_{k,3} + n_1^k, \\ r_2^k &= -\sqrt{P} c_2^* h_{k,1} + \sqrt{P} c_1^* h_{k,2} + \sqrt{P} c_3 h_{k,4} + n_2^k, \\ r_3^k &= -\sqrt{P} c_3^* h_{k,1} + \sqrt{P} c_1^* h_{k,3} - \sqrt{P} c_2 h_{k,4} + n_3^k, \\ r_4^k &= -\sqrt{P} c_3^* h_{k,2} + \sqrt{P} c_2^* h_{k,3} + \sqrt{P} c_1 h_{k,4} + n_4^k, \end{aligned} \quad (3)$$

in which  $r_i^k$  is the data received by the  $k$ -th receiving antenna in the  $i$ -th time slot, and  $n_i^k$  is the corresponding receiver noise. In a vector form, let  $\mathbf{r}_k = [r_1^k, r_2^k, r_3^k, r_4^k]^T$ ,  $\mathbf{n}_k = [n_1^k, n_2^k, n_3^k, n_4^k]^T$ , and  $\mathbf{h}_k = [h_{k,1}, h_{k,2}, h_{k,3}, h_{k,4}]^T$ , and we have

$$\mathbf{r}_k = \sqrt{P} \mathbf{c} \mathbf{h}_k + \mathbf{n}_k, \quad (4)$$

where  $\mathbf{r}_k$ ,  $\mathbf{n}_k$  and  $\mathbf{h}_k$  are the received signal vector, the noise vector, and the wireless channel vector of the  $k$ -th antenna, respectively.

Eq. (4) reveals the basic principle of the 4-transmit space-time transmission system based on the dual-polarized RIS. It is consistent with the traditional space-time transmission system in form but different in hardware composition and channel structure. The proposed system modulates the information onto the carrier in the process of reflecting the signal through the dual-polarized RIS, which does not require the conventional RF chains. Therefore, the 4-transmit space-time transmission scheme based on the dual-polarized RIS provides a new space-time transmitter design with the advantages of low hardware costs and complexity.

### 3 Design of 4-Transmit Space-Time Transmission Based on Dual-Polarized RIS

This section will introduce in detail the scheme of a 4-transmit space-time transmission wireless communication system based on the dual-polarized RIS, including the design of waveforms, the wireless frame structure, and the receiver combining method.

#### 3.1 Waveform Design of Transmitted Symbols

Space-time coding needs to realize the complex conjugate operation of the transmitted symbols. However, the amplitude response and phase shift of the dual-polarized phase-adjustable RIS used in this paper are coupled. If the phase of RIS's unit cell is directly regulated, the envelope of the transmitted symbol will be non-constant and the complex conjugate operation of the symbol cannot be realized. Therefore, the non-linear modulation technology is employed in this work<sup>[17]</sup>.

During one symbol period, the transmitted symbol waveform is designed as

$$x_{p,n,m}(t) = \begin{cases} e^{j \frac{\Delta\varphi_{p,n,m}}{T_s} (T_s - t_{p,n,m} - t)}, & t \in [0, T_s - t_{p,n,m}] \\ e^{j \frac{\Delta\varphi_{p,n,m}}{T_s} (2T_s - t_{p,n,m} - t)}, & t \in (T_s - t_{p,n,m}, T_s], \end{cases} \quad (5)$$

where  $T_s$  is the symbol period,  $t_{p,n,m} \in [0, T_s]$  is the circular time delay, and the symbol phase decreases linearly with  $\frac{\Delta\varphi_{p,n,m}}{T_s}$  as the slope.

According to the designed symbol waveform, the amplitude  $A_{p,n,m}^{-1}$  and phase  $\varphi_{p,n,m}^{-1}$  of the  $-1$ -st harmonic with frequency  $f_s - \frac{1}{T_s}$  are

$$A_{p,n,m}^{-1} = \left| \text{sinc} \left( \frac{\Delta\varphi_{p,n,m}}{2} - \pi \right) \right|, \quad (6)$$

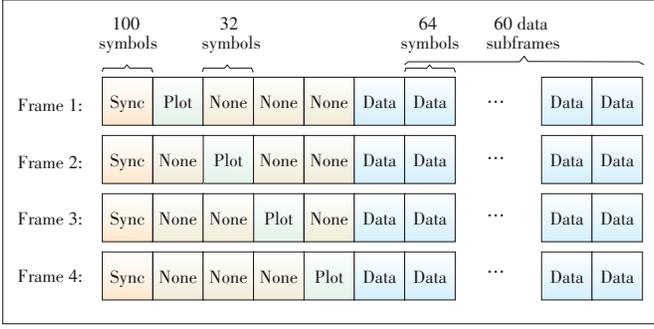
and

$$\begin{aligned} \varphi_{p,n,m}^{-1} &= -\frac{2\pi t_{p,n,m}}{T_s} + \frac{\Delta\varphi_{p,n,m}}{2} - \pi + \pi \cdot \\ &\text{mod} \left( \left\lfloor \frac{\Delta\varphi_{p,n,m}}{2\pi} - 1 \right\rfloor, 2 \right) + \pi \cdot \varepsilon(2\pi - \Delta\varphi_{p,n,m}), \end{aligned} \quad (7)$$

in which  $\text{sinc}(\cdot)$ ,  $\text{mod}(\cdot)$ ,  $\lfloor \cdot \rfloor$ , and  $\varepsilon(\cdot)$  denote sinc function, modulus operation, round down function and step function, respectively.

#### 3.2 Wireless Frame Structure Design

The frame structure is designed as shown in Fig. 2 for 4-



▲ Figure 2. Frame structure design of 4-transmit space-time transmission system based on dual-polarized RIS

transmit space-time coding transmission based on the dual-polarized RIS. It includes one synchronization subframe, one pilot subframe and 60 data subframes. The pilots of four transmitting data frames are orthogonal in time for ease of channel estimation. The four data frames are transmitted by the four parts of the RIS respectively. Each pilot subframe includes 32 Binary Phase Shift Keying (BPSK) symbols. There are 64 space-time coded 16-quadrature amplitude modulation (QAM) symbols in each data subframe, and all data symbols are based on the previously described waveform. 11 520 ( $4 \times 60 \times 64 \times 3/4 = 11\ 520$ ) data bits are transmitted per frame duration.

### 3.3 Combining Scheme of Receiver

When the receiver has only one receiving antenna ( $K=1$ ), we have

$$\begin{aligned}
 r_1^1 &= \sqrt{P} c_1 h_{1,1} + \sqrt{P} c_2 h_{1,2} + \sqrt{P} c_3 h_{1,3} + n_1^1, \\
 r_2^1 &= -\sqrt{P} c_2^* h_{1,1} + \sqrt{P} c_1^* h_{1,2} + \sqrt{P} c_3 h_{1,4} + n_2^1, \\
 r_3^1 &= -\sqrt{P} c_3^* h_{1,1} + \sqrt{P} c_1^* h_{1,3} - \sqrt{P} c_2 h_{1,4} + n_3^1, \\
 r_4^1 &= -\sqrt{P} c_3^* h_{1,2} + \sqrt{P} c_2^* h_{1,3} + \sqrt{P} c_1 h_{1,4} + n_4^1,
 \end{aligned} \quad (8)$$

where  $c_1$ ,  $c_2$ , and  $c_3$  are the source symbols satisfying  $\mathbb{E}\{|c_1|^2\} = \mathbb{E}\{|c_2|^2\} = \mathbb{E}\{|c_3|^2\} = 1$ , and  $\mathbb{E}\{\cdot\}$  represents expectation operation.  $n_1^1$ ,  $n_2^1$ ,  $n_3^1$ , and  $n_4^1$  denote independent complex Gaussian noises with zero mean and variance  $\sigma^2$ .

By applying the combination scheme in Ref. [22], we have

$$\begin{aligned}
 \tilde{c}_1 &= \frac{r_1^1 h_{1,1}^* + r_2^1 h_{1,2}^* + r_3^1 h_{1,3}^* + r_4^1 h_{1,4}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2\right) \sqrt{P}} = c_1 + \\
 &\frac{n_1^1 h_{1,1}^* + n_2^1 h_{1,2}^* + n_3^1 h_{1,3}^* + n_4^1 h_{1,4}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2\right) \sqrt{P}},
 \end{aligned} \quad (9)$$

$$\begin{aligned}
 \tilde{c}_2 &= \frac{r_1^1 h_{1,2}^* - r_2^1 h_{1,1}^* - r_3^1 h_{1,4}^* + r_4^1 h_{1,3}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2\right) \sqrt{P}} = c_2 + \\
 &\frac{n_1^1 h_{1,2}^* - n_2^1 h_{1,1}^* - n_3^1 h_{1,4}^* + n_4^1 h_{1,3}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2\right) \sqrt{P}},
 \end{aligned}$$

$$\begin{aligned}
 \tilde{c}_3 &= \frac{r_1^1 h_{1,3}^* + r_2^1 h_{1,4}^* - r_3^1 h_{1,1}^* - r_4^1 h_{1,2}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2\right) \sqrt{P}} = c_3 + \\
 &\frac{n_1^1 h_{1,3}^* + n_2^1 h_{1,4}^* - n_3^1 h_{1,1}^* - n_4^1 h_{1,2}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2\right) \sqrt{P}},
 \end{aligned}$$

where  $\tilde{c}_1$ ,  $\tilde{c}_2$  and  $\tilde{c}_3$  represent the recovered symbols of three data streams. According to the above formulas, when the receiver has only one receiving antenna, the signal-to-noise ratio (SNR) of stream 1, stream 2 and stream 3 is

$$\begin{aligned}
 \text{SNR}_1^{\text{IRx}} = \text{SNR}_2^{\text{IRx}} = \text{SNR}_3^{\text{IRx}} &= \\
 \frac{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2\right) P}{\sigma^2}.
 \end{aligned} \quad (10)$$

Similarly, when there are two receiving antennas, according to the combining scheme in Ref. [22], there are

$$\begin{aligned}
 \tilde{c}_1 &= \frac{r_1^1 h_{1,1}^* + r_2^1 h_{1,2}^* + r_3^1 h_{1,3}^* + r_4^1 h_{1,4}^* + r_1^{2*} h_{2,1}^* + r_2^{2*} h_{2,2}^* + r_3^{2*} h_{2,3}^* + r_4^{2*} h_{2,4}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2\right) \sqrt{P}} = \\
 c_1 &+ \frac{n_1^1 h_{1,1}^* + n_2^1 h_{1,2}^* + n_3^1 h_{1,3}^* + n_4^1 h_{1,4}^* + n_1^{2*} h_{2,1}^* + n_2^{2*} h_{2,2}^* + n_3^{2*} h_{2,3}^* + n_4^{2*} h_{2,4}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2\right) \sqrt{P}},
 \end{aligned} \quad (11)$$

$$\begin{aligned}
 \tilde{c}_2 &= \frac{r_1^1 h_{1,2}^* - r_2^1 h_{1,1}^* - r_3^1 h_{1,4}^* + r_4^1 h_{1,3}^* + r_1^{2*} h_{2,2}^* - r_2^{2*} h_{2,1}^* - r_3^{2*} h_{2,4}^* + r_4^{2*} h_{2,3}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2\right) \sqrt{P}} = \\
 c_2 &+ \frac{n_1^1 h_{1,2}^* - n_2^1 h_{1,1}^* - n_3^1 h_{1,4}^* + n_4^1 h_{1,3}^* + n_1^{2*} h_{2,2}^* - n_2^{2*} h_{2,1}^* - n_3^{2*} h_{2,4}^* + n_4^{2*} h_{2,3}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2\right) \sqrt{P}},
 \end{aligned} \quad (12)$$

$$\begin{aligned}
 \tilde{c}_3 &= \frac{r_1^1 h_{1,3}^* + r_2^1 h_{1,4}^* - r_3^1 h_{1,1}^* - r_4^1 h_{1,2}^* + r_1^{2*} h_{2,3}^* + r_2^{2*} h_{2,4}^* - r_3^{2*} h_{2,1}^* - r_4^{2*} h_{2,2}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2\right) \sqrt{P}} = \\
 c_3 &+ \frac{n_1^1 h_{1,3}^* + n_2^1 h_{1,4}^* - n_3^1 h_{1,1}^* - n_4^1 h_{1,2}^* + n_1^{2*} h_{2,3}^* + n_2^{2*} h_{2,4}^* - n_3^{2*} h_{2,1}^* - n_4^{2*} h_{2,2}^*}{\left(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2\right) \sqrt{P}}.
 \end{aligned} \quad (13)$$

Therefore, when there are two receiving antennas at the receiver, the SNR of stream 1, stream 2 and stream 3 is

$$\text{SNR}_1^{2\text{Rx}} = \text{SNR}_2^{2\text{Rx}} = \text{SNR}_3^{2\text{Rx}} = \frac{\left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right) P}{\sigma^2}. \quad (14)$$

From the two SNR formulas of Eqs. (10) and (14), it can be found that the 4-transmit space-time transmission scheme based on the dual-polarized RIS proposed in this paper is basically consistent with the traditional space-time transmission system. The same diversity order and diversity gain can be obtained.

## 4 Polarization Coupling Modeling and BER Performance Analysis

### 4.1 Polarization Coupling Modeling for Dual-Polarized RIS

The authors of Ref. [19] have preliminarily revealed the phenomenon of polarization coupling in the unit cell of the dual-polarized RIS. Based on the coupling constellation and the coupling voltage of the unit cell given in Ref. [19], a linear symbolic coupling model is proposed as follows to analyze the polarization coupling phenomenon in the 4-transmit space-time transmission based on the dual-polarized RIS.

Theorem 1. The coupling model between two polarization directions of the unit cell  $U_{n,m}$  can be written as

$$\begin{cases} x_{couple}^{0,n,m} = x_{0,n,m} + ax_{1,n,m} \\ x_{couple}^{1,n,m} = x_{1,n,m} + bx_{0,n,m} \end{cases}, \quad (15)$$

where  $a$  and  $b$  are the coefficients coupled to the horizontal and vertical polarization,  $x_{0,n,m}$  and  $x_{1,n,m}$  are the original transmitted symbols in horizontal and vertical polarization, and  $x_{couple}^{0,n,m}$  and  $x_{couple}^{1,n,m}$  are the actual transmitted symbols in horizontal and vertical polarization.

Theorem 1 reveals that the transmitted symbols of two polarization directions are mutually coupled in one unit cell of the dual-polarized RIS. In particular, Theorem 1 indicates that the actual transmitted symbols are the linear combination of original transmitted symbols and coupled transmitted symbols from another polarization direction.

Based on the system model in Section 2, when the system operates normally and has no polarization coupling interference, the signal received from  $U_{n,m}$  by the  $k$ -th receiving antenna in polarization  $p$  is  $y_k^{p,n,m} = h_k^{p,n,m} \sqrt{P} x_{p,n,m}$ .

According to Theorem 1, the signal that considers polarization coupling interference received from  $U_{n,m}$  by the  $k$ -th receiving antenna in polarization  $p$  is  $y_{couple}^{k,p,n,m} = h_k^{p,n,m} \sqrt{P} x_{p,n,m} + qh_k^{p,n,m} \sqrt{P} x_{1-p,n,m}$ , where  $x_{1-p,n,m}$  is the transmitted symbol in

another polarization, and  $q$  is the coupling coefficient between unit cell's control voltages for horizontal and vertical polarization directions. According to Ref. [19],  $q = a = \frac{1.01}{6}$  is the coefficient coupled to the horizontal polarization and  $q = b = \frac{0.92}{8}$  is the coefficient coupled to the vertical polarization for the employed dual-polarized RIS.

### 4.2 BER Performance Without Polarization Coupling Interference

For the sake of analysis, it is assumed that the wireless channel parameters  $(h_{1,1}, h_{1,2}, h_{1,3}, h_{1,4}, h_{2,1}, h_{2,2}, h_{2,3}, h_{2,4})$  remain constant during transmission. Due to the standardly mapped 16-QAM with Gray coding and the traditional QAM demodulation method, the theoretical total BER of the 4-transmit space-time RIS-based transmission system can be expressed as<sup>[17, 23]</sup>:

$$\text{BER}_{1\text{Rx}} \cong \frac{3}{8} \text{erfc} \left( \sqrt{\frac{4\text{SNR}_{\text{Rx1}}}{30}} \right) + \frac{1}{4} \text{erfc} \left( 3 \sqrt{\frac{4\text{SNR}_{\text{Rx1}}}{30}} \right), \quad (16)$$

$$\text{SNR}_{\text{Rx1}} = \frac{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right) P}{4\sigma^2}, \quad (17)$$

where  $\text{erfc}(\cdot)$  is the complementary error function and  $\text{SNR}_{\text{Rx1}}$  is the received SNR for one receiving antenna.

Similarly, the system BER with two receiving antennas is<sup>[17, 23]</sup>:

$$\text{BER}_{2\text{Rx}} \cong \frac{3}{8} \text{erfc} \left( \sqrt{\frac{4\beta \text{SNR}_{\text{Rx1}}}{30}} \right) + \frac{1}{4} \text{erfc} \left( 3 \sqrt{\frac{4\beta \text{SNR}_{\text{Rx1}}}{30}} \right), \quad (18)$$

$\beta =$

$$\frac{|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2}{|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2}, \quad (19)$$

where  $\beta$  is the diversity gain obtained when the receiver has two receiving antennas.

### 4.3 BER Performance Considering Polarization Coupling Interference

Take the horizontal polarization as an example, when the symbols  $X_1, X_2, X_3$  and  $X_4$  are respectively transmitted by the yellow, orange, blue and indigo parts of RIS as shown in Fig. 1 in a certain time slot, the signal received by the  $k$ -th receiving antenna in horizontal polarization from the yellow part is

$$\begin{aligned}
 y_{yellow}^{k,0} &= \sum_{m=1}^{\frac{M}{2}} \sum_{n=1}^N h_k^{0,n,m} \sqrt{P} X_1 + \\
 &a \sum_{m=1}^{\frac{M}{2}} \sum_{n=\frac{N}{2}+1}^N h_k^{0,n,m} \sqrt{P} X_3 + a \sum_{m=1}^{\frac{M}{2}} \sum_{n=1}^{\frac{N}{2}} h_k^{0,n,m} \sqrt{P} X_4 = \\
 &h_{k,1} \sqrt{P} X_1 + a \sum_{m=1}^{\frac{M}{2}} \sum_{n=\frac{N}{2}+1}^N h_k^{0,n,m} \sqrt{P} X_3 + a \sum_{m=1}^{\frac{M}{2}} \sum_{n=1}^{\frac{N}{2}} h_k^{0,n,m} \sqrt{P} X_4 = \\
 &h_{k,1} \sqrt{P} X_1 + ah_k^{upper,yellow} \sqrt{P} X_3 + ah_k^{lower,yellow} \sqrt{P} X_4, \quad (20)
 \end{aligned}$$

where  $h_k^{upper,yellow} = \sum_{m=1}^{\frac{M}{2}} \sum_{n=\frac{N}{2}+1}^N h_k^{0,n,m}$  and  $h_k^{lower,yellow} = \sum_{m=1}^{\frac{M}{2}} \sum_{n=1}^{\frac{N}{2}} h_k^{0,n,m}$

are the channels from the upper region and lower region of the yellow part to the  $k$ -th receiving antenna in horizontal polarization.

And the signal received by the  $k$ -th receiving antenna in horizontal polarization from the orange part is

$$\begin{aligned}
 y_{orange}^{k,0} &= \sum_{m=\frac{M}{2}+1}^M \sum_{n=1}^N h_k^{0,n,m} \sqrt{P} X_2 + \\
 &a \sum_{m=\frac{M}{2}+1}^M \sum_{n=\frac{N}{2}+1}^N h_k^{0,n,m} \sqrt{P} X_3 + a \sum_{m=\frac{M}{2}+1}^M \sum_{n=1}^{\frac{N}{2}} h_k^{0,n,m} \sqrt{P} X_4 = \\
 &h_{k,2} \sqrt{P} X_2 + a \sum_{m=\frac{M}{2}+1}^M \sum_{n=\frac{N}{2}+1}^N h_k^{0,n,m} \sqrt{P} X_3 + \\
 &a \sum_{m=\frac{M}{2}+1}^M \sum_{n=1}^{\frac{N}{2}} h_k^{0,n,m} \sqrt{P} X_4 = \\
 &h_{k,2} \sqrt{P} X_2 + ah_k^{upper,orange} \sqrt{P} X_3 + ah_k^{lower,orange} \sqrt{P} X_4, \quad (21)
 \end{aligned}$$

where  $h_k^{upper,orange} = \sum_{m=\frac{M}{2}+1}^M \sum_{n=\frac{N}{2}+1}^N h_k^{0,n,m}$  and  $h_k^{lower,orange} = \sum_{m=\frac{M}{2}+1}^M \sum_{n=1}^{\frac{N}{2}} h_k^{0,n,m}$  are the channels from the upper region and

lower region of the orange part to the  $k$ -th receiving antenna in horizontal polarization.

Due to the proposed wireless frame structure and transmission scheme,  $h_k^{upper,yellow}$ ,  $h_k^{lower,yellow}$ ,  $h_k^{upper,orange}$  and  $h_k^{lower,orange}$  cannot be estimated and obtained. Because receiving antennas are placed approximately on the central normal line of the RIS, we can assume that  $h_k^{upper,yellow} \approx \frac{1}{2} h_{k,1}$ ,  $h_k^{lower,yellow} \approx \frac{1}{2} h_{k,1}$ ,  $h_k^{upper,orange} \approx \frac{1}{2} h_{k,2}$ , and  $h_k^{lower,orange} \approx \frac{1}{2} h_{k,2}$ , and Eqs. (18) and (19) can be written as

$$\begin{aligned}
 y_{yellow}^{k,0} &= h_{k,1} \sqrt{P} X_1 + a \sum_{m=1}^{\frac{M}{2}} \sum_{n=\frac{N}{2}+1}^N h_k^{0,n,m} \sqrt{P} X_3 + \\
 &a \sum_{m=1}^{\frac{M}{2}} \sum_{n=1}^{\frac{N}{2}} h_k^{0,n,m} \sqrt{P} X_4 \approx h_{k,1} \sqrt{P} X_1 + \\
 &a \left( \frac{1}{2} h_{k,1} \right) \sqrt{P} X_3 + a \left( \frac{1}{2} h_{k,1} \right) \sqrt{P} X_4 = \\
 &h_{k,1} \sqrt{P} \left( X_1 + \frac{1}{2} a X_3 + \frac{1}{2} a X_4 \right), \quad (22)
 \end{aligned}$$

and

$$y_{orange}^{k,0} \approx h_{k,2} \sqrt{P} \left( X_2 + \frac{1}{2} a X_3 + \frac{1}{2} a X_4 \right). \quad (23)$$

Similarly, for the blue and indigo parts, we have

$$y_{blue}^{k,1} \approx h_{k,3} \sqrt{P} \left( X_3 + \frac{1}{2} b X_1 + \frac{1}{2} b X_2 \right), \quad (24)$$

and

$$y_{indigo}^{k,1} \approx h_{k,4} \sqrt{P} \left( X_4 + \frac{1}{2} b X_1 + \frac{1}{2} b X_2 \right). \quad (25)$$

For the  $k$ -th receiving antenna, the received signal in a certain time slot is

$$\begin{aligned}
 y_k &= y_{yellow}^{k,0} + y_{orange}^{k,0} + y_{blue}^{k,1} + y_{indigo}^{k,1} \\
 &\approx \sqrt{P} \begin{bmatrix} h_{k,1} \left( X_1 + \frac{1}{2} a X_3 + \frac{1}{2} a X_4 \right) + \\ h_{k,2} \left( X_2 + \frac{1}{2} a X_3 + \frac{1}{2} a X_4 \right) + \\ h_{k,3} \left( X_3 + \frac{1}{2} b X_1 + \frac{1}{2} b X_2 \right) + h_{k,4} \left( X_4 + \frac{1}{2} b X_1 + \frac{1}{2} b X_2 \right) \end{bmatrix} = \\
 &\sqrt{P} \mathbf{h}_k \mathbf{X} \mathbf{Q}_{couple} = \sqrt{P} \mathbf{h}_k \mathbf{X}_{couple}, \quad (26)
 \end{aligned}$$

where  $\mathbf{h}_k = [h_{k,1}, h_{k,2}, h_{k,3}, h_{k,4}]^T$ ,  $\mathbf{X} = [X_1, X_2, X_3, X_4]$ , and

$$\mathbf{Q}_{couple} = \begin{bmatrix} 1 & 0 & \frac{1}{2} b & \frac{1}{2} b \\ 0 & 1 & \frac{1}{2} b & \frac{1}{2} b \\ \frac{1}{2} a & \frac{1}{2} a & 1 & 0 \\ \frac{1}{2} a & \frac{1}{2} a & 0 & 1 \end{bmatrix}, \quad (27)$$

$$\mathbf{X}_{couple} = \mathbf{X}\mathbf{Q}_{couple} = \begin{bmatrix} X_1 + \frac{1}{2}aX_3 + \frac{1}{2}aX_4, X_2 + \frac{1}{2}aX_3 + \frac{1}{2}aX_4, X_3 + \frac{1}{2}bX_1 + \frac{1}{2}bX_2, X_4 + \frac{1}{2}bX_1 + \frac{1}{2}bX_2 \end{bmatrix}.$$

When the RIS-based transmitter transmits data according to the STBC matrix  $\mathbf{c}$  of Eq. (2), we let  $u = \frac{1}{2}a$ ,  $v = \frac{1}{2}b$  and there is

$$\mathbf{c}_{couple} = \mathbf{c}\mathbf{Q}_{couple} = \begin{bmatrix} c_1 + uc_3 & c_2 + uc_3 & c_3 + vc_1 + vc_2 & vc_1 + vc_2 \\ -c_3^* + uc_3^* & c_1^* + uc_3^* & -vc_2^* + vc_1^* & c_3 - vc_2^* + vc_1^* \\ -c_3^* + uc_1^* - uc_2^* & uc_1^* - uc_2^* & c_1^* - vc_3^* & -c_2 - vc_3^* \\ uc_2^* + uc_1^* & -c_3^* + uc_2^* + uc_1^* & c_2^* - vc_3^* & c_1 - vc_3^* \end{bmatrix}. \quad (28)$$

Then we have

$$\mathbf{r}_{couple}^k = \sqrt{P} \mathbf{c}_{couple} \mathbf{h}_k + \mathbf{n}_k, \quad (29)$$

where  $\mathbf{r}_{couple}^k = [r_{couple}^{k,1}, r_{couple}^{k,2}, r_{couple}^{k,3}, r_{couple}^{k,4}]^T$  is the received signal vector considering polarization coupling interference.

Eq. (27) gives the system model that considers polarization coupling interference, and the influence of the proposed symbolic coupling model is mainly reflected in the STBC matrix  $\mathbf{c}_{couple}$ . Then the BER performance of the system is discussed as follows.

**Proposition 1.** When considering polarization coupling interference and there is only one receiving antenna at the receiver, the upper bound of system BER can be written as

$$\text{BER}_{upbound}^{1\text{Rx}} \cong \frac{3}{8} \text{erfc} \left( \sqrt{\frac{\text{SINR}^{1\text{Rx}}}{10}} \right) + \frac{1}{4} \text{erfc} \left( 3 \sqrt{\frac{\text{SINR}^{1\text{Rx}}}{10}} \right), \quad (30)$$

$$\text{SINR}^{1\text{Rx}} = \frac{1}{\frac{3}{4\text{SNR}_{\text{Rx1}}} + \frac{4(u^2|h_{1,1}|^2 + u^2|h_{1,2}|^2 + v^2|h_{1,3}|^2 + v^2|h_{1,4}|^2)}{3(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2)} + \frac{(u^2 + v^2)(|h_{1,1}|^2 + |h_{1,2}|^2)(|h_{1,3}|^2 + |h_{1,4}|^2)}{3(|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2)^2}}, \quad (31)$$

where  $\text{SINR}^{1\text{Rx}}$  is the signal-to-interference plus noise ratio (SINR) of the data stream with a single receiving antenna.

**Proof:** See Appendix A.

Eq. (28) in Proposition 1 and Eq. (16) have the same structure, but the difference is that Eq. (28) is related to SINR due to the polarization coupling interference while Eq. (16) is just related to SNR. Proposition 1 reveals that the BER of the proposed system is not only influenced by the SNR of the system,

but also by polarization coupling interference. When the interference terms in SINR are 0, Eq. (28) will degenerate into Eq. (16). However, because of the interference terms, the SINR of the system will decrease, which leads to the deterioration of the system BER performance and this is consistent with the intuition.

For the case of two receiving antennas, we can get Proposition 2 as follows.

**Proposition 2:** When considering polarization coupling interference and there are two receiving antennas at the receiver, the upper bound of system BER can be written as

$$\text{BER}_{upbound}^{2\text{Rx}} \cong \frac{3}{8} \text{erfc} \left( \sqrt{\frac{\text{SINR}^{2\text{Rx}}}{10}} \right) + \frac{1}{4} \text{erfc} \left( 3 \sqrt{\frac{\text{SINR}^{2\text{Rx}}}{10}} \right), \quad (32)$$

$$\text{SINR}^{2\text{Rx}} = \frac{1}{\frac{3}{4\beta\text{SNR}_{\text{Rx1}}} + d_1 + d_2 + d_3}, \quad (33)$$

where  $\text{SINR}^{2\text{Rx}}$  is the SINR of system data stream with two receiving antennas and

$$d_1 = \frac{4 \left[ (u^2|h_{1,1}|^2 + u^2|h_{1,2}|^2 + v^2|h_{1,3}|^2 + v^2|h_{1,4}|^2) (|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2) \right]}{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right)^2},$$

$$d_2 = \frac{4 \left[ (u^2|h_{2,1}|^2 + u^2|h_{2,2}|^2 + v^2|h_{2,3}|^2 + v^2|h_{2,4}|^2) (|h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2) \right]}{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right)^2},$$

$$d_3 = \frac{(u^2 + v^2) \left[ (|h_{1,1}|^2 + |h_{1,2}|^2) (|h_{1,3}|^2 + |h_{1,4}|^2) + (|h_{2,1}|^2 + |h_{2,2}|^2) (|h_{2,3}|^2 + |h_{2,4}|^2) \right]}{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right)^2}.$$

**Proof:** See Appendix B.

According to Proposition 2, diversity gain  $\beta$  exists in both SNR terms and interference terms ( $d_1, d_2, d_3$ ), which means that even if more receiving antennas are equipped, polarization coupling interference will still prominently affect the BER performance of the system. Besides, Eqs. (28) and (29) also indicate that if we want to improve the BER performance, we need to reduce coupling coefficients  $u$  and  $v$ , which leads to less polarization coupling interference. And that puts forward new requirements for the design of the unit cell of a dual-polarized RIS. In the following section, we will validate the proposed symbolic polarization coupling model by the experi-

mental measurements based on our prototype system.

## 5 Implementation of 4-Transmit Space-Time Transmission Based on Dual-Polarized RIS

### 5.1 Prototype Setup

Fig. 3 shows our prototype of 4-transmit space-time transmission based on the dual-polarized RIS. The system consists of a 4-transmit space-time RIS-based transmitter and a traditional receiver. The 4-transmit space-time transmitter is composed of a dual-polarized phase-adjustable RIS that is controlled by a PXIe platform, a horn antenna and an RF signal source. The traditional receiver consists of two receiving antennas and a software defined radio (SDR) platform.

The dual-polarized RIS is a phase programmable metasurface with an operating frequency of 2.7 GHz. There are 144 (12×12) unit cells in total, and each unit cell has two pairs of metal patches and two varactor diodes are connected across it. The bias voltage is applied to the varactor diodes through the corresponding metal patches to control the reflection coefficient of the unit cell in two polarization directions<sup>[18-19]</sup>. The RIS fabricated with varactors can continuously regulate the phase, which makes the transmission waveform design in Section 4 achievable.

Specifically, three bit streams are generated by the PXIe platform, mapped into three 16-QAM symbol streams, and encoded by the STBC matrix to generate four baseband frames. The four frames are converted into four analog voltage signals by the digital-to-analog converters (DACs) of the PXIe platform and loaded on the varactors of the corresponding RIS's unit cells respectively. When the 2.7 GHz single tone carrier generated by the RF signal source is transmitted and irradiated on the RIS through the horn antenna, the dual-polarized RIS modulates four data frames onto the carrier. The transmitting horn antenna is placed at an inclination of 45° so that there are carrier components in both horizontal and vertical polarization directions. The receiver receives the signal

through two single polarization antennas inclined at 45°. The received signal is down converted and sampled on the SDR platform, and the corresponding baseband data symbols are then decoded and demodulated.

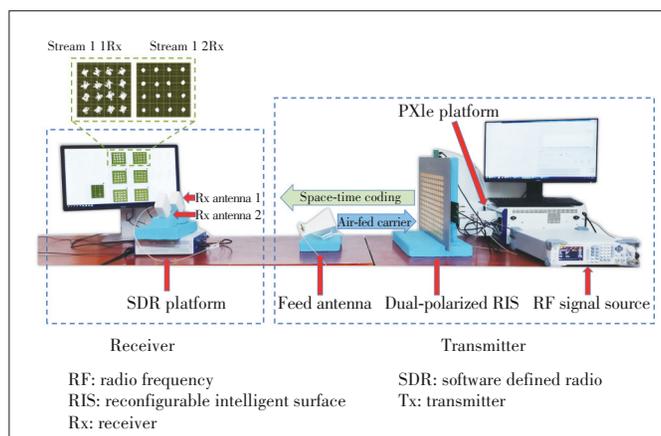
### 5.2 Experimental Results

As shown in Fig. 3, the 4-transmit space-time transmission system based on the dual-polarized RIS can successfully realize data transmission. The prototype is in an indoor environment with a transmission distance of about 1.5 m. When the symbol rate of the system is 2.5 million symbol/s, the transmission rate can reach 7.5 Mbit/s, which is determined by the adopted space-time coding scheme. It can be seen from Fig. 3 that the constellations of three data streams at the receiver can be well recovered. By adjusting the transmission power to change the received SNR, the BER performance is recorded simultaneously. The measurement results are shown in Fig. 4. In Fig. 4, the blue line is the theoretical BER curve without polarization coupling interference, the red line is the upper bound of the theoretical BER considering polarization coupling interference, and the orange line is the actual measured BER curve. It can be seen that the measured BER is obviously worse than the theoretical one without polarization coupling interference. This is because the control voltages regulating the unit cell's phase in different polarization directions are coupled<sup>[19]</sup>, which makes the four data frames doped with each other in the transmission process, resulting in additional interference and degradation of the BER performance of the actual system. Fig. 4 also shows that the theoretical BER upper bounds given in Eqs. (28) and (29) are in good agreement with the measured curve, which validates our proposed linear polarization coupling model for the dual-polarized RIS.

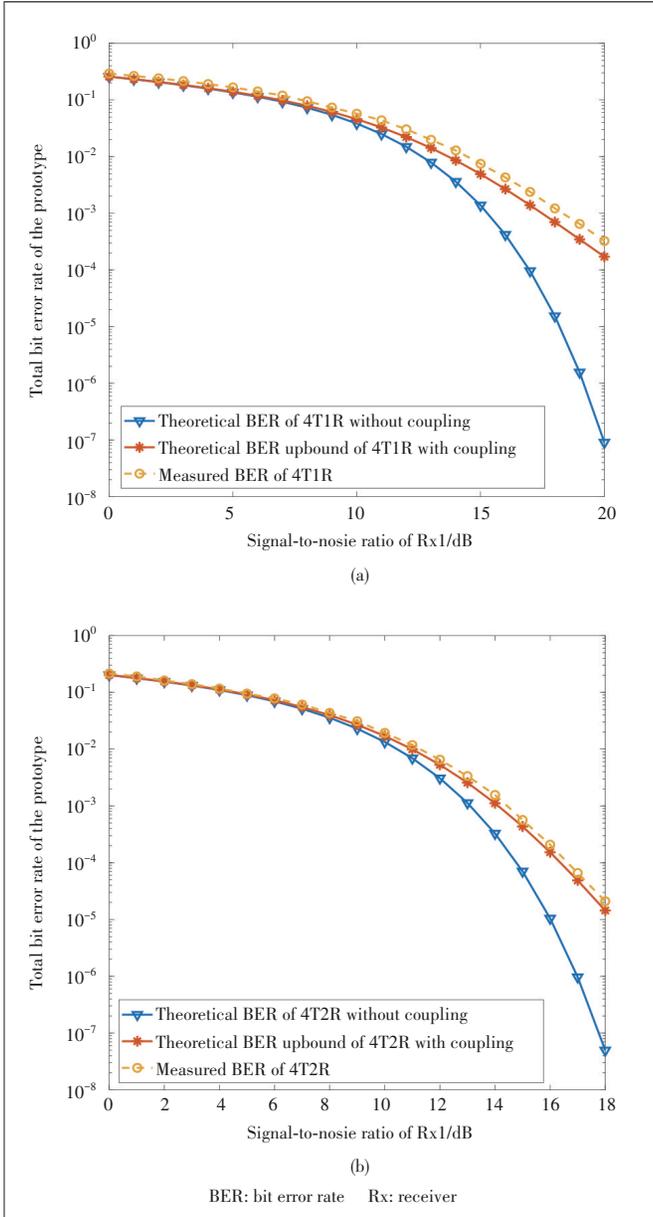
The above results indicate that the 4-transmit space-time transmission scheme based on the dual-polarized phase-adjustable RIS is feasible, but the mutual coupling of control voltages of the unit cell for different polarization directions will degrade the system performance. This enlightens that the polarization isolation performance should be carefully considered when designing dual-polarized RISs, and it is worth further exploring in the future.

## 6 Conclusions

In this paper, a 4-transmit space-time transmission scheme based on the dual-polarized RIS is proposed. The prototype system based on a dual-polarized phase-adjustable RIS can transmit data over the air at a rate of 7.5 Mbit/s, which verifies the feasibility of the proposed scheme. In particular, a simple linear symbolic polarization coupling model is proposed for theoretical analysis, and the BER performance upper bound of the system is given. The measured curves are in good agreement with the theoretical ones, which proves the effectiveness of the proposed model and scheme. In general, a multi-transmit space-time transmitter based on the dual-polarized



▲ Figure 3. Prototype of 4-transmit space-time transmission based on dual-polarized RIS



▲ Figure 4. Comparison of system theoretical and measured BER-SNR performance: (a) one receiving antenna and (b) two receiving antennas

RIS is a promising scheme, which can flexibly build different space-time transmission systems to satisfy the transmission requirements with low hardware complexity and cost.

## Appendix A

### Proof of Proposition 1

The STBC matrix considering polarization coupling interference is shown in Eq. (26). When  $K=1$ , which means that the receiver has only one receiving antenna, it can be obtained from Eq. (27) that

$$r_{couple}^{1,1} = \sqrt{P} h_{1,1} (c_1 + uc_3) + \sqrt{P} h_{1,2} (c_2 + uc_3) + \sqrt{P} h_{1,3} (c_3 + vc_1 + vc_2) + \sqrt{P} h_{1,4} (vc_1 + vc_2) + n_1^1,$$

$$r_{couple}^{1,2} = \sqrt{P} h_{1,1} (-c_2^* + uc_3) + \sqrt{P} h_{1,2} (c_1^* + uc_3) + \sqrt{P} h_{1,3} (-vc_2^* + vc_1^*) + \sqrt{P} h_{1,4} (c_3 - vc_2^* + vc_1^*) + n_2^1,$$

$$r_{couple}^{1,3} = \sqrt{P} h_{1,1} (-c_3^* + uc_1^* - uc_2) + \sqrt{P} h_{1,2} (uc_1^* - uc_2) + \sqrt{P} h_{1,3} (c_1^* - vc_3^*) + \sqrt{P} h_{1,4} (-c_2 - vc_3^*) + n_3^1,$$

$$r_{couple}^{1,4} = \sqrt{P} h_{1,1} (uc_2^* + uc_1) + \sqrt{P} h_{1,2} (-c_3^* + uc_2^* + uc_1) + \sqrt{P} h_{1,3} (c_2^* - vc_3^*) + \sqrt{P} h_{1,4} (c_1 - vc_3^*) + n_4^1.$$

According to Ref. [22], there are the following combination formulas.

$$\tilde{c}_{1,couple} = r_{couple}^{1,1} h_{1,1}^* + r_{couple}^{1,2*} h_{1,2} + r_{couple}^{1,3*} h_{1,3} + r_{couple}^{1,4} h_{1,4}^*,$$

$$\tilde{c}_{2,couple} = r_{couple}^{1,1} h_{1,2}^* - r_{couple}^{1,2*} h_{1,1} - r_{couple}^{1,3*} h_{1,4}^* + r_{couple}^{1,4*} h_{1,3}^*,$$

$$\tilde{c}_{3,couple} = r_{couple}^{1,1} h_{1,3}^* + r_{couple}^{1,2*} h_{1,4}^* - r_{couple}^{1,3*} h_{1,1} - r_{couple}^{1,4*} h_{1,2}^*.$$

By substituting  $r_{couple}^{1,1}$ ,  $r_{couple}^{1,2}$ ,  $r_{couple}^{1,3}$ , and  $r_{couple}^{1,4}$  into the combination formula of  $\tilde{c}_{1,couple}$ , we have:

$$\begin{aligned} \tilde{c}_{1,couple} &= r_{couple}^{1,1} h_{1,1}^* + r_{couple}^{1,2*} h_{1,2} + r_{couple}^{1,3*} h_{1,3} + r_{couple}^{1,4} h_{1,4}^* = \\ &= \sqrt{P} \left[ (c_1 + uc_3) |h_{1,1}|^2 + (c_2 + uc_3) h_{1,2} h_{1,1}^* + \right. \\ &\quad \left. (c_3 + vc_1 + vc_2) h_{1,3} h_{1,1}^* + (vc_1 + vc_2) h_{1,4} h_{1,1}^* \right] + n_1^1 h_{1,1}^* = \\ &= \sqrt{P} \left[ (-c_2^* + uc_3) h_{1,1}^* h_{1,2} + (c_1^* + uc_3) |h_{1,2}|^2 + \right. \\ &\quad \left. (-vc_2^* + vc_1^*) h_{1,3} h_{1,2} + (c_3 - vc_2^* + vc_1^*) h_{1,4} h_{1,2} \right] + n_2^1 h_{1,2} + \\ &= \sqrt{P} \left[ (-c_3^* + uc_1^* - uc_2) h_{1,1}^* h_{1,3} + (uc_1^* - uc_2) h_{1,2}^* h_{1,3} + \right. \\ &\quad \left. (c_1^* - vc_3^*) |h_{1,3}|^2 + (-c_2 - vc_3^*) h_{1,4}^* h_{1,3} \right] + n_3^1 h_{1,3} + \\ &= \sqrt{P} \left[ (uc_2^* + uc_1) h_{1,1} h_{1,4}^* + (-c_3^* + uc_2^* + uc_1) h_{1,2} h_{1,4}^* + \right. \\ &\quad \left. (c_2^* - vc_3^*) h_{1,3} h_{1,4}^* + (c_1 - vc_3^*) |h_{1,4}|^2 \right] + n_4^1 h_{1,4}^* . \end{aligned}$$

Sort out the expression of  $\tilde{c}_{1,couple}$  and we have

$$\begin{aligned} \tilde{c}_{1,couple} &= c_1 \sqrt{P} \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right) + \\ &c_1 \sqrt{P} \left[ u \left( h_{1,2} h_{1,4}^* + h_{1,1}^* h_{1,3} + h_{1,1} h_{1,4}^* + h_{1,2}^* h_{1,3} \right) + \right. \\ &\left. v \left( h_{1,2} h_{1,4}^* + h_{1,1}^* h_{1,3} + h_{1,3}^* h_{1,2} + h_{1,4} h_{1,1}^* \right) \right] + \\ &c_2 \sqrt{P} v \left( h_{1,1}^* h_{1,3} + h_{1,4} h_{1,1}^* - h_{1,3}^* h_{1,2} - h_{1,2} h_{1,4}^* \right) + \\ &c_2^* \sqrt{P} u \left( h_{1,1} h_{1,4}^* + h_{1,2} h_{1,4}^* - h_{1,1}^* h_{1,3} - h_{1,2}^* h_{1,3} \right) + \\ &c_3 \sqrt{P} \left[ u \left( |h_{1,1}|^2 + h_{1,1}^* h_{1,2} \right) - v \left( |h_{1,3}|^2 + h_{1,3} h_{1,4}^* \right) \right] + \\ &c_3^* \sqrt{P} \left[ u \left( |h_{1,2}|^2 + h_{1,1}^* h_{1,2} \right) - v \left( |h_{1,4}|^2 + h_{1,3} h_{1,4}^* \right) \right] + \\ &n_1^1 h_{1,1}^* + n_2^1 h_{1,2} + n_3^1 h_{1,3} + n_4^1 h_{1,4}^* . \end{aligned}$$

$$\begin{aligned} \text{Let } \tilde{c}_{signal}^{1,couple} &= c_1 \sqrt{P} \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right), \\ \tilde{c}_{noise}^{1,couple} &= n_1^1 h_{1,1}^* + n_2^1 h_{1,2} + n_3^1 h_{1,3} + n_4^1 h_{1,4}^* , \end{aligned}$$

$$\begin{aligned} \tilde{c}_{interference}^{1,couple} &= c_1 \sqrt{P} \left[ u \left( h_{1,2} h_{1,4}^* + h_{1,1}^* h_{1,3} + h_{1,1} h_{1,4}^* + h_{1,2}^* h_{1,3} \right) + \right. \\ &\left. v \left( h_{1,2} h_{1,4}^* + h_{1,1}^* h_{1,3} + h_{1,3}^* h_{1,2} + h_{1,4} h_{1,1}^* \right) \right] + \\ &c_2 \sqrt{P} v \left( h_{1,1}^* h_{1,3} + h_{1,4} h_{1,1}^* - h_{1,3}^* h_{1,2} - h_{1,2} h_{1,4}^* \right) + \\ &c_2^* \sqrt{P} u \left( h_{1,1} h_{1,4}^* + h_{1,2} h_{1,4}^* - h_{1,1}^* h_{1,3} - h_{1,2}^* h_{1,3} \right) + \\ &c_3 \sqrt{P} \left[ u \left( |h_{1,1}|^2 + h_{1,1}^* h_{1,2} \right) - v \left( |h_{1,3}|^2 + h_{1,3} h_{1,4}^* \right) \right] + \\ &c_3^* \sqrt{P} \left[ u \left( |h_{1,2}|^2 + h_{1,1}^* h_{1,2} \right) - v \left( |h_{1,4}|^2 + h_{1,3} h_{1,4}^* \right) \right] . \end{aligned}$$

From the inequality property  $(x_1 + x_2)^2 \geq x_1^2 + x_2^2$ , we can get

$$\begin{aligned} &\left| u \left( h_{1,2} h_{1,4}^* + h_{1,1}^* h_{1,3} + h_{1,1} h_{1,4}^* + h_{1,2}^* h_{1,3} \right) + \right. \\ &\left. v \left( h_{1,2} h_{1,4}^* + h_{1,1}^* h_{1,3} + h_{1,3}^* h_{1,2} + h_{1,4} h_{1,1}^* \right) \right|^2 \geq \\ &(u^2 + v^2) \left( |h_{1,2}|^2 |h_{1,4}|^2 + |h_{1,1}|^2 |h_{1,3}|^2 + \right. \\ &\left. |h_{1,1}|^2 |h_{1,4}|^2 + |h_{1,2}|^2 |h_{1,3}|^2 \right) , \end{aligned}$$

$$\begin{aligned} &\left| h_{1,1}^* h_{1,3} + h_{1,4} h_{1,1}^* - h_{1,3}^* h_{1,2} - h_{1,2} h_{1,4}^* \right|^2 \geq |h_{1,2}|^2 |h_{1,4}|^2 + \\ &|h_{1,1}|^2 |h_{1,3}|^2 + |h_{1,1}|^2 |h_{1,4}|^2 + |h_{1,2}|^2 |h_{1,3}|^2 , \end{aligned}$$

$$\begin{aligned} &\left| h_{1,1} h_{1,4}^* + h_{1,2} h_{1,4}^* - h_{1,1}^* h_{1,3} - h_{1,2}^* h_{1,3} \right|^2 \geq |h_{1,2}|^2 |h_{1,4}|^2 + \\ &|h_{1,1}|^2 |h_{1,3}|^2 + |h_{1,1}|^2 |h_{1,4}|^2 + |h_{1,2}|^2 |h_{1,3}|^2 , \end{aligned}$$

$$\begin{aligned} &\left| u \left( |h_{1,1}|^2 + h_{1,1}^* h_{1,2} \right) - v \left( |h_{1,3}|^2 + h_{1,3} h_{1,4}^* \right) \right|^2 \geq u^2 \left( |h_{1,1}|^4 + \right. \\ &\left. |h_{1,1}|^2 |h_{1,2}|^2 \right) + v^2 \left( |h_{1,3}|^4 + |h_{1,3}|^2 |h_{1,4}|^2 \right) , \end{aligned}$$

$$\begin{aligned} &\left| u \left( |h_{1,2}|^2 + h_{1,1}^* h_{1,2} \right) - v \left( |h_{1,4}|^2 + h_{1,3} h_{1,4}^* \right) \right|^2 \geq u^2 \left( |h_{1,2}|^4 + \right. \\ &\left. |h_{1,1}|^2 |h_{1,2}|^2 \right) + v^2 \left( |h_{1,4}|^4 + |h_{1,3}|^2 |h_{1,4}|^2 \right) . \end{aligned}$$

Then we have

$$\begin{aligned} P_{interference}^{\tilde{c}_{1,couple}} &= E \left[ \left( \tilde{c}_{interference}^{1,couple} \right)^2 \right] \geq \\ &P \left[ 2(u^2 + v^2) \left( |h_{1,2}|^2 |h_{1,4}|^2 + |h_{1,1}|^2 |h_{1,3}|^2 + |h_{1,1}|^2 |h_{1,4}|^2 \right) + \right. \\ &\left. |h_{1,2}|^2 |h_{1,3}|^2 \right. \\ &\left. u^2 \left( |h_{1,1}|^4 + |h_{1,1}|^2 |h_{1,2}|^2 \right) + v^2 \left( |h_{1,3}|^4 + |h_{1,3}|^2 |h_{1,4}|^2 \right) + \right. \\ &\left. u^2 \left( |h_{1,2}|^4 + |h_{1,1}|^2 |h_{1,2}|^2 \right) + v^2 \left( |h_{1,4}|^4 + |h_{1,3}|^2 |h_{1,4}|^2 \right) \right] = \\ &P \left[ u^2 |h_{1,1}|^4 + u^2 |h_{1,2}|^4 + v^2 |h_{1,3}|^4 + v^2 |h_{1,4}|^4 + \right. \\ &2u^2 \left( |h_{1,1}|^2 |h_{1,2}|^2 + |h_{1,1}|^2 |h_{1,3}|^2 + |h_{1,1}|^2 |h_{1,4}|^2 + \right. \\ &\left. |h_{1,2}|^2 |h_{1,3}|^2 + |h_{1,2}|^2 |h_{1,4}|^2 \right) + \\ &\left. 2v^2 \left( |h_{1,1}|^2 |h_{1,2}|^2 + |h_{1,1}|^2 |h_{1,3}|^2 + |h_{1,1}|^2 |h_{1,4}|^2 + \right. \right. \\ &\left. \left. |h_{1,2}|^2 |h_{1,3}|^2 + |h_{1,2}|^2 |h_{1,4}|^2 \right) \right] = \\ &P(u^2 + v^2) \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right) \left( |h_{1,3}|^2 + |h_{1,4}|^2 \right) + \\ &P \left( u^2 |h_{1,1}|^2 + u^2 |h_{1,2}|^2 + v^2 |h_{1,3}|^2 + \right. \\ &\left. v^2 |h_{1,4}|^2 \right) \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right) , \end{aligned}$$

$$P_{signal}^{\tilde{c}_{1,couple}} = E \left[ \left( \tilde{c}_{signal}^{1,couple} \right)^2 \right] = P \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right)^2 ,$$

$$P_{noise}^{\tilde{c}_{1,couple}} = E \left[ \left( \tilde{c}_{noise}^{1,couple} \right)^2 \right] = \sigma^2 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right) .$$

Similarly, for

$$\tilde{c}_{2,couple} = r_{couple}^{1,1} h_{1,2}^* - r_{couple}^{1,2*} h_{1,1} - r_{couple}^{1,3} h_{1,4}^* + r_{couple}^{1,4*} h_{1,3} ,$$

$$\tilde{c}_{3,couple} = r_{couple}^{1,1} h_{1,3}^* + r_{couple}^{1,2} h_{1,4}^* - r_{couple}^{1,3*} h_{1,1} - r_{couple}^{1,4*} h_{1,2} ,$$

there are

$$P_{\text{interference}}^{\tilde{c}_{2,\text{couple}}} = E \left[ \left( \tilde{c}_{2,\text{couple}}^{\text{interference}} \right)^2 \right] \geq$$

$$P(u^2 + v^2) \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right) \left( |h_{1,3}|^2 + |h_{1,4}|^2 \right) +$$

$$P \left( u^2 |h_{1,1}|^2 + u^2 |h_{1,2}|^2 + v^2 |h_{1,3}|^2 + v^2 |h_{1,4}|^2 \right) \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right).$$

$$P_{\text{signal}}^{\tilde{c}_{2,\text{couple}}} = E \left[ \left( \tilde{c}_{2,\text{couple}}^{\text{signal}} \right)^2 \right] = P \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right)^2,$$

$$P_{\text{noise}}^{\tilde{c}_{2,\text{couple}}} = E \left[ \left( \tilde{c}_{2,\text{couple}}^{\text{noise}} \right)^2 \right] = \sigma^2 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right),$$

$$P_{\text{interference}}^{\tilde{c}_{3,\text{couple}}} = E \left[ \left( \tilde{c}_{3,\text{couple}}^{\text{interference}} \right)^2 \right] \geq$$

$$P \left[ u^2 \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right)^2 + v^2 \left( |h_{1,3}|^2 + |h_{1,4}|^2 \right)^2 \right] +$$

$$P \left( u^2 |h_{1,1}|^2 + u^2 |h_{1,2}|^2 + v^2 |h_{1,3}|^2 + v^2 |h_{1,4}|^2 \right) \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right),$$

$$P_{\text{signal}}^{\tilde{c}_{3,\text{couple}}} = E \left[ \left( \tilde{c}_{3,\text{couple}}^{\text{signal}} \right)^2 \right] = P \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right)^2,$$

$$P_{\text{noise}}^{\tilde{c}_{3,\text{couple}}} = E \left[ \left( \tilde{c}_{3,\text{couple}}^{\text{noise}} \right)^2 \right] = \sigma^2 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right).$$

Then,

$$\text{SINR}^{\text{Rx}} = \frac{P_{\text{signal}}^{\tilde{c}_{1,\text{couple}}} + P_{\text{signal}}^{\tilde{c}_{2,\text{couple}}} + P_{\text{signal}}^{\tilde{c}_{3,\text{couple}}}}{P_{\text{noise}}^{\tilde{c}_{1,\text{couple}}} + P_{\text{noise}}^{\tilde{c}_{2,\text{couple}}} + P_{\text{noise}}^{\tilde{c}_{3,\text{couple}}} + P_{\text{interference}}^{\tilde{c}_{1,\text{couple}}} + P_{\text{interference}}^{\tilde{c}_{2,\text{couple}}} + P_{\text{interference}}^{\tilde{c}_{3,\text{couple}}}} \leq$$

$$\frac{1}{3} + \frac{4 \left( u^2 |h_{1,1}|^2 + u^2 |h_{1,2}|^2 + v^2 |h_{1,3}|^2 + v^2 |h_{1,4}|^2 \right)}{4 \text{SNR}_{\text{Rx1}} \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right)} +$$

$$\frac{(u^2 + v^2) \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right) \left( |h_{1,3}|^2 + |h_{1,4}|^2 \right)}{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right)^2}.$$

## Appendix B

### Proof of Proposition 2

When  $K=2$ , the received signals on the second receiving antenna can be obtained from Eq. (27):

$$r_{\text{couple}}^{2,1} = \sqrt{P} h_2 (c_1 + uc_3) + \sqrt{P} h_{2,2} (c_2 + uc_3) +$$

$$\sqrt{P} h_{2,3} (c_3 + vc_1 + vc_2) + \sqrt{P} h_{2,4} (vc_1 + vc_2) + n_1^2,$$

$$r_{\text{couple}}^{2,2} = \sqrt{P} h_{2,1} (-c_2^* + uc_3) + \sqrt{P} h_{2,2} (c_1^* + uc_3) +$$

$$\sqrt{P} h_{2,3} (-vc_2^* + vc_1^*) + \sqrt{P} h_{2,4} (c_3 - vc_2^* + vc_1^*) + n_2^2,$$

$$r_{\text{couple}}^{2,3} = \sqrt{P} h_{2,1} (-c_3^* + uc_1^* - uc_2) + \sqrt{P} h_{2,2} (uc_1^* -$$

$$uc_2) + \sqrt{P} h_{2,3} (c_1^* - vc_3^*) + \sqrt{P} h_{2,4} (-c_2 - vc_3^*) + n_3^2,$$

$$r_{\text{couple}}^{2,4} = \sqrt{P} h_{2,1} (uc_2^* + uc_1) + \sqrt{P} h_{2,2} (-c_3^* + uc_2^* +$$

$$uc_1) + \sqrt{P} h_{2,3} (c_2^* - vc_3^*) + \sqrt{P} h_{2,4} (c_1 - vc_3^*) + n_4^2.$$

According to Ref. [22], for two receiving antennas, there are the following combination formulas:

$$\tilde{c}_{1,\text{couple}} = r_1^1 h_{1,1}^* + r_2^1 h_{1,2}^* + r_3^1 h_{1,3}^* + r_4^1 h_{1,4}^* + r_1^2 h_{2,1}^* +$$

$$r_2^2 h_{2,2}^* + r_3^2 h_{2,3}^* + r_4^2 h_{2,4}^*,$$

$$\tilde{c}_{2,\text{couple}} = r_1^1 h_{1,2}^* - r_2^1 h_{1,1}^* - r_3^1 h_{1,4}^* + r_4^1 h_{1,3}^* + r_1^2 h_{2,2}^* -$$

$$r_2^2 h_{2,1}^* - r_3^2 h_{2,4}^* + r_4^2 h_{2,3}^*,$$

$$\tilde{c}_{3,\text{couple}} = r_1^1 h_{1,3}^* + r_2^1 h_{1,4}^* - r_3^1 h_{1,1}^* - r_4^1 h_{1,2}^* + r_1^2 h_{2,3}^* +$$

$$r_2^2 h_{2,4}^* - r_3^2 h_{2,1}^* - r_4^2 h_{2,2}^*.$$

Since the form and structure of the received signals of the two antennas are consistent, the signal processing of the two receiving antennas is consistent in the combination formula. Therefore, it is easy to obtain

$$P_{\text{interference}}^{\tilde{c}_{1,\text{couple}}} = P_{\text{interference}}^{\tilde{c}_{2,\text{couple}}} \geq$$

$$P(u^2 + v^2) \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right) \left( |h_{1,3}|^2 + |h_{1,4}|^2 \right) +$$

$$P \left( u^2 |h_{1,1}|^2 + u^2 |h_{1,2}|^2 + v^2 |h_{1,3}|^2 + v^2 |h_{1,4}|^2 \right) \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right) + P(u^2 + v^2) \left( |h_{2,1}|^2 + |h_{2,2}|^2 \right) \left( |h_{2,3}|^2 + |h_{2,4}|^2 \right) +$$

$$P \left( u^2 |h_{2,1}|^2 + u^2 |h_{2,2}|^2 + v^2 |h_{2,3}|^2 + v^2 |h_{2,4}|^2 \right) \left( |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right),$$

$$P_{interference}^{\tilde{c}_{3,couple}} \geq P \left[ u^2 \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right)^2 + v^2 \left( |h_{1,3}|^2 + |h_{1,4}|^2 \right)^2 \right] +$$

$$P \left( u^2 |h_{1,1}|^2 + u^2 |h_{1,2}|^2 + v^2 |h_{1,3}|^2 + v^2 |h_{1,4}|^2 \right) \left( |h_{1,1}|^2 +$$

$$|h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right) + P \left[ u^2 \left( |h_{2,1}|^2 + |h_{2,2}|^2 \right)^2 +$$

$$v^2 \left( |h_{2,3}|^2 + |h_{2,4}|^2 \right)^2 \right] + P \left( u^2 |h_{2,1}|^2 + u^2 |h_{2,2}|^2 +$$

$$v^2 |h_{2,3}|^2 + v^2 |h_{2,4}|^2 \right) \left( |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right),$$

$$P_{signal}^{\tilde{c}_{1,couple}} = P_{signal}^{\tilde{c}_{2,couple}} = P_{signal}^{\tilde{c}_{3,couple}} = P \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 +$$

$$|h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right),$$

$$P_{noise}^{\tilde{c}_{1,couple}} = P_{noise}^{\tilde{c}_{2,couple}} = P_{noise}^{\tilde{c}_{3,couple}} = \sigma^2 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 +$$

$$|h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right).$$

Then we have

$$\text{SINR}^{2\text{Rx}} = \frac{P_{signal}^{\tilde{c}_{1,couple}} + P_{signal}^{\tilde{c}_{2,couple}} + P_{signal}^{\tilde{c}_{3,couple}}}{\frac{P_{noise}^{\tilde{c}_{1,couple}} + P_{noise}^{\tilde{c}_{2,couple}} + P_{noise}^{\tilde{c}_{3,couple}} + P_{interference}^{\tilde{c}_{1,couple}} + P_{interference}^{\tilde{c}_{2,couple}} + P_{interference}^{\tilde{c}_{3,couple}}}{1}} \leq \beta =$$

$$\frac{3}{4\beta \text{SNR}_{\text{Rx1}}} + d_1 + d_2 + d_3, \quad \text{where}$$

$$\frac{|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2}{|h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2} \text{ is}$$

the diversity gain and

$$d_1 = \frac{4 \left[ \left( u^2 |h_{1,1}|^2 + u^2 |h_{1,2}|^2 + v^2 |h_{1,3}|^2 + v^2 |h_{1,4}|^2 \right) \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 \right) \right]}{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right)^2},$$

$$d_2 = \frac{4 \left[ \left( u^2 |h_{2,1}|^2 + u^2 |h_{2,2}|^2 + v^2 |h_{2,3}|^2 + v^2 |h_{2,4}|^2 \right) \left( |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right) \right]}{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right)^2},$$

$$d_3 = \frac{(u^2 + v^2) \left[ \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right) \left( |h_{1,3}|^2 + |h_{1,4}|^2 \right) + \left( |h_{2,1}|^2 + |h_{2,2}|^2 \right) \left( |h_{2,3}|^2 + |h_{2,4}|^2 \right) \right]}{3 \left( |h_{1,1}|^2 + |h_{1,2}|^2 + |h_{1,3}|^2 + |h_{1,4}|^2 + |h_{2,1}|^2 + |h_{2,2}|^2 + |h_{2,3}|^2 + |h_{2,4}|^2 \right)^2}.$$

## References

- [1] LARSSON E G, EDFORS O, TUFVESSON F, et al. Massive MIMO for next generation wireless systems [J]. IEEE communications magazine, 2014, 52(2): 186 – 195. DOI: 10.1109/MCOM.2014.6736761
- [2] HAN C, JORNET J M, AKYILDIZ I. Ultra-massive MIMO channel modeling for graphene-enabled terahertz-band communications [C]//2018 IEEE 87th Vehicular Technology Conference (VTC Spring). IEEE, 2018: 1 – 5
- [3] HUANG C W, ZAPPONE A, ALEXANDROPOULOS G C, et al. Reconfigurable intelligent surfaces for energy efficiency in wireless communication [J]. IEEE transactions on wireless communications, 2019, 18(8): 4157 – 4170. DOI: 10.1109/TWC.2019.2922609
- [4] WU Q, ZHANG R. Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming [J]. IEEE transactions on wireless communications, 2019, 18(11): 5394 – 5409
- [5] TANG W K, CHEN M Z, CHEN X Y, et al. Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement [J]. IEEE transactions on wireless communications, 2021, 20(1): 421 – 439. DOI: 10.1109/TWC.2020.3024887
- [6] TAO Q, WANG J W, ZHONG C J. Performance analysis of intelligent reflecting surface aided communication systems [J]. IEEE communications letters, 2020, 24(11): 2464 – 2468. DOI: 10.1109/LCOMM.2020.3011843
- [7] CUI T J, QI M Q, WAN X, et al. Coding metamaterials, digital metamaterials and programmable metamaterials [J]. Light: science & applications, 2014, 3 (10): 218. DOI: 10.1038/lsa.2014.99
- [8] CUI T J, LIU S, ZHANG L. Information metamaterials and metasurfaces [J]. Journal of materials chemistry, 2017, 5(15): 3644 – 3668. DOI: 10.1039/c7tc00548b
- [9] ZHANG L, CHEN X Q, LIU S, et al. Space-time-coding digital metasurfaces [J]. Nature communications, 2018, 9(1): 1 – 11. DOI: 10.1038/s41467-018-06802-0
- [10] ZHAO J, CHENG Q, CHEN J, et al. A tunable metamaterial absorber using varactor diodes [J]. New journal of physics, 2013, 15(4): 043049. DOI: 10.1088/1367-2630/15/4/043049
- [11] HAN Y, TANG W K, JIN S, et al. Large intelligent surface-assisted wireless communication exploiting statistical CSI [J]. IEEE transactions on vehicular technology, 2019, 68(8): 8238 – 8242. DOI: 10.1109/TVT.2019.2923997
- [12] DI RENZO M, ZAPPONE A, DEBBAH M, et al. Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead [J]. IEEE journal on selected areas in communications, 2020, 38(11): 2450 – 2525. DOI: 10.1109/JSAC.2020.3007211
- [13] ZHAO J, YANG X, DAI J Y, et al. Programmable time-domain digital-coding metasurface for non-linear harmonic manipulation and new wireless communication systems [J]. National science review, 2018, 6(2): 231 – 238. DOI: 10.1093/nsr/nwy135
- [14] TANG W K, LI X, DAI J Y, et al. Wireless communications with programmable metasurface: transceiver design and experimental results [J]. China communications, 2019, 16(5): 46 – 61. DOI: 10.23919/j.cc.2019.05.004
- [15] DAI J Y, TANG W K, ZHAO J, et al. Wireless communications through a simplified architecture based on time-domain digital coding metasurface [J]. Advanced materials technologies, 2019, 4(7): 1900044. DOI: 10.1002/admt.201900044
- [16] TANG W K, DAI J Y, CHEN M Z, et al. Programmable metasurface-based RF chain-free 8PSK wireless transmitter [J]. Electronics letters, 2019, 55(7): 417 – 420. DOI: 10.1049/el.2019.0400
- [17] TANG W K, DAI J Y, MING Z C, et al. Realization of reconfigurable intelligent surface-based alamouti space-time transmission [C]//Proceedings of 2020 International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2020: 904 – 909. DOI: 10.1109/WCSP49889.2020.9299726
- [18] KE J C, DAI J Y, CHEN M Z, et al. Linear and nonlinear polarization syntheses and their programmable controls based on anisotropic time-domain digital coding metasurface [J]. Small structures, 2021, 2(1): 2000060. DOI: 10.1002/ssr.202000060
- [19] CHEN X Y, KE J C, TANG W K, et al. Design and implementation of MIMO transmission based on dual-polarized reconfigurable intelligent surface [J]. IEEE wireless communications letters, 2021, 10(10): 2155 – 2159. DOI: 10.1109/LWC.2021.3095172
- [20] HAN Y, LI X, TANG W K, et al. Dual-polarized RIS-assisted mobile commu-

nications [J]. *IEEE transactions on wireless communications*, 2022, 21(1): 591 – 606. DOI: 10.1109/TWC.2021.3098521

- [21] PATIL S T, SHINDE P N. A review of space-time block codes from real and complex orthogonal designs [J]. *International journal of computer applications*, 2014, 100(11): 16 – 19. DOI: 10.5120/17568-8239
- [22] LI K, WANG Z. Analysis of space-time block coding (STBC) performance in MIMO system [J]. *Wireless communication technology*, 2008, 17(4): 11 – 16
- [23] CHO K, YOON D. On the general BER expression of one- and two-dimensional amplitude modulations [J]. *IEEE transactions on communications*, 2002, 50(7): 1074 – 1080. DOI: 10.1109/TCOMM.2002.800818

### Biographies

**ZHOU Mingyong** received his B.S. degree in communication engineering from the School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, China in 2020. He is currently pursuing his M.S. degree with the National Mobile Communications Research Laboratory, Southeast University, China. His research interests include metasurface-based wireless communication system design and implementation.

**CHEN Xiangyu** received his B.S. degree in information engineering from the School of Information Science and Engineering, Southeast University, China in 2019. He is currently pursuing his M.S. degree with the National Mobile Communications Research Laboratory, Southeast University. His research interests include metasurface-based wireless communication system design and measurement.

**TANG Wankai** received his B.S. degree in information engineering, M.S. degree in circuits and systems, and Ph.D. degree in communications and information systems from Southeast University, China in 2011, 2014, and 2021, respectively. He worked at National Instruments (Shanghai), USA from 2014 to 2016. He received the Electronics Letters Best Paper Award in 2020 and the China Communications Best Paper Award in 2021. His research interests include modeling and prototyping of metasurface-based wireless communication systems.

**KE Jun Chen** received his B.S. degree from the School of Instrument Science & Engineering, Southeast University, China in 2016, where he is currently pursuing his Ph.D. degree with the State Key Laboratory of Millimeter Waves, De-

partment of Radio Engineering, Southeast University, China. His research interests include metamaterials, metasurfaces, and array antennas.

**JIN Shi** (jinshi@seu.edu.cn) received his Ph.D. degree in communications and information systems from Southeast University, China in 2007. From June 2007 to October 2009, he was a research fellow with the Adastral Park Research Campus, University College London, United Kingdom. He is currently with the faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include space-time wireless communications, information theory, intelligent communications, and reconfigurable intelligent surfaces. He and his coauthors received the 2010 Young Author Best Paper Award from the IEEE Signal Processing Society and the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory.

**CHENG Qiang** received his B.S. and M.S. degrees from Nanjing University of Aeronautics and Astronautics, China in 2001 and 2004, respectively, and his Ph.D. degree from Southeast University, China in 2008. He is currently a full professor with the State Key Laboratory of Millimeter Waves, Southeast University. His research interests include metamaterials, tunable microwaves circuits, microwave imaging, and terahertz systems. He was a recipient of the 2010 Best Paper Award from *New Journal of Physics*. He was awarded China's Top Ten Scientific Advances of 2010 and the second class National Natural Science Award in 2014.

**CUI Tie Jun** received his B.S., M.S., and Ph.D. degrees in electrical engineering from Xidian University, China in 1987, 1990, and 1993, respectively. In 1993, he joined the Department of Electromagnetic Engineering, Xidian University, and was promoted to associate professor in 1993. From 1995 to 1997, he was a research fellow with the Institut für Hochfrequenztechnik und Elektronik, University of Karlsruhe, Germany. In 1997, he joined the Center for Computational Electromagnetics, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, USA, first as a postdoctoral research associate and then as a research scientist. In 2001, he was a Cheung-Kong Professor with the Department of Radio Engineering, Southeast University, China. Currently he is the Chief Professor of Southeast University. He is also the founding director of the Institute of Electromagnetic Space, Southeast University. His current research interests include metamaterials and computational electromagnetics. He has published over 500 peer-review journal papers, which have been cited more than 40 000 times (H-Factor 102; Google Scholar). He is the Academician of the Chinese Academy of Science.

# Metric Learning for Semantic-Based Clothes Retrieval



YANG Bo<sup>1</sup>, GUO Caili<sup>1,2</sup>, LI Zheng<sup>1</sup>

(1. Beijing Laboratory of Advanced Information Networks, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China)

DOI: 10.12142/ZTECOM.202201010

<https://kns.cnki.net/kcms/detail/34.1294.TN.20220307.1444.002.html>, published online March 9, 2022

Manuscript received: 2021-11-15

**Abstract:** Existing clothes retrieval methods mostly adopt binary supervision in metric learning. For each iteration, only the clothes belonging to the same instance are positive samples, and all other clothes are “indistinguishable” negative samples, which causes the following problem. The relevance between the query and candidates is only treated as relevant or irrelevant, which makes the model difficult to learn the continuous semantic similarities between clothes. Clothes that do not belong to the same instance are completely considered irrelevant and are uniformly pushed away from the query by an equal margin in the embedding space, which is not consistent with the ideal retrieval results. Motivated by this, we propose a novel method called semantic-based clothes retrieval (SCR). In SCR, we measure the semantic similarities between clothes and design a new adaptive loss based on these similarities. The margin in the proposed adaptive loss can vary with different semantic similarities between the anchor and negative samples. In this way, more coherent embedding space can be learned, where candidates with higher semantic similarities are mapped closer to the query than those with lower ones. We use Recall@K and normalized Discounted Cumulative Gain (nDCG) as evaluation metrics to conduct experiments on the DeepFashion dataset and have achieved better performance.

**Keywords:** clothes retrieval; metric learning; semantic-based retrieval

**Citation** (IEEE Format): B. Yang, C. L. Guo, and Z. Li, “Metric learning for semantic-based clothes retrieval,” *ZTE Communications*, vol. 20, no. 1, pp. 76 – 82, Mar. 2022. doi: 10.12142/ZTECOM.202201010.

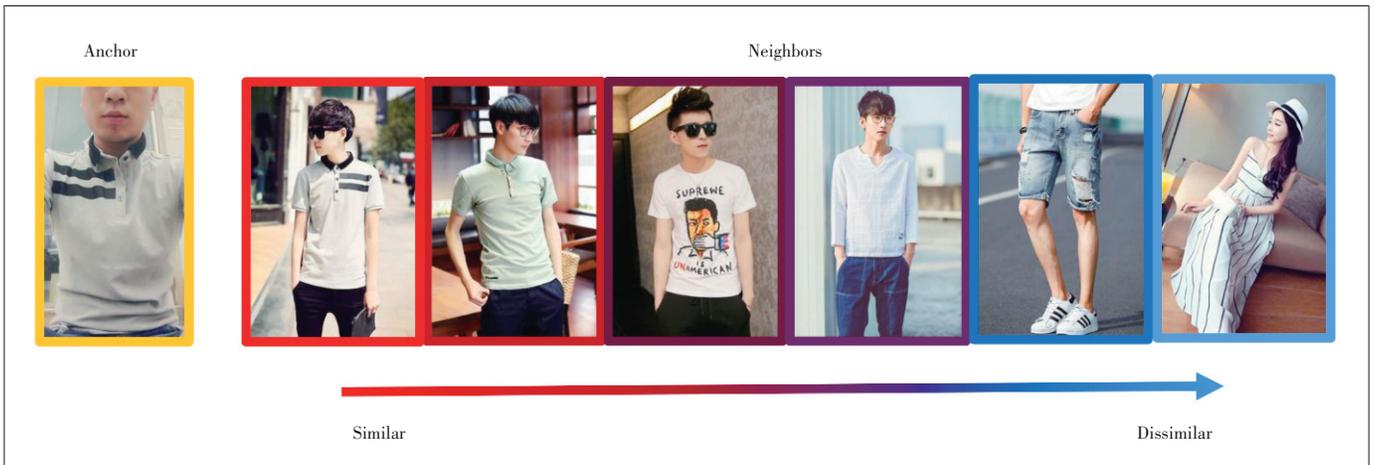
## 1 Introduction

Clothes retrieval, commonly associated with visual search, has received a lot of attention recently, which is also an interesting challenge from both commercial and academic perspectives. The task of clothing retrieval is to find the exact or very similar products from the gallery to the given query image. It mainly has two scenarios, namely, in-shop clothes retrieval and consumer-to-shop clothes retrieval. In the former case, query and gallery come from the same domain while it is opposite in the latter case. In consumer-to-shop clothes retrieval, photos in the gallery usually only contain a single product and are taken by professionals using high-quality equipment. On the contrary, the photos taken by the user are usually of low quality, possibly with cluttered backgrounds and multiple unrelated objects<sup>[1]</sup>.

The above uncertain factors have brought great challenges to consumer-to-shop clothes retrieval. In this complex scenario, only the model which can discover the differences between clothes from a semantic point of view will not be affected by environmental factors, so as to obtain better retrieval results. However, most of the current clothing retrieval works belong to instance-based retrieval, and they are all based on

the following assumptions: Only the clothes belonging to the same instance are considered relevant, while all the other clothes are irrelevant. However, this assumption appears conflicted with our common sense. For example, as shown in Fig. 1, the semantic similarity between clothes is more than relevant or irrelevant. Instead, it shows a continuous decreasing trend. From a category point of view, when we use a polo shirt as a query, except for the same polo shirt, the other polo shirts should theoretically be the most similar. Ordinary short-sleeved and long-sleeved may follow next, and finally it comes to other different categories such as pants and skirts. However, under the premise of the above assumption, models are more likely to learn the ranking relationship among clothes, rather than their semantic similarities<sup>[2]</sup>. Moreover, the learned embedding space appears not continuous either, which may lead to worse user experience.

In order to address the above mentioned problems, Refs. [2 – 4] propose several metric learning methods using data with continuous labels, which is called semantic-based retrieval. Our work is inspired by and completed based on the research in these fields. In this paper, we propose a new method called Semantic-Based Clothes Retrieval (SCR) to measure the semantic similarity between clothes, and these



▲ Figure 1. Ideal relationship between clothes in different categories

similarities can guide the learning process of continuous embedding space. In this way, we realize the semantic-based clothes retrieval eventually.

Our contributions can be summarized as follows.

- We expose the problem in the instance-based clothes retrieval. The relationship between clothes is simply regarded as relevant or irrelevant, which is not consistent with the actual situation.

- We propose a novel method called SCR to measure the semantic similarity between clothes and design a new adaptive loss based on these similarities. As a result, clothes in a gallery are ranked by their similarity to the query, making multiple clothes relevant.

- We use Recall@k and normalized discounted cumulative gain (nDCG) as the evaluation metrics and conduct experiments on the DeepFashion dataset, and have achieved better results.

## 2 Related Work

### 2.1 Instance-Based Clothes Retrieval

Pioneer works<sup>[5-8]</sup> in clothing retrieval utilized traditional features like the scale invariant feature transform (SIFT). Later, due to the wide application of deep neural networks, the development of computer vision was greatly promoted. In Refs. [9 - 16], these methods usually extracted both global features and local features, combining them for similarity calculation and matching. Recently, the authors in Ref. [17] have proposed a graph reasoning network (GRNet), which first represented the multi-scale regional similarities and their relationships as a graph and then performed graph convolutional neural network (CNN) based reasoning over the graph to adaptively adjust both the local and global similarities.

All the aforementioned methods focus on the feature extraction stage of instance-based clothes retrieval. Closest to our work, Ref. [1] transferred the leading ReID model called the ReID strong baseline (RST)<sup>[18]</sup> to fashion retrieval task, signifi-

cantly outperforming previous state-of-the-art results despite a much simpler architecture. In this paper, we use this ReID model as our backbone for subsequent semantic-based clothes retrieval.

### 2.2 Metric Learning for Instance-Based Clothes Retrieval

Metric learning attempts to map data to an embedding space, where similar data are close together while dissimilar data are far apart<sup>[1]</sup>. In general, metric learning can be achieved by means of embedding and classification losses, and both of them are often utilized at the same time in most retrieval methods. Many state-of-the-art methods combine ID and triplet losses to constrain the same feature  $f$ . Combining these two losses always makes the model achieve better performance<sup>[18]</sup>.

The contrastive loss<sup>[19]</sup> and the triplet loss<sup>[20]</sup>, as two typical embedding losses, provide the foundation of metric learning. Given an image pair, the contrastive loss minimizes their distance in the embedding space if their classes are the same, and separates them with a fixed margin away otherwise. The triplet loss takes triplets of anchor, positive, and negative images, and enforces the distance between the anchor and the positive to be smaller than that between the anchor and the negative image. A wide variety of losses has since been built on these fundamental concepts such as quadruplet loss,  $n$ -pair loss, and lifted structured loss.

Although the above mentioned losses substantially improve the quality of the learned embedding space, they are commonly based on binary relations between image pairs, thus they are not directly applicable for metric learning in semantic-based retrieval.

### 2.3 Metric Learning for Semantic-Based Retrieval Using Continuous Labels

There have been several metric learning methods for semantic-based retrieval using data with continuous labels in some research areas such as image caption<sup>[21]</sup>, place recogni-

tion<sup>[22]</sup> and camera relocalization<sup>[23]</sup>.

Similar to our work, KIM et al.<sup>[2]</sup> propose a log-ratio loss to learn embedding space in image retrieval. Their work primarily focuses on human poses, in which they use the distance between joints to rank images. They also explore within-modal image retrieval using word mover's distance, as a proxy for semantic similarity. For cross-modal retrieval, ZHOU et al.<sup>[3]</sup> propose to measure the relevance degree among images and sentences and design a ladder loss to learn coherent embedding space. In the ladder loss, these relevance degree values are divided into several levels, but the relevance in each level is still formulated as a binary variable. WRAY et al.<sup>[4]</sup> propose the task of semantic similarity video retrieval (SVR), which allows multiple captions to be relevant to a video and vice-versa, and defines non-binary similarity among items. In addition, they also propose several proxies to estimate semantic similarities and introduce nDCG as the evaluation metric.

The above methods all implement semantic-based retrieval by metric learning with continuous labels, and the premise is that the dataset provides continuous labels, or the similarity can be measured by well-designed proxies. While, in the field of clothing retrieval, it is difficult to measure the semantic similarities among clothes because of the semantic gap between visual similarity and semantic similarity. Therefore, semantic-based clothing retrieval has not been realized so far.

### 3 Semantic-Based Clothes Retrieval

In this section, we propose to move beyond instance-based clothes retrieval towards that uses semantic similarity among clothes. In our proposed SCR, we first measure the semantic similarity for clothes in Section 3.1. And then we modify classic triplet loss to adaptive loss for clothes retrieval using semantic similarity in Section 3.2. In this way, coherent embedding space can be learned, where candidates with higher semantic similarities are mapped closer to the query than those with lower semantic similarities.

#### 3.1 Semantic Similarity for Clothes

The semantic gap exists between the raw image and the full semantic understanding of the image's content. In our method, we take full advantage of annotation information in DeepFashion to bridge the gap.

Various tags of clothes are labeled in the DeepFashion dataset, in addition to categories, bounding boxes, key points, and the attributes of clothes. Taking the consumer-to-shop benchmark as an example, each piece of clothing has 303 attribute tags. These attribute tags are attached to 18 categories, including clothing length, thickness, material, style and other categories. According to these attributes, clothing can be described from multiple points of view. Previous works used these attribute annotations as supervision for the multi-label classification task, and combined clothing classification and retrieval loss to jointly train a neural network. However, due to the com-

plexity and diversity of attribute tags, the multi-label classification task is considered an extraordinary challenge. In order to fully utilize the semantic information hidden in the annotated labels, we do not use attribute labels as the supervision for attribute classification, but directly use them to guide the learning process of the embedding space.

Specifically, given a set of clothes instances  $C$ , we let  $c_i$  be a clothes instance. We use a 303-dimensional vector of 0 or 1 to represent the attribute vector of  $c_i$ , denoted by  $s_i$ . The similarity between  $s_i$  and other clothes  $s_j$  can be measured by the inner product of two vectors,

$$S(c_i, c_j) = \langle s_i, s_j \rangle. \quad (1)$$

That is to say, if the corresponding position of  $s_i$  and  $s_j$  is 1, which means two clothes both have a certain attribute, and after accumulation of all positions, the final value is used to represent the semantic similarity between them. The inner product of the vector and itself is always the largest, which is consistent with "clothing is always the most similar to itself". Besides, the attribute labels are characterized by fine granularity as well as high accuracy, so they are able to quantitatively measure the semantic similarities between clothes. We use the calculated similarities to guide the follow-up learning process of embedding space as well as evaluate the performance of the retrieval model.

#### 3.2 Adaptive Loss

The classic triplet loss takes a triplet of an anchor, a positive, and a negative image as input. It is designed to penalize triplets violating the rank constraint. The distance between the anchor and the positive images must be smaller than that between the anchor and the negative images in the embedding space. The loss is formulated as:

$$L_{\text{Triplet}} = \left[ \lambda + D(c_a, c_p) - D(c_a, c_n) \right]_+, \quad (2)$$

where  $(c_a, c_p)$  is the positive pair,  $(c_a, c_n)$  is the negative pair for a query,  $D(\cdot)$  means the squared Euclidean distance of the embedding vector,  $\lambda$  is a fixed margin, and  $[\cdot]_+$  denotes the hinge function. Note that the embedding vectors should be L2 normalized since their magnitudes tend to diverge and the margin becomes trivial without such a normalization. For the RST model used in clothes retrieval<sup>[1]</sup>, the mining strategy selects the most difficult positive and negative samples, that is the farthest positive sample and the closest negative sample in embedding space.

The triplet loss tends to treat the relevance between query and candidates in a bipolar way: for a query  $c_a$ , only the exactly same clothes  $c_p$  are regarded as relevant, and other clothes  $c_n$  are all regarded as irrelevant. Therefore, only  $c_p$  is pulled closed to  $c_a$ , while others are pushed away by a fixed

margin equally. However, as mentioned in Section 1, the semantic similarity between samples should not be a binary variable.

In ideal embedding space, the difference in distance between positive/negative samples and anchor should be proportional to the difference in semantic similarities between them. In other words, a negative pair with lower semantic similarity should be pulled farther apart. Therefore, it is beneficial to introduce the semantic similarities to determine how far negative samples will be pushed away. We design a novel adaptive loss for clothes retrieval based on a classic triplet loss as follows:

$$L_{\text{Adaptive}} = \left[ M_{an} + D(c_a, c_p) - D(c_a, c_n) \right]_+, \quad (3)$$

$$M_{an} = \left( 1 - \frac{S(c_a, c_n)}{S_{\max}} \right) * \lambda, \quad (4)$$

where  $S_{\max}$  is a normalization factor to guarantee that the maximum value of semantic similarity is 1. 1 minus the normalized value, as the coefficient of  $\lambda$ , can allow those negative samples that are very similar to the anchor to be pushed away by a small margin. On the contrary, those negative samples that are not similar to the anchor will be pushed away by a large margin.

In the proposed adaptive loss, the margin between the anchor and the corresponding negative sample is no longer a fixed value, but a dynamic distance varying with semantic similarity computed in Eq. (1). In this way, a model is trained under metric learning beyond binary supervision and a coherent embedding space is learned as a result.

## 4 Experiments

### 4.1 Dataset and Experiment Settings

We evaluate our loss on the consumer-to-shop clothes retrieval benchmark of DeepFashion<sup>[11]</sup>. This benchmark aims at matching consumer-taken photos with their shop counterparts. It contains 33 881 clothing items, 239 557 consumer/shop clothes images, 195 540 cross-domain pairs, and each image is annotated by the bounding box, clothing type and source type. In our experiments, we use 96 708 images for testing and the remaining for training.

In this section, we select the RST model<sup>[1]</sup> that has obtained the best performance in clothes retrieval as the backbone network for the following experiments. Our loss function consists of two parts, which are obtained by adding the classification loss and the ranking loss (adaptive loss). We keep the identity loss in the original RST model and use it as the classification loss. As for the adaptive loss, we first select the closest negative sample to the anchor and the farthest positive sample in a

batch, and then use the dynamic margin calculated by the semantic similarity between the negative sample and anchor to replace the fixed  $\lambda$  in the triplet. The adaptive loss we proposed can be used as an improved version of triplet loss in any backbone. In this paper, RST with the original classic triplet loss is considered as a comparison method.

Except for the loss function, several variants of the RST model were proposed in Ref. [1], and we select one with ResNet50-IBN-A backbone, 320×320 input image size, and no re-ranking setting as our baseline. We follow the warmup learning rate strategy proposed in the original RST model<sup>[18]</sup>. Models are trained using Adam for 120 epochs, with a batch size of 64. Hyper-parameters are set as  $\lambda = 0.3$ , and  $\epsilon = 0.1$ .

For instance-based retrieval, we use the Recall@K (R@K) as the evaluation metric for the task. R@K indicates the percentage of queries for which the model returns the correct item in its top  $K$  results.

For semantic-based retrieval, we use nDCG<sup>[24]</sup> as evaluation metrics. The nDCG has been used previously for information retrieval<sup>[25]</sup>. It requires similarity scores among all items in the test set. We calculate discounted cumulative gain (DCG) for a query  $q_i$  and the set of items  $Z$ , ranked according to their distance from  $q_i$  in the learned embedding space:

$$\text{DCG@K}(q_i) = \sum_{j=1}^K \frac{2^{S(q_i, z_j)} - 1}{\log(j + 1)}, \quad (5)$$

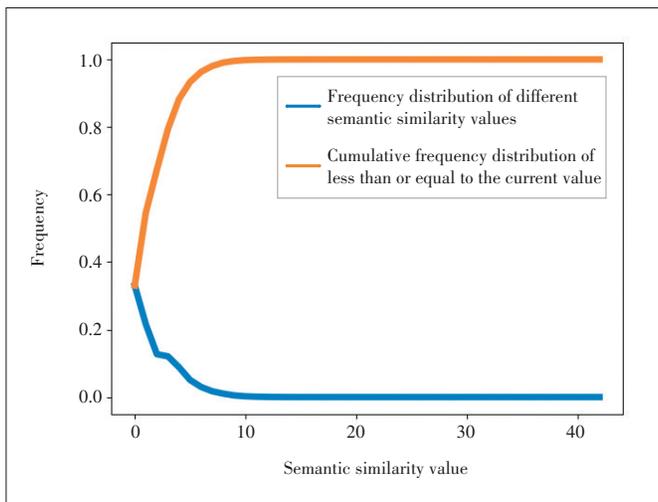
where  $K$  is the length of the list returned by the retrieval system. Note that this equation would give the same value when items of the same similarity  $S_S$  are retrieved in any order. It also captures different levels of semantic similarity. The nDCG can then be calculated by normalizing the DCG score so that it lies in the range [0, 1]:

$$\text{nDCG@K}(q_i) = \frac{\text{DCG@K}(q_i)}{\text{IDCG@K}(q_i)}, \quad (6)$$

where  $\text{IDCG@K}(q_i)$  is calculated from  $\text{DCG}$  and  $Z$  ordered by relevance to the query  $q_i$ .

### 4.2 Analysis of Semantic Similarity

In this paper, we follow the method introduced in Section 3.1 to obtain the inner product of two attribute vectors as the semantic similarity among clothes. After statistics, the minimum value of similarity is 0 and the maximum is 42, as shown in Fig. 2, where the blue curve represents the frequency distribution of different semantic similarity values and the sum of these values is 1; The orange line shows the cumulative frequency distribution of less than or equal to the current value. In this range, the greater the similarity, the lower the frequency of occurrence. Among them, the proportion of similarity of 0 or 1 is 54%, and the proportion of similarity of less



▲ Figure 2. Frequency statistical graph of semantic similarity

than or equal to 8 is as high as 99%. In other words, the similarity among most clothes is generally very low, and the similarity of only a few clothes can be close to or equal to their own similarity. Our goal is to use these high similarity samples to learn their common features from a semantic point of view, so as to optimize the learning quality of the embedding space.

Fig. 3 shows the normalized semantic similarities calculated by the inner product of the 303-dimensional attribute vectors. Note that similarities here are normalized by  $S_{ap}$  instead of  $S_{max}$  for better understanding. If clothes possess all the attribute tags of anchor, their semantic similarity with anchor will come to 1. While, in the training process, we use the semantic similarity normalized by  $S_{max}$  to determine the distance that negative samples are pushed away from the query, that is,



▲ Figure 3. Normalized semantic similarity visualization

the lower the semantic similarity, the farther will be pushed away from the query.

### 4.3 Comparison with State-of-the-Art Methods

Table 1 compares the proposed SCR with state-of-the-art methods, including FashionNet<sup>[11]</sup>, Visual Attention Model (VAM) and its variants (VAM+ImgDrop, VAM+Product, and VAM+Nonshared)<sup>[26]</sup>, DREML<sup>[27]</sup>, KPM<sup>[28]</sup> and GRNet<sup>[17]</sup>. The proposed SCR obtains competitive results. Specifically, it obtains an accuracy of 29.2, 51.0 and 61.4 and outperforms the existing methods at R@1.

### 4.4 Comparison with Triplet Loss and Quadruplet Loss

Since the adaptive loss we proposed is an improved version of the triplet loss, in order to verify whether the proposed adaptive loss is effective, we compare it with the triplet loss in both instance-based and semantic-based retrieval. Moreover, a comparison with quadruplet loss is also revealed below.

#### 1) Instance-based retrieval results

As shown in Table 2, the adaptive loss has a significant improvement compared with a triplet loss function. This means that the information provided by the calculated semantic similarity can improve the quality of the learned embedding space and performance better in the retrieval process. However, Recall@ $k$  only cares about the appearance of the first  $k$  results, and does not care about their sorting positions, so the measurement results for our optimization based on semantic similarity are limited.

#### 2) Semantic-based retrieval results

Table 3 summarizes the results evaluated by nDCG. The proposed loss function has an obvious improvement. In terms of nDCG@5, nDCG@10, and nDCG@50, the adaptive loss separately increases by 4.3%, 7.6%, 6.7%, compared with the classical triplet loss. Since the nDCG metric can simultaneously consider the relevance degree and ranking position, it reflects semantic-based retrieval results more accurately.

Fig. 4 shows the qualitative comparison between the triplet loss and adaptive loss on the consumer-to-shop benchmark using RST. For each retrieval, given an image query, we show the top-5 retrieved products. The correct retrieval items for each query are outlined in green boxes. Compared with the triplet loss, the correct query result ranks higher. Moreover, the search results based on the

adaptive loss possess higher semantic similarity with the query. For example, in the first query results based on the adaptive loss, although the first two candidates are not considered ground truth, they are indeed indistinguishable from the query, which also proves that it is not in line with the actual retrieval requirements to only consider ground truth to be relevant.

In addition, even if it is an incorrect query result, it will be

▼ **Table 1. Comparison with state-of-the-art methods on DeepFashion consumer-to-shop benchmark**

Methods	R@1	R@20	R@50
FashionNet <sup>[11]</sup>	7.0	18.8	22.8
VAM+Nonshared <sup>[26]</sup>	11.3	38.8	51.5
VAM+Product <sup>[26]</sup>	13.4	43.6	56.7
VAM+ImageDrop(192, 48) <sup>[26]</sup>	13.7	43.9	56.9
DREML <sup>[27]</sup>	18.6	51.0	59.1
KPM <sup>[28]</sup>	21.3	54.1	65.2
GRNeT <sup>[17]</sup>	25.7	64.4	75.0
SCR (Adaptive)	29.2	51.0	61.4

GRNet: graph reasoning network  
KPM: Kronecker-product matching

SCR: semantic-based clothes retrieval  
VAM: visual attention model

▼ **Table 2. Instance-based retrieval results on DeepFashion**

Loss Functions	R@1	R@5	R@10	mAP	Mean
Triplet	26.9	35.9	41.0	33.9	1 275
Quadruplet	26.3	35.1	40.3	33.3	1 348
Adaptive	29.2	38.6	44.0	36.6	1 091

▼ **Table 3. Semantic-based retrieval results on DeepFashion**

Loss Functions	NDCG@1	NDCG@10	NDCG@50
Triplet	22.2	21.3	16.4
Quadruplet	21.8	20.9	16.1
Adaptive	23.9	22.8	17.5

nDCG: normalized discounted cumulative gain



▲ **Figure 4. Qualitative retrieval comparison between triplet loss and our adaptive loss on consumer-to-shop benchmark**

semantically closer to the query. For example, in the third query result, when using a shirt with a distinctive pattern as a query, the network trained by the triplet loss pays too much attention to the local similarity because four of the top five search results are pants with similar patterns. From the perspective of the clothing category, the search results are completely contrary to the user's intention. The network trained by the adaptive loss focuses on the overall semantic similarity so the searching scope tends to be restricted to the same category, which is crucial to the user experience. This benefits from the fact that the margin we use in the adaptive loss can adjust dynamically according to different samples. In this way, the model is easier to explore the semantic relationship among samples and finally a more semantic coherent embedding space is learned.

## 5 Conclusions

This paper focuses on semantic-based clothes retrieval and proposes a novel method called SCR to measure the semantic similarity between clothes. Motivated by metric learning with continuous labels in other research areas, we modify the classic triplet loss using semantic similarity and design an adaptive loss for clothes retrieval. As a result, more reasonable embedding space is learned, where candidates with higher semantic similarities are mapped closer to the query than those with lower similarities, which is more in line with the actual user experience of the retrieval system. Our method outperforms the baseline and obtains competitive semantic-based retrieval results on consumer-to-shop retrieval benchmarks of DeepFashion.

## Acknowledgment

We would like to acknowledge LIU Xuean, HE Baolin, ZHANG Yuanjian from ZTE Corporation for their contribution to this paper.

## References

- [1] WIECZOREK M, MICHALOWSKI A, WROBLEWSKA A, et al. A strong baseline for fashion retrieval with person re-identification models [EB/OL]. (2020-03-09) [2021-11-05]. <https://www.arxiv-vanity.com/papers/2003.04094/>. DOI: 10.1007/978-3-030-63820-7\_33
- [2] KIM S, SEO M, LAPTEV I, et al. Deep metric learning beyond binary supervision [EB/OL]. (2019-04-21) [2021-11-05]. <https://arxiv.org/abs/1904.09626>. DOI: 10.1109/CVPR.2019.00239
- [3] ZHOU M, NIU Z X, WANG L, et al. Ladder loss for coherent visual-semantic embedding [J]. Proceedings of the AAAI conference on artificial intelligence, 2020, 34(7): 13050 - 13057. DOI: 10.1609/aaai.v34i07.7006

- [4] WRAY M, DOUGHTY H, DAMEN. On semantic similarity in video retrieval [EB/OL]. [2021-11-05]. <https://arxiv.org/abs/2103.10095>
- [5] WANG X W, ZHANG T. Clothes search in consumer photos via color matching and attribute learning [C]//MM '11: Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011: 1353 - 1356. DOI: 10.1145/2072298.2072013
- [6] DI W, WAH C, BHARDWAJ A, et al. Style finder: fine-grained clothing style detection and retrieval [C]//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2013: 8 - 13. DOI: 10.1109/CVPRW.2013.6
- [7] FU J L, WANG J Q, LI Z C, et al. Efficient clothing retrieval with semantic preserving visual phrases [C]//Asian Conference on Computer Vision. ACCV, 2012: 420 - 431
- [8] GARCIA N, VOGIATZIS G. Dress like a star: Retrieving fashion products from videos [C]//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. IEEE, 2017: 2293 - 2299. DOI: 10.1109/ICCVW.2017.270
- [9] HUANG J, FERIS R S, CHEN Q, et al. Cross-domain image retrieval with a dual attribute-aware ranking network [C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015: 1062-1070. DOI: 10.1109/ICCV.2015.127
- [10] KIAPOUR M H, HAN X F, LAZEBNIK S, et al. Where to buy it: matching street clothing photos in online shops [C]//Proceedings of 2015 IEEE International Conference on Computer Vision. IEEE, 2015: 3343 - 3351. DOI: 10.1109/ICCV.2015.382
- [11] LIU Z W, LUO P, QIU S, et al. DeepFashion: powering robust clothes recognition and retrieval with rich annotations [J]. 2016 IEEE conference on computer vision and pattern recognition (CVPR), 2016: 1096 - 1104. DOI: 10.1109/CVPR.2016.124
- [12] JI X, WANG W, ZHANG M H, et al. Cross-domain image retrieval with attention modeling [C]//Proceedings of the 25th ACM international conference on Multimedia. ACM, 2017: 1654 - 1662. DOI: 10.1145/3123266.3123429
- [13] SONG Y, LI Y, WU B, et al. Learning unified embedding for apparel recognition [EB/OL]. (2017-07-19) [2021-11-05]. <https://arxiv.org/abs/1707.05929>
- [14] CORBIÈRE C, BEN-YOUNES H, RAMÉ A, et al. Leveraging weakly annotated data for fashion image retrieval and label prediction [C]//Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. IEEE, 2017: 2268 - 2274. DOI: 10.1109/ICCVW.2017.266
- [15] CHENG Z Q, WU X, LIU Y, et al. Video2Shop: exact matching clothes in videos to online shopping images [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 4169 - 4177. DOI: 10.1109/CVPR.2017.444
- [16] ZHANG Y H, PAN P, ZHENG Y, et al. Visual search at Alibaba [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 993 - 1001. DOI: 10.1145/3219819.3219820
- [17] KUANG Z H, GAO Y M, LI G B, et al. Fashion retrieval via graph reasoning networks on a similarity pyramid [C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 3066 - 3075. DOI: 10.1109/ICCV.2019.00316
- [18] LUO H, JIANG W, GU Y Z, et al. A strong baseline and batch normalization neck for deep person re-identification [J]. IEEE transactions on multimedia, 2020, 22(10): 2597 - 2609. DOI: 10.1109/TMM.2019.2958756
- [19] HADSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping [C]//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2006: 1735 - 1742. DOI: 10.1109/CVPR.2006.100
- [20] WEINBERGER K Q, SAUL L K. Distance metric learning for large margin nearest neighbor classification [J]. Journal of machine learning research, 2009, 10(2): 207 - 244
- [21] GORDO A, LARLUS D. Beyond instance-level image retrieval: leveraging captions to learn a global visual representation for semantic retrieval [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 5272 - 5281. DOI: 10.1109/CVPR.2017.560
- [22] ARANDJELOVIC R, GRONAT P, TORII A, et al. NetVLAD: CNN architecture for weakly supervised place recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 5297 - 5307. DOI: 10.1109/CVPR.2016.572
- [23] BALNTAS V, LI S D, PRISACARIU V. RelocNet: continuous metric learning relocalisation using neural nets [C]//Computer Vision. ECCV, 2018: 751 - 767. DOI: 10.1007/978-3-030-01264-9\_46
- [24] JÄRVELIN K, KEKÄLÄINEN J. Cumulated gain-based evaluation of IR techniques [J]. ACM transactions on information systems, 2002, 20(4): 422 - 446. DOI: 10.1145/582415.582418
- [25] LIU T Y. Learning to rank for information retrieval [M]. Berlin, Germany: Springer, 2011
- [26] WANG Z H, GU Y J, ZHANG Y, et al. Clothing retrieval with visual attention model [C]//Proceedings of 2017 IEEE Visual Communications and Image Processing. IEEE, 2017: 1 - 4. DOI: 10.1109/VICIP.2017.8305144
- [27] XUAN H, SOUVENIR R, PLESS R. Deep randomized ensembles for metric learning [C]//Computer vision. ECCV, 2018: 723 - 734. DOI: 10.1007/978-3-030-01270-0\_44
- [28] SHEN Y, XIAO T, LI H, et al. End-to-end deep kronecker-product matching for person re-identification [C]//Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition. IEEE, 2018: 6886 - 6895. DOI: 10.48550/arXiv.1807.11182

### Biographies

**YANG Bo** received the B.S. degree in communication engineering from Beijing University of Posts and Telecommunications (BUPT), China in 2019. He is currently pursuing the M.S. degree in information and communication engineering at BUPT. His current research interests include computer vision and image retrieval.

**GUO Caili** (guocaili@bupt.edu.cn) received the Ph.D. degree in communication and information systems from Beijing University of Posts and Telecommunications (BUPT), China in 2008. She is currently a professor in the School of Information and Communication Engineering, BUPT. Her general research interests include machine learning and statistical signal processing, with current emphasis on semantic communications, deep learning, and intelligence visual computing. In the related areas, she has published over 200 papers and holds over 30 granted patents. She won Diamond Best Paper Award of IEEE ICME 2018 and Best Paper Award of IEEE WCNC 2021.

**LI Zheng** received the B.S. degree in telecommunication engineering from Shandong University, China in 2016, and the M.S. degree in information and communication engineering from Beijing University of Posts and Telecommunications (BUPT), China in 2019. He is currently pursuing the Ph.D. degree in information and communication engineering at BUPT. His current research interests include computer vision and multimedia retrieval.

# ZTE Communications Guidelines for Authors

## Remit of Journal

*ZTE Communications* publishes original theoretical papers, research findings, and surveys on a broad range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics and industry researchers from around the world.

## Manuscript Preparation

Manuscripts must be typed in English and submitted electronically in MS Word (or compatible) format. The word length is approximately 3 000 to 8 000, and no more than 8 figures or tables should be included. Authors are requested to submit mathematical material and graphics in an editable format.

## Abstract and Keywords

Each manuscript must include an abstract of approximately 150 words written as a single paragraph. The abstract should not include mathematics or references and should not be repeated verbatim in the introduction. The abstract should be a self-contained overview of the aims, methods, experimental results, and significance of research outlined in the paper. Five carefully chosen keywords must be provided with the abstract.

## References

Manuscripts must be referenced at a level that conforms to international academic standards. All references must be numbered sequentially in-text and listed in corresponding order at the end of the paper. References that are not cited in-text should not be included in the reference list. References must be complete and formatted according to *ZTE Communications* Editorial Style. A minimum of 10 references should be provided. Footnotes should be avoided or kept to a minimum.

## Copyright and Declaration

Authors are responsible for obtaining permission to reproduce any material for which they do not hold copyright. Permission to reproduce any part of this publication for commercial use must be obtained in advance from the editorial office of *ZTE Communications*. Authors agree that a) the manuscript is a product of research conducted by themselves and the stated co-authors; b) the manuscript has not been published elsewhere in its submitted form; c) the manuscript is not currently being considered for publication elsewhere. If the paper is an adaptation of a speech or presentation, acknowledgement of this is required within the paper. The number of co-authors should not exceed five.

## Content and Structure

*ZTE Communications* seeks to publish original content that may build on existing literature in any field of communications. Authors should not dedicate a disproportionate amount of a paper to fundamental background, historical overviews, or chronologies that may be sufficiently dealt with by references. Authors are also requested to avoid the overuse of bullet points when structuring papers. The conclusion should include a commentary on the significance/future implications of the research as well as an overview of the material presented.

## Peer Review and Editing

All manuscripts will be subject to a two-stage anonymous peer review as well as copyediting, and formatting. Authors may be asked to revise parts of a manuscript prior to publication.

## Biographical Information

All authors are requested to provide a brief biography (approx. 100 words) that includes email address, educational background, career experience, research interests, awards, and publications.

## Acknowledgements and Funding

A manuscript based on funded research must clearly state the program name, funding body, and grant number. Individuals who contributed to the manuscript should be acknowledged in a brief statement.

## Address for Submission

<http://mc03.manuscriptcentral.com/ztecom>

# ZTE COMMUNICATIONS

## 中兴通讯技术(英文版)

### ZTE Communications has been indexed in the following databases:

- Abstract Journal
- Cambridge Scientific Abstracts (CSA)
- China Science and Technology Journal Database
- Chinese Journal Fulltext Databases
- Index of Copernicus
- Ulrich's Periodicals Directory
- Wanfang Data
- WJCI 2021

---

### Industry Consultants:

DUAN Xiangyang, GAO Yin, HU LiuJun, LIU Xinyang,  
LU Ping, SHI Weiqiang, WANG Huitao, XIONG Xiankui,  
ZHU Fang, ZHU Xiaoguang, ZHAO Zhiyong

---

### ZTE COMMUNICATIONS

Vol. 20 No. 1 (Issue 78)

Quarterly

First English Issue Published in 2003

#### Supervised by:

Anhui Publishing Group

#### Sponsored by:

Time Publishing and Media Co., Ltd.

Shenzhen Guangyu Aerospace Industry Co., Ltd.

#### Published by:

Anhui Science & Technology Publishing House

### Edited and Circulated (Home and Abroad) by:

Magazine House of ZTE Communications

#### Staff Members:

General Editor: WANG Xiyu

Editor-in-Chief: JIANG Xianjun

Executive Editor-in-Chief: HUANG Xinming

Editorial Director: LU Dan

Editor-in-Charge: ZHU Li

Editors: REN Xixi, XU Ye, YANG Guangxi

Producer: XU Ying

Circulation Executive: WANG Pingping

Assistant: WANG Kun

---

### Editorial Correspondence:

Add: 12F Kaixuan Building, 329 Jinzhai Road,  
Hefei 230061, P. R. China

Tel: +86-551-65533356

Email: magazine@zte.com.cn

Website: <http://zte.magtechjournal.com>

### Annual Subscription: RMB 120

#### Printed by:

Hefei Tiancai Color Printing Company

Publication Date: March 25, 2022

China Standard Serial Number:  $\frac{\text{ISSN } 1673-5188}{\text{CN } 34-1294/\text{TN}}$