João Henrique Inacio de Souza

Physical Layer Optimization, Multiple Access, and Energy Efficiency in Extra-Large Scale Massive MIMO and Wireless Networks Aided by Reconfigurable Intelligent Surfaces

Londrina
2024

João Henrique Inacio de Souza

Physical Layer Optimization, Multiple Access, and Energy Efficiency in Extra-Large Scale Massive MIMO and Wireless Networks Aided by Reconfigurable Intelligent Surfaces

Doctoral thesis presented to the Programa de Pós-Graduação Associado em Engenharia Elétrica UEL/UTFPR-CP at the Universidade Estadual de Londrina for obtaining the degree of Doctor of Science.

Universidade Estadual de Londrina – UEL
Centro de Tecnologia e Urbanismo – CTU
Departamento de Engenharia Elétrica – DEEL

Programa de Pós-Graduação Associado em
Engenharia Elétrica UEL/UTFPR-CP (Doutorado)

Supervisor: Taufik Abrão
Co-supervisor: José Carlos Marinello Filho

Londrina
2024

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Inacio de Souza, João Henrique.
Physical Layer Optimization, Multiple Access, and Energy Efficiency in Extra-Large Scale Massive MIMO and Wireless Networks Aided by Reconfigurable Intelligent Surfaces / João Henrique Inacio de Souza. - Londrina, 2024.
141 f. : il.

Orientador: Taufik Abrão.
Coorientador: José Carlos Marinello.
Tese (Doutorado em Engenharia Elétrica) - Universidade Estadual de Londrina, Centro de Tecnologia e Urbanismo, Programa de Pós-Graduação em Engenharia Elétrica, 2024.
Inclui bibliografia.

1. Massive multiple-input multiple-output (MIMO) - Tese. 2. Extra-large scale massive MIMO (XL-MIMO) - Tese. 3. Reconfigurable intelligent surface (RIS) - Tese. 4. Energy efficiency - Tese. I. Abrão, Taufik. II. Marinello, José Carlos. III. Universidade Estadual de Londrina. Centro de Tecnologia e Urbanismo. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

CDU 62

João Henrique Inacio de Souza

Physical Layer Optimization, Multiple Access, and Energy Efficiency in Extra-Large Scale Massive MIMO and Wireless Networks Aided by Reconfigurable Intelligent Surfaces

Doctoral thesis presented to the Programa de Pós-Graduação Associado em Engenharia Elétrica UEL/UTFPR-CP at the Universidade Estadual de Londrina for obtaining the degree of Doctor of Science.

Approved thesis. Londrina, 1st March 2024:

Aldebaro Barreto da Rocha Klautau Júnior
Universidade Federal do Pará

Djamel Fawzi Hadj Sadok
Universidade Federal de Pernambuco

João Luiz Rebelatto
Universidade Tecnológica Federal do Paraná

Richard Demo Souza
Universidade Federal de Santa Catarina

Taufik Abrão (Supervisor)
Universidade Estadual de Londrina

Londrina
2024

Acknowledgements

This thesis is the result of a long journey that began ten years ago, in 2014, when I stepped into UEL for the first time. For this reason, I couldn't start without thanking UEL and all the people who make it a reality. UEL is my second home that has gifted me with incredible opportunities and became, luckily for me, the place where I could take my first steps as a junior researcher.

I thank Professor Taufik for his guidance, encouragement, and hard work. Furthermore, I thank Professor Marinello for his valuable collaboration. I thank Professor Petar for the inestimable opportunity to work in Aalborg. I also thank my friends from Londrina and Aalborg for the intense technical discussions balanced with memorable and fun times.

Most importantly, I thank my parents, Vera and José, for their unconditional support. Finally, I thank my brothers, Magda and José, and my partner Jorge for their love.

This work was supported in by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and by the National Council for Scientific and Technological Development (CNPq) of Brazil under Grant 141445/2020-3.

Abstract

INACIO DE SOUZA, João Henrique. **Physical Layer Optimization, Multiple Access, and Energy Efficiency in Extra-Large Scale Massive MIMO and Wireless Networks Aided by Reconfigurable Intelligent Surfaces**. 2024. Thesis (Doctorate in Electrical Engineering) – Universidade Estadual de Londrina, Londrina, 2024.

The mobile networks beyond the fifth generation (B5G) must be designed to supply an increasing demand for connectivity, coming not only from established applications but also from emerging ones that surge with progressively ambitious communication requirements. To that end, physical layer technologies such as extra-large scale massive multiple-input multiple-output (XL-MIMO) and reconfigurable intelligent surfaces (RISs) have been considered for integrating the specifications of the next-generation networks aiming to boost the key performance indicators (KPIs). XL-MIMO encompasses communication systems where the base station is equipped with a physically large antenna array, with hundreds to thousands of half-wavelength-spaced antennas. Seen as an evolution of massive MIMO, in XL-MIMO systems, the array provides extra spatial degrees of freedom (DoFs) that can be explored to spatially multiplex many users with high data rates. Furthermore, an RIS is a thin sheet of composite material equipped with electronic circuits that can be programmed to change the characteristics of the incoming electromagnetic field. It can be used to coherently reflect radio signals, enhancing the wireless channel in arbitrary spots of the service area, relying on low-cost and energy-efficient hardware. To unlock the potential of XL-MIMO and RIS, it is necessary to solve open research questions concerning channel modeling, system design, and optimization strategies. Addressing the challenges faced in the development of these technologies, this thesis investigates the applicability of XL-MIMO and RIS to leverage the communication KPIs in the B5G networks. Based on the B5G usage scenarios, it focuses on four objectives: 1) providing high data rates and supporting high connection density for gigabit per second communication; 2) enabling energy-efficient wireless communication for the Internet of Things (IoT); 3) providing reliable and low-latency communication for mission-critical applications; and 4) investigating the potential of RISs to program aspects of the wireless channel. Regarding XL-MIMO, to support multi-user communication in crowded environments, a strategy for system optimization is proposed to efficiently explore all the DoFs provided by the high-aperture array. Furthermore, concerning RIS, the energy efficiency of RIS-aided IoT networks is analyzed, producing relevant insights for system design aiming to extend the devices' battery lifetime and enhance the network coverage. Moreover, a multiple access scheme to multiplex hybrid traffic is proposed, using the RIS to support the coexistence through network resource sharing of mission-critical services along with broadband communication services. Finally, tackling the integration of the RIS controllability and exploring its capability of shaping the wireless channel, a method is proposed to control the channel

temporal statistics by using an RIS with time-varying stochastic configurations. In short, this thesis presents methods, procedures, and algorithms to implement XL-MIMO systems and RISs in the B5G networks, accompanied by comprehensive evaluations of the system trade-offs in terms of relevant KPIs, including spectral efficiency, energy efficiency, outage probability, and latency.

Keywords: Massive multiple-input multiple-output (MIMO), extra-large scale massive MIMO (XL-MIMO), reconfigurable intelligent surface (RIS), energy efficiency, Internet of Things (IoT), random access protocols.

Resumo

INACIO DE SOUZA, João Henrique. **Physical Layer Optimization, Multiple Access, and Energy Efficiency in Extra-Large Scale Massive MIMO and Wireless Networks Aided by Reconfigurable Intelligent Surfaces**. 2024. Tese (Doutorado em Engenharia Elétrica) – Universidade Estadual de Londrina, Londrina, 2024.

As redes de comunicações móveis da próxima geração devem ser projetadas para suprir uma demanda crescente por conectividade, provocada não apenas por aplicações consolidadas, mas também por aplicações emergentes, que apresentam requisitos de comunicação progressivamente ambiciosos. Neste sentido, visando impulsionar os indicadores desempenho, tecnologias de camada física conhecidas como *massive multiple-input multiple-output* (XL-MIMO) e *reconfigurable intelligent surfaces* (RISs) têm sido consideradas para integrar as especificações das redes da próxima geração. Em um sistema XL-MIMO, a estação rádio base é equipada com um arranjo de antenas com grandes dimensões físicas, possuindo de centenas a milhares de antenas espaçadas por meio comprimento de onda. Tratado como evolução do sistema *massive* MIMO, em um sistema XL-MIMO, o arranjo de antenas fornece graus de liberdade espacial adicionais, que podem ser explorados para aplicar multiplexação espacial de diversos usuários a altas taxas. Além disso, uma RIS consiste em uma superfície delgada constituída por materiais compósitos e equipada com circuitos eletrônicos, capaz de ser controlada para alterar características do campo eletromagnético incidente. Especificamente, esta superfície pode ser empregada para refletir sinais de rádio coerentemente, melhorando a qualidade do canal sem fio em pontos arbitrários da área se serviço, dependendo de um hardware de baixo custo e energeticamente eficiente. Para explorar o potencial dos sistemas XL-MIMO e das RISs, é necessário solucionar questões de pesquisa em aberto, relacionadas à modelagem do canal, projeto de sistemas e estratégias de otimização. Focando nos desafios encontrados durante o desenvolvimento destas tecnologias, esta tese investiga a aplicabilidade de XL-MIMO e RIS para melhorar os indicadores de desempenho nas redes da próxima geração. Baseando-se nos cenários de uso definidos para a próxima geração, ela concentra-se em quatro objetivos: 1) fornecer altas taxas de transmissão e suportar alta densidade de conexões para comunicação a gigabit por segundo; 2) viabilizar comunicação sem fio energeticamente eficiente para internet das coisas; 3) fornecer comunicação confiável e de baixa latência para aplicações de missão crítica; e 4) investigar o potencial das RISs para programar aspectos do canal de propagação sem fio. Em relação a XL-MIMO, para suportar comunicação multi-usuário em ambientes com alta densidade de conexões, uma estratégia para otimização do sistema é proposta visando explorar eficientemente todos os graus de liberdade fornecidos pelo arranjo de antenas. Por outro lado, envolvendo redes auxiliadas por RIS, é analisada a eficiência energética de redes para aplicações de internet das coisas, produzindo observações relevantes para o projeto de sistemas visando ampliar o tempo de bateria dos dispositivos

e expandir a cobertura da rede. Além disso, é proposto um esquema de múltiplo acesso para multiplexar tráfico híbrido, usando a RIS para viabilizar a coexistência de serviços de missão crítica e comunicação de banda larga através do compartilhamento de recursos. Por fim, visando estudar a integração da controlabilidade da RIS e explorando sua habilidade de moldar o canal de comunicação, é proposto um método para o controle da estatística temporal do canal adotando-se configurações da RIS estocásticas e variantes no tempo. Em resumo, esta tese apresenta métodos, procedimentos e algoritmos para implementar sistemas XL-MIMO e RISs nas redes da próxima geração, acompanhados por avaliações extensivas dos compromissos oferecidos pelo sistema de comunicação em termos de indicadores de desempenho relevantes, incluindo eficiência espectral, eficiência energética, probabilidade de interrupção e latência.

Palavras-chave: *Massive multiple-input multiple-output* (MIMO), *extra-large scale massive* MIMO (XL-MIMO), *reconfigurable intelligent surface* (RIS), eficiência energética, internet das coisas, protocolos de acesso aleatório.

List of Figures

List of Tables

List of Abbreviations and Acronyms

3GPP	3rd Generation Partnership Project
5G	Fifth generation
5G NR	5G new radio
6G	Sixth generation
ACF	Autocorrelation function
AF	Amplify-and-forward
AI	Artificial intelligence
AR(1)	First-order autoregressive
AWGN	Additive white Gaussian noise
B5G	Beyond the fifth generation
BS	Base station
CCF	Cross-correlation function
CSI	Channel state information
DF	Decode-and-forward
DL	Downlink
DMRS	Demodulation reference signal
DoF	Degree of freedom
EBL	Expected battery lifetime
EE	Energy efficiency
eMBB	Enhanced mobile broadband
IMT-2020	International Mobile Telecommunications for 2020 and beyond
IMT-2030	International Mobile Telecommunications for 2030 and beyond
IN	Interference nulling
IoT	Internet of Things

ITU	International Telecommunication Union
KPI	Key performance indicator
LoS	Line-of-sight
MAC	Medium access control
MIMO	Multiple-input multiple-output
mMTC	Massive machine type communication
NLoS	Non-line-of-sight
NP	Nondeterministic polynomial-time
OFDM	Orthogonal frequency-division multiplexing
PR	Phasors rotation
QoS	Quality of service
RA	Random access
RF	Radio frequency
RHS	Right-hand side
RIS	Reconfigurable intelligent surface
Rx	Receiver
RZF	Regularized zero-forcing
SDG	Sustainable development goal
SE	Spectral efficiency
SIC	Successive interference cancellation
SINR	Signal-to-interference-plus-noise ratio
SNR	Signal-to-noise ratio
TTI	Transmission time interval
Tx	Transmitter
UE	User's equipment
UL	Uplink

UN	United Nations
UPA	Uniform planar array
URLLC	Ultra-reliable-low-latency communication
UWG	Undirected weighted graph
VR	Cisibility region
WSS	Wide-sense stationary
XL-MIMO	Extra-large scale massive MIMO
XR	Extended reality
ZF	Zero-forcing

Contents

Part I

Introductory Chapters

1 Introduction and Motivation

The demand for mobile communication services is growing quickly in the last decades, both in terms of connected devices and traffic volume. Moreover, as the process of digitization of the society progresses, new use cases and applications are emerging with increasingly complex communication constraints. The total global mobile data traffic is projected to grow from 118 EB to 472 EB per month in the interval from the end of 2022 to 2028, a growth by a factor of 4 [1]. Especially, a relevant contributor to this growth should be the extended reality (XR) services, newly data-intensive applications that require the exchange of video, audio, and other multisensory data streams. To address such applications, improvements on the data rates, communication latency, and radio unit capacity are required, taking the mobile networks of the fifth generation (5G) as reference. Furthermore, due to its global widespread, the environmental impacts of the mobile communication infrastructure and the end-user devices need to be assessed and reduced to meet the target global emissions by 2030 and, subsequently, reach the net-zero emissions by 2050 [2]. In summary, the networks beyond the fifth generation (B5G) must address the emerging use cases and applications, complying with the United Nations (UN) sustainable development goals (SDGs)[1] to inflict low impact on the environment and deliver improvements in the quality of life.

The International Telecommunication Union (ITU) framework for the B5G networks have proposed 6 usage scenarios, 3 considered to be evolutions of the 5G usage scenarios and 3 brand-new ones [4]. With applications in the areas of education, healthcare, media, transportation, manufacturing, agriculture, and energy, each usage scenario presents its own relevant key performance indicators (KPIs), and the interplay between them forms the basis to supply the application's demands. For instance, the evolution of enhanced mobile broadband (eMBB), the *immersive communication* usage scenario, should provide high data rate communication with improved latency and high connection density to enable the simultaneous operation of many XR devices equipped with haptic feedback. On the other hand, the newly introduced *ubiquitous connectivity* usage scenario should bridge the digital divide by providing connectivity to the unconnected or scarcely covered areas, focusing on the coverage KPIs. Therefore, to address the emerging applications, the mobile networks must have pervasive presence and provide uniform quality of service (QoS) to the end-user devices, while supporting compound services that depend on multiple, potentially conflicting, communication constraints.

The global threat of climate change urges the necessity to cut the greenhouse gas

[1] The 17 SDGs were established in 2015 by the UN in the 2030 Agenda for Sustainable Development, accompanied by 169 economical, social, and environmental targets to, *e.g.*, eradicate poverty, achieve gender equality, mitigate the climate change and its impacts, and implement the sustainable use of the natural resources [3].

emissions, which includes replacing fossil fuels by renewable energy sources. A study shows that most of the energy consumed by mobile operators is used to power up the radio access network [5]. This outcome sheds light on the importance of increasing the energy efficiency (EE) and reducing the energy consumption of the mobile networks by improving hardware designs and optimizing their operation. Hence, to develop B5G networks that have low environmental impact, EE must be introduced as a foundational KPI, while the principles of environmental sustainability must be incorporated in their design and operation [2].

In Brazil, it is estimated that the smartphone is present in 96.6% of the homes, while a lower percentage of homes, precisely, 92%, is covered by a mobile broadband network. Such a coverage statistic is especially alarming in the rural area, where only 69.4% of the homes are in the footprint of a network [6]. Furthermore, it is estimated that 25% of the population that lives in areas covered by a mobile broadband network doesn't use the service. These numbers show that there are *coverage* and *usage gaps* that must be closed to deliver full connectivity to the Brazilian population. Directions to close these gaps include reducing mobile tariffs, improving the device affordability, and incentivizing investments to expand the network coverage, with emphasis on remote and sparsely-populated areas [7]. It is important to stress that such gaps cannot be read only as a revenue opportunity, but also as a duty to provide the necessary connectivity for a significant part of the population to benefit from the blessings of the digital society. The mobile networks enable cheap internet access to the population, being an important tool to improve the quality of life indicators by building universal and meaningful private and public services.

Despite the deployment of the 5G networks in countries like Brazil is still in the early stages, it is time to research, design, and evaluate novel technologies to support the forthcoming applications and address the gaps left by 5G, paving the way for the next-generation mobile networks. From the technological perspective, the challenges described previously can be addressed by the strategies in the following. The support to several compound services can be addressed by developing new network architectures, methods, and algorithms capable of delivering the specific requirements of each service. Moreover, the decreasing of the environmental impact can be attained by optimizing the EE and energy consumption of the infrastructure and end-user devices. Finally, the coverage and usage gaps can be bridged by, respectively, developing cheap technologies to extend the network coverage and reducing the cost of the network infrastructure, which potentially decreases the cost of the mobile broadband service and boosts the capacity of public and private investments on the network expansion. Guided by these directions, this thesis explores technologies for the radio interface intending to develop energy-efficient B5G networks, ready to thrive in the future demanding applications.

1.1 Thesis Objectives

The aim of this thesis is to contribute to the technological development of solutions to the challenges introduced in the previous section, with emphasis on the physical layer aspects of the B5G networks. Precisely, the thesis objectives are:

1) **General:** Develop methods, procedures, and algorithms for physical layer optimization, multiple access, and EE improvement in B5G systems equipped with an extra-large scale massive MIMO (XL-MIMO) transceiver or aided by an RIS, aiming to supply the communication requirements of the compound services and enable the coexistence of them.

2) **Specific:**

    a) Develop optimization methods, procedures, and algorithms in addition to multiple access schemes to provide high data rates and high connection density for gigabit per second communications;

    b) Analyze the performance and trade-offs of energy-efficient system deployments through RISs for Internet of Things (IoT) applications;

    c) Develop multiple access schemes to provide reliable and low-latency communication for mission-critical applications served by RIS-aided systems;

    d) Analyze the potential and develop frameworks to program aspects of the wireless propagation environment through the control of RISs.

1.2 Methodology

Numerical simulations and analytical methods have been used in the development of the contributions of this thesis. Implemented by MATLAB, the simulations are based on stochastic channel models to represent the communications systems dynamics. Moreover, they rely on the Monte-Carlo approach, having as primary sources of randomness the position of the mobile terminals in the environment, the fading channel realizations, and the additive white Gaussian noise (AWGN) at the received signals. When applied, the analytical methods helped to obtain bounds and approximations for performance metrics, and also to develop the frameworks for system design.

Individually, each contribution has its own set of relevant metrics for performance assessment. Fundamental metrics comprise spectral efficiency (SE), which measures the achievable information rate per spectrum unit, and EE, which also measures the achievable information rate, but per unit of consumed energy.[2] Other more specific metrics, but still

[2] Alternatively, the EE can be expressed as the amount of energy required to transmit and correctly receive an information bit, in J/bit.

related to SE and EE, include sum-rate, throughput, outage probability, and expected battery lifetime (EBL). The algorithms and communication schemes proposed in the contributions are analyzed using these metrics, and subsequently compared to state-of-the-art benchmarks.

1.3 Thesis Outline

The thesis is formatted as a collection of papers, and it is divided into two parts. In the remainder of the Part I, Chapters 2 to 4 cover the relevant theoretical background necessary to develop the research. Specifically, Chapter 2 examines the vision for the B5G mobile networks. Meanwhile, Chapters 3 and 4 give an introduction about XL-MIMO and RISs, respectively. Chapter 5 summarizes the contributions of the thesis, describing the main findings of each scientific paper that make up this thesis. Then, Chapter 6 provides the conclusions and future work directions. Part II contains the full version of the published and submitted scientific papers summarized in Chapter 5 of Part I, arranged in order of conception.

2 Mobile Networks Beyond the Fifth Generation

This chapter reports on the current status of the research of B5G networks by academia and industry. Initially, a summary of the vision for the next-generation mobile networks given by relevant initiatives is provided. Then, the usage scenarios of the International Mobile Telecommunications for 2030 and beyond (IMT-2030) are presented, with their main KPIs along with the suggested target values. Finally, enabling technologies with potential to integrate the B5G networks are introduced.

The research and development of technologies for the next-generation networks have been conducted by several initiatives comprising universities, governments, and industries. To identify the demands and challenges with the aim of finding promising paths for such a process, each initiative builds a particular vision about the next-generation networks. Specifically, the vision of the Hexa-X[1] initiative comprises B5G networks designed to connect the human, digital, and physical worlds based on three values: *trustworthiness*, *digital inclusiveness*, and *sustainability* [8]. They highlight that, ensuring privacy, security, and reliability, the B5G networks will be enablers of large-scale deployments of artificial intelligence (AI) services, transforming AI into a vital technology for the end-user life. From a different perspective, the Next G Alliance[2] has built a vision of the sixth generation (6G) systems based on six goals: 1) trust, security, and resilience; 2) digital world experience; 3) cost efficient solutions; 4) distributed clouding and communications systems; 5) AI-native network; and 6) sustainability [9]. Also giving high priority to AI, they argue that increased robustness, performance, and efficiencies can be attained by incorporating AI into the system's design and development cycle. To that end, AI tools should be used for the design, deployment, management, and operation of network functions and devices, including users' equipment (UEs), base stations (BSs), and network core equipment.

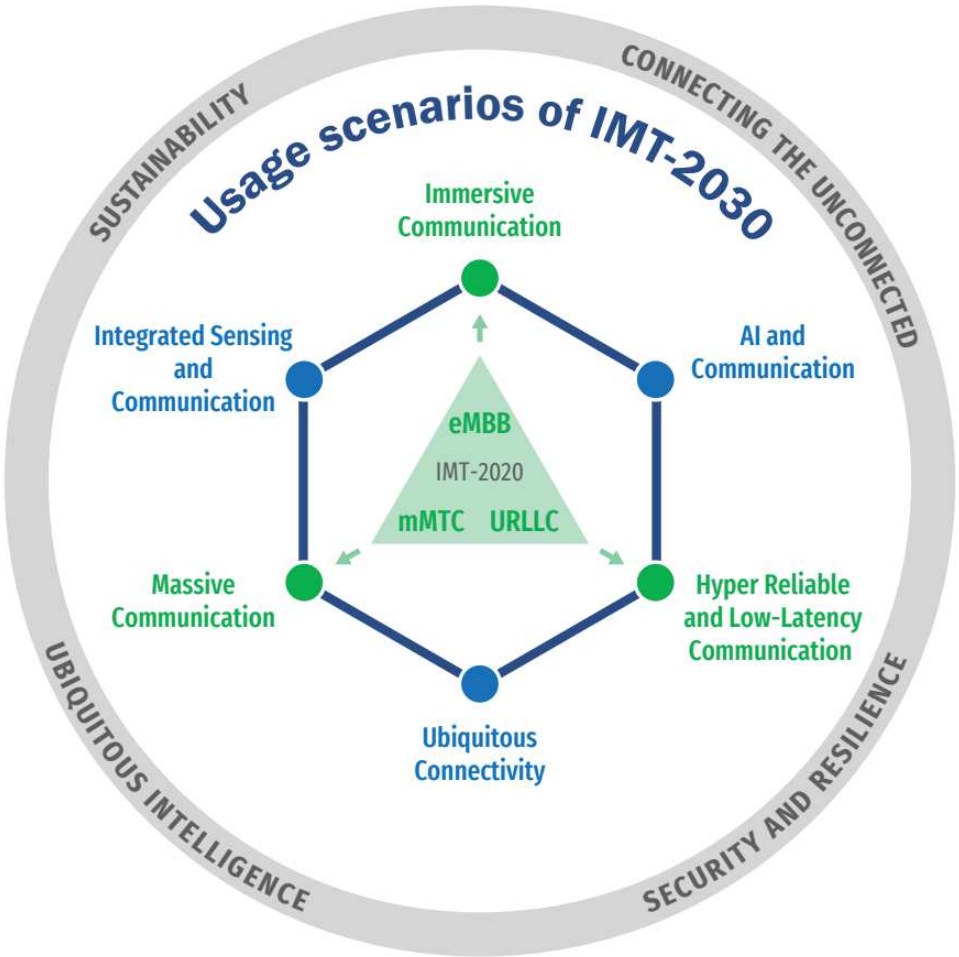
Another essential point shared by the visions of these two initiatives is the need to account for environmental sustainability, both in the B5G networks and the applications that use them. To address this aim, new KPIs must be incorporated in the design and operation phases to reduce the networks' environmental impact on material, water, land, and air. Furthermore, the networks' environmental impact will be decreased with the adoption of system architectures and optimization strategies to reduce the dependence on non-renewable energy sources, prioritizing renewable sources, to improve the EE of networks and devices, and to reduce the carbon footprint of the network infrastructure hardware and software [9].

[1] `https://hexa-x.eu`
[2] `https://www.nextgalliance.org`

2.1 IMT-2030 Usage Scenarios

In the development of 5G, the usage scenarios of IMT for 2020 and beyond (IMT-2020) have defined 3 services with different combinations of requirements that oriented the research and standardization, *i.e.*, eMBB, URLLC, and massive machine type communication (mMTC) [10]. For the B5G networks, the framework for IMT-2030 proposes 6 usage scenarios, 3 based on an evolution of the IMT-2020 scenarios to provide improved 5G services, and 3 brand-new ones to accommodate the forthcoming applications [4]. The IMT-2030 usage scenarios are depicted in Fig. 1 with the IMT-2030 design principles: sustainability, security and resilience, connecting the unconnected, and ubiquitous intelligence. Furthermore, Table 1 shows a comparison of the minimum technical performance requirements defined by ITU for the 5G and B5G networks.

Figure 1 – IMT-2030 usage scenarios and design principles.



Source: [4, Fig. 1].

A description of each usage scenario, containing the respective target applications, projected features, and main performance metrics, is given in the following.

1) **Immersive communication:** As an evolution of eMBB, this usage scenario em-

Table 1 – Minimum technical performance requirements defined by ITU for the 5G and B5G networks.

Capability	**5G**	**B5G**
Peak data rate (Gbps)	10 and 20	50, 100, and 200
User experienced data rate (Mbps)	50 and 100	300 and 500
Average spectral efficiency (bps/Hz/TRxP*)	1.6 to 9	1.5 to 3 times higher
Area traffic capacity (Mbps/m^2)	10	30 and 50
Connection density (devices/km^2)	10^6	10^6 to 10^8
Mobility (km/h)	500	500 to 1000
Latency (ms)	1	0.1 to 1
Reliability	$1-10^{-5}$	$1-10^{-5}$ to $1-10^{-7}$
Positioning accuracy (cm)	N/A	1 to 10

*TRxP: transmission reception point

Source: [11, Section 4] and [4, Section 4].

braces interactive applications related to XR, requiring the support of transmission and reception of time-synchronized multisensory data streams, including audio, video, and environment data collected by sensors. The main communication requirements involve increased data rates, mobility, and system capacity. In particular cases, combined enhanced reliability and low latency are required for responsive and accurate interaction. Hence, the associated KPIs are SE and data rate, mobility measured by the maximum supported speed, and connection density.

2) **Hyper reliable and low-latency communication:** Clearly an evolution of URLLC, this usage scenario covers mission-critical applications, such as emergency alert and disaster recovery systems, and also applications for full industrial automation, control, and operation. Reliability and latency are the fundamental KPIs of this usage scenario. However, depending on the application, positioning precision and connection density might be additional relevant KPIs.

3) **Massive communication:** Extending the features offered by mMTC, this usage scenario comprises the IoT applications, which have an important role in the fields of healthcare, transportation, environmental monitoring, smart cities, and energy. To better address these applications, an improvement in the supported number of connected devices is required, making connection density the main KPI of this scenario. Nonetheless, coverage, power consumption, security, and reliability are potential extra KPIs.

4) **Ubiquitous connectivity:** This novel usage scenario was introduced with the aim of bridging the digital divide by connecting the unconnected or scarcely covered areas, delivering to them the minimum connectivity necessary to use communication services. The foreseen feature to realize this objective is the integration of the

mobile networks with other terrestrial and non-terrestrial networks, such as radio local access networks and satellite communications systems, respectively. Certainly, network coverage is the relevant KPI in this scenario.

5) **Artificial intelligence and communication:** Also a novel usage scenario, it was conceived envisioning mobile networks optimized for applications that require distributed computing and AI services. To this end, the mobile networks must be equipped with features related to, *e.g.*, AI model training and orchestration of computing resources. Particularly, this scenario comprises a multitude of applications that can benefit from AI services, such as autonomous driving and digital twins. Furthermore, it covers applications in which the terminals need to offload computation tasks across other devices to deal with resource-intensive operations. The main KPIs associated to this usage scenario are area traffic capacity and user experienced data rate.

6) **Integrated sensing and communication:** The last newly-introduced scenario embraces applications that will be able to exploit contextual information about the environment, measured and communicated by the mobile network components. Such applications include, among others, assisted navigation, presence detection, gesture recognition, fall detection, and environmental monitoring. Especially in this scenario, the KPIs are tied to the type of required sensing information, being positioning precision an illustrative example.

2.2 Enabling Technologies

Several enabling technologies have been investigated with the aim of developing systems capable of attaining the communication requirements organized in Table 1, enabling the support to the applications described in Section 2.1 To further improve the expressive SE gains achieved by transceivers with multiple antennas, advanced massive multiple-input multiple-output (MIMO) techniques have been extensively studied. On this matter, the cell-free massive MIMO architecture has potential of enhancing performance and coverage simultaneously. This architecture eliminates the cell boundaries by spreading many access points over the service area, allowing the UEs to communicate through limited sets of access points selected according to the channel conditions. As a result, cell-free massive MIMO yields a high macro-diversity degree, providing uniform QoS in the whole service area. Importantly, practical implementations of cell-free massive MIMO depend on the development of advanced techniques for synchronizing the geographically distributed access points with the central processing unit [12]. On the other hand, the XL-MIMO systems simultaneously increase the SE and the BS capacity by deploying physically large arrays with hundreds to thousands of antennas. The feasibility of this architecture relies on the

better characterization of the wireless channel [13], in addition to the development of low-complexity algorithms for channel estimation and signal processing to tackle the huge numbers of antennas [14–16].

The list of promising enabling technologies for the B5G networks also includes the RISs. In particular, software-controlled intelligent surfaces capable of altering the signal propagation can be used to enable a programmable wireless environment [17]. Additionally, substantial gains in SE, network coverage and EE can be attained by implementing RIS-aided communications systems. The success of RISs comes from their power-efficient and low-cost hardware, combined with a low-profile construction that enables flexible deployments, both in indoor and outdoor environments. To realize practical RIS-aided systems, there is a need for efficient strategies to optimize the RIS phase shift configuration, preferably using partial channel state information (CSI) or side information that generates small control overhead. Moreover, more accurate channel models to account for the RIS operation in the radiative near-field region are also needed [18]. Finally, it is necessary to explore other RIS architectures, testing different combinations in terms of number and arrangement of active and passive elements, and to develop hardware designs aiming to obtain RIS elements capable of changing several properties of the electromagnetic fields simultaneously.

3 Extra-Large Scale Massive MIMO

Massive MIMO was one of the main drivers of SE in 5G networks. In massive MIMO systems, the BS is equipped with an array of many antennas to support many UEs simultaneously. In short, its benefits come from three fundamental properties: *channel hardening*, which makes the fading channel behaves as deterministic; *favorable propagation*, resulting in asymptotic mutual orthogonality among the UEs' channel vectors; and large array gain, proportional to the number of BS antennas [19]. These properties helped to deliver high area SE and increased connection density through spatial multiplexing, implemented with low computational complexity by linear transmit precoding and receive combining schemes. The current massive MIMO implementations are restricted to compact planar arrays with tens of antennas. There is an interest to increase both the number of antennas and the array size to extend the number of spatial dimensions offered by the channel, further improving SE and connection density. However, the increase in both the number of antennas and the physical dimensions of the array implies significant changes in channel modelling, hardware design, signal processing considerations, and fundamental limits for the system performance.

In this regard, XL-MIMO systems use arrays with hundreds to thousands of antennas, attaining dimensions of several tens of meters in sub-6 GHz frequency bands. Designed to address crowded environments due to the many spatial dimensions offered by their channels, the physically large arrays can be installed close to UEs and integrated into large structures, such as the walls and ceiling of airports, stadiums, and shopping malls. Importantly, XL-MIMO is treated as a distinct operation regime of massive MIMO mainly due to the properties of the wireless channel. To clarify the differences between the technologies, Table 2 presents a comparison between massive MIMO and XL-MIMO regarding transceiver architecture, channel characteristics, and signal processing aspects. In particular, as the physical dimensions of the array increase, *spatial non-stationarities* appear on the channel, since different parts of the array experience the same multipath channel components with different powers, or even distinct components [20]. Moreover, the XL-MIMO channel is characterized by *visibility regions (VRs)* created by path loss and the presence of obstacles in the environment [21]. In practice, the energy of the transmitted signal by a single UE will be significantly concentrated on a portion of the array, which is called the UE's VR. Hence, uplink (UL) detection and downlink (DL) transmission can be done efficiently by processing only the antennas inside the UE's VR. Also, due to the increased aperture of the XL-MIMO arrays, the UEs are likely located in the *radiative near-field region* of the BS array they are connected to. In this case, the amplitude and phase variations over the array are characterized by the exact distances between each pair of BS-UE antennas. Therefore, the emitted waves that arrive to or depart from

the array are better modelled by *spherical wavefronts* [22]. The evident consequence of these propagation properties is the necessity of dedicated XL-MIMO channel modelling, different from the conventional models employed for massive MIMO systems. Besides, impressively, these same properties can be exploited to improve the BS capabilities, further expanding the gains obtained by transceivers with multiple antennas. Fig. 2 illustrates an XL-MIMO deployment in an indoor enviroment with high user density. The properties of the XL-MIMO channel are represented by the several beams departing from the highlighted users.

Table 2 – Comparison of main aspects between massive MIMO and XL-MIMO.

Aspect	**Massive MIMO**	**XL-MIMO**
Transceiver architecture	• Tens of antennas, up to few hundreds • Co-located deployments of compact uniform planar arrays	• Hundreds to thousands of antennas • Arrays of large dimensions embedded in environment structures • Distributed deployments based on sub-arrays
Channel characteristics	• UEs likely experience propagation in the array radiative *far-field* region • Emitted waves are better modelled by *plane wavefronts* • *Stationary* spatial correlation over the array	• UEs likely experience propagation in the array radiative *near-field* region • Emitted waves are better modelled by *spherical wavefronts* • *Non-stationary* spatial correlation over the array • Visibility regions
Signal processing	• *Linear* beamforming schemes • *Angular-domain* channel estimation	• *Non-linear* beamforming schemes optimized to deal with extremely high numbers of antennas • *Polar-domain* channel estimation • Distributed subarray-based processing

Source: The author.

The properties of the XL-MIMO channel unlock opportunities to develop more efficient signal processing algorithms, multiplexing schemes, and resource allocation strategies. For instance, the UEs' VRs can be seen as an additional dimension for spatial multiplexing, improving signals' separability. Exploiting this characteristic, the authors in [23] proposed low-complexity receivers for a system architecture based on subarrays. Specifically, they resort to the fact that the UL signal of each UE reaches only part of the BS, arriving to a limited number of subarrays. Therefore, using a strategy inspired by *coded random access*, zero-forcing (ZF) detection is carried out locally at each subarray, and then the interference produced by the detected symbols is removed from the other subarrays via successive interference cancellation (SIC). The VRs also motivated the authors in [24] to develop a random access (RA) protocol with gains on sum-rate and access latency. These gains are attained by the possibility of assigning the same communication resource to several UEs, owing to the fact that UEs with non-overlapping VRs have mutually-orthogonal channels. Furthermore, communication in the radiative near-field region of the array enables

Figure 2 – Example of XL-MIMO deployment for providing communication in an indoor environment with high user density using six physically large arrays. Due to the properties of the XL-MIMO channel, the signals transmitted by the highlighted mobile users reach each subarray with different power levels.

Source: The author, with original photo by Benoit Dujardin.

unprecedented multiplexing schemes, such as *depth-domain multiplexing* of many UEs and angular multiplexing of data streams to a single UE [22]. In depth-domain multiplexing, the UEs' channels are separable not only in the *polar* and *azimuth* dimensions, but also in the *radial distance*, increasing the BS multiplexing capacity. These examples demonstrate that, when properly considered, the odd properties of the XL-MIMO wireless channel unveil novel strategies to improve the communication performance.

One of the main challenges on the development of XL-MIMO systems is managing the high computational complexity required to process the many signals transmitted and received by the thousands of antennas at the BS. Accordingly, such extremely high data volume generated by the antennas restricts the hardware architecture on building scalable systems, since fully-connected architectures might require a prohibitive interconnection bandwidth [25]. To circumvent these technical difficulties, low-complexity signal processing algorithms have been developed, along with distributed transceiver architectures. Particularly, in [15, 16], the authors proposed decentralized XL-MIMO receivers focusing on reducing complexity and attaining scalable designs. The receivers, based on *belief propagation* and the *variational message-passing algorithm*, operate independently at each local processing unit connected to a single BS subarray. Ensuring scalability, these units do not rely on a central processing unit, exchanging information solely with neighboring local processing units to improve detection through SIC. These solutions are innovative

examples of scalable and modular system designs that can be applied to different scenarios and adapted according to environment changes.

In XL-MIMO systems, there is a noticeable interest on developing *non-linear beamforming* schemes, *i.e.*, transmit precoding and receive combining, since the linear schemes from massive MIMO suffer from high complexity due to the number of BS antennas. Illustrating, the regularized zero-forcing (RZF) receiver relies on the inversion of a matrix with size equal to the number of BS antennas, a resource-intensive procedure when using standard inversion algorithms. Addressing this specific issue, the authors in [14] proposed non-linear XL-MIMO receivers that emulate the RZF based on *randomized Kaczmarz algorithms*. Specifically, the receivers exploit that the original RZF optimization problem can be reformulated as a system of linear equations. Then, the randomized Kaczmarz algorithms, which are employed to efficiently solve large systems of linear equations, are applied to approximate the RZF solution with lower computational complexity and marginal performance degradation. Other strategies to develop low-complexity signal processing algorithms incorporate methods such as *polynomial expansion* and *Neumann series*, iterative methods to calculate matrix inversions such as *Gauss-Seidel*, *Jacobi over relaxation*, and *conjugate gradient*, and even AI methods [26]. Continuing to investigate strategies to further decrease the computational complexity of signal processing algorithms is essential to enable XL-MIMO systems, leveraging the benefits of the extremely high number of antennas.

Beyond signal processing algorithms, XL-MIMO systems also face challenges and opportunities regarding resource allocation. Exploiting the UEs' channels separability in the polar-domain due to the propagation regime in the radiative near-field, efficient user scheduling strategies can be developed [27]. Furthermore, due to the massive architecture of XL-MIMO systems, it is recommended to design mechanisms for scaling the system capacity according to the network demand, prioritizing EE. On this matter, to reduce power consumption in times of low system load, system modules responsible for the exceeding capacity can be temporarily turned off. To maximize the EE of an XL-MIMO system, the authors in [28] proposed approaches for *antenna selection* based on the long-term fading parameters of the channel. Based on *evolutionary optimization*, *randomized algorithms*, and *heuristic algorithms* developed from model specificities, the approaches provide a vast collection of trade-offs between EE and computational complexity. Noticeably, the authors devised an iterative method to determine the optimal number of active antennas, corroborating that, for attaining maximum EE, the number of active antennas must be adjusted according to the number of connected UEs. Still on antenna selection, aiming to maximize the DL SE, the authors in [29] developed approaches for jointly determining the set of active antennas and the allocated power levels in transceivers with a limited number of radio frequency (RF) chains. In particular, to preserve system scalability and cut hardware costs, the BS array follows a subarray-based architecture, where each subarray

is equipped with fewer antennas than RF chains and has a partially-connected switching structure. The authors demonstrate via a numerical analysis that, with a limited number of RF chains, proper antenna selection is crucial to exploit the potential of XL-MIMO systems. Moreover, the authors devised a distributed *genetic algorithm* that substantially reduces the volume of exchanged information between the subarrays and the central processing unit, addressing the mentioned issue of high interconnection bandwidth. In summary, the characteristics of the XL-MIMO systems have created paths to develop novel resource allocation strategies, delivering improved communication performance through, *e.g.*, power control, user scheduling and grouping, and antenna selection.

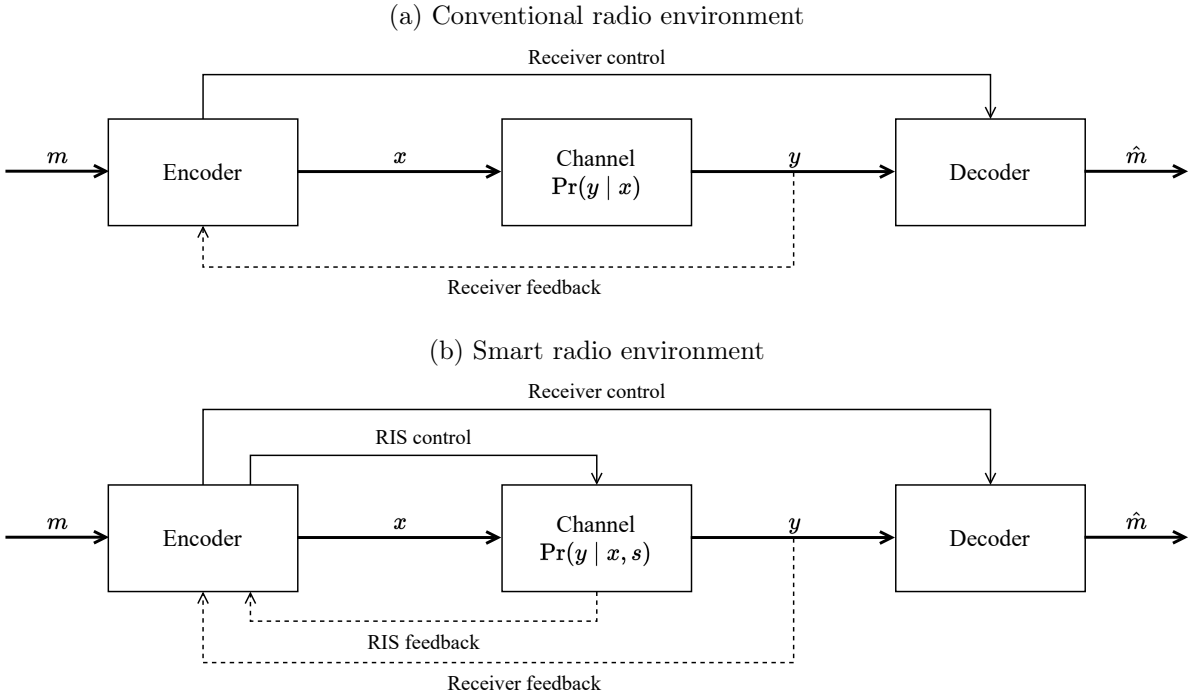
There are still open research challenges that require attention to achieve practical implementations of XL-MIMO systems. A better characterization of the wireless channel is necessary, encompassing the effects of spatial non-stationarities, VRs, and propagation in the array radiative near-field region [13]. In addition, the signal processing algorithms must be tailored according to the channel properties observed under near-field propagation. Another promising research path is the study of array geometries to achieve all the spatial degrees of freedom (DoFs) available in the XL-MIMO channel [22]. Furthermore, the computational complexity reduction can be attained by using AI techniques for channel estimation, beamforming schemes, and resource allocation strategies [26]. Indispensably, future research must pursue the main characteristics desired for an XL-MIMO system, which can be summarized as scalable transceiver architecture, constrained computational complexity, and energy-efficient and low-cost hardware.

4 Reconfigurable Intelligent Surfaces

The development and analysis of communications systems fundamentally bases on the information-theoretic model depicted in Fig. 3a. Considering a point-to-point wireless system, in such a model, the communication endpoints, *i.e.*, the transmitter and receiver, are entities fully planned and optimized by the system designer, while the communication channel, which embodies the propagation environment, is given by nature according to the laws of physics. However, the emergence of technologies such as RISs have shed light on other models in which the channel state can be changed by the system designer, despite still being defined by nature. The differences between these models become clear when comparing them in Fig. 3. On the one hand, in the conventional model in Fig. 3a, the channel is modelled through a fixed *transition probability*, which cannot be changed by other factors than the environment setup. On the other hand, in the model of Fig. 3b, the transition probability is a variable programmed by the designer, and can be optimized accordingly. In particular, models similar to the latter one inspired the concept of *smart radio environment*, where the transmitter, receiver, and communication channel can be jointly optimized, unveiling unprecedented gains on communication performance [18]. A promising path to effectively achieve smart radio environments relies on the development of practical deployments of communications systems aided by RISs, the technology introduced and discussed by this chapter.

An RIS is a thin sheet of composite material equipped with electronic circuits that can be programmed and reconfigured to change characteristics of the incoming electromagnetic field. By appropriately changing such a field, an RIS is capable of dynamically altering the wireless propagation channel, which makes it a relevant enabler for smart radio environments. Still, this capability can also be used to modulate or encode the impinging radio signals. The basic architecture of an RIS comprises an *electromagnetic surface*, a *microcontroller*, and a *network interface controller*. Being the core component of an RIS, the electromagnetic surface is a metamaterial-based continuous sheet or a discrete sheet made upon arranged patches, built with the ability of changing characteristics of the incoming radio waves. Besides, the microcontroller operates the circuits that ensure the RIS reconfigurability, while the network controller provides communication to transmit and receive information about the environment state, side information, and configuration signals. These three components work together to deliver the most attractive feature of RISs, which is the capability of being reconfigured after their deployment in the network [18]. Moreover, the lightweight and low profile construction of the components enables flexible deployments, making it easy to install RISs on indoor and outdoor environments, close to the users that will benefit from their operation. Attractively, the deployment of an RIS can be made transparent to the UEs, as it does not necessarily require any hardware and

Figure 3 – Information-theoretic models for a point-to-point wireless communication system where the relationship between the transmitted and received signals is characterized by the channel transition probability.



Source: Adapted from [18, Fig. 9].

software update at the end-user devices [30]. In principle, an RIS can be developed to apply different elementary functions to the impinging radio waves, such as reflection, refraction, absorption, polarization, collimation, and splitting [18]. By orchestrating these changes all over the electromagnetic surface, there are many possibilities of attaining expressive gains on communication performance due to, *e.g.*, enhancement of the communication channel, suppression of interference, and extension of network coverage. Fig. 4 depicts an example of how an RIS can be deployed to improve network coverage in an indoor environment. This is a promising scheme to enable consistent communication performance in regions of the service area where the UEs don't present line-of-sight (LoS) radio links with the BS, or suffer from severe path loss.

In the literature, the most popular type of RIS presents an architecture with a surface comprised by discrete passive reflecting elements. Specifically, these elements are called passive because they don't apply power amplification to the impinging radio signals and have extremely low power consumption, resulted from the circuits used to configure their states. Individually, each element reflects the impinging radio waves to an arbitrary reflecting angle. Therefore, the reflecting angles applied across the RIS can be configured to collaboratively alter the wireless channel. For instance, if the angles are determined so that the reflected waves interact constructively in a region of the service area, the RIS provides amplification gains on the received signal, improving communication performance in this

Figure 4 – Example of RIS deployment for network coverage improvement in an indoor environment. In this case, the RIS phase shift configuration is optimized to coherently reflect the signals emitted by the BS array, enhancing the received signal power at the highlighted mobile user.

Source: The author, with original photo by Ralph Chang.

specific region. Such a principle is explored to deploy *passive beamforming* strategies, which provides moderate array gain by using low-power passive reflecting elements rather than antennas that need power-hungry RF chains.[1] Moreover, this signal amplification effect can be used to combat severe path loss, extending the network coverage. Conversely, if the reflecting angles are determined so that the waves interact destructively, the RIS decreases the received signal power, potentially improving security if an eavesdropper is located in the target region. Furthermore, due to channel reciprocity, the same principle can be applied to suppress the interference caused by potential undesired transmitters. These two use cases also fall under the topic of passive beamforming, but now with the objective of mitigating the signals received by or transmitted to, *e.g.*, eavesdroppers, interfering devices, and neighboring communication cells. Noticeably, high gains on communication performance can be achieved with a simple RIS architecture based solely on passive reflecting elements. Nonetheless, other more complex architectures, composed by elements that can change different characteristics of the radio waves simultaneously, have been motivated innovative use cases. Exemplifying, a self-configuring RIS architecture is proposed in [31], eliminating the necessity of an active control channel to operate the RIS. Specifically,

[1] In general, RISs are not considered to be a candidate technology to completely replace massive MIMO. In fact, they are seen as an alternative of further increasing the number of antennas at the BS, attaining moderate array gain and providing spatial multiplexing capability through the deployment of large low-cost and low-power passive RISs.

the architecture relies on elements that simultaneously reflect and absorb part of the incident signals, enabling local channel estimation at the RIS and, based on such an information, the autonomous selection of the RIS phase shift configuration. Differently, the authors in [32] investigated an RIS architecture where the elements simultaneously reflect and refract/transmit the impinging waves so that the signals reach both sides of the space surrounding the RIS. Finally, in [33], presenting an innovative application of RISs equipped with reflecting elements, the authors used time-modulated reflection coefficients to enable precise control of the incident waves both in space and frequency domains. Therefore, in practice, the impinging signal with predetermined carrier frequency scatters to several spatial directions with distinct frequency offsets.

The research on RISs mainly concentrates on developing schemes, methods, and algorithms to optimize the RIS configuration aiming to improve the communication performance and leverage the system resources. In particular, the authors in [30] investigated the problem of joint optimization of the BS beamforming and RIS phase shift configuration to minimize the BS transmit power, subject to UEs individual signal-to-interference-plus-noise ratio (SINR) constraints. Considering CSI knowledge at the BS, the proposed schemes approximate solutions to the formulated optimization problem based on *semidefinite relaxation* and *alternating optimization*, resulting in efficient low-complexity algorithms. Remarkably, the results on the single-user scenario revealed that the receive signal-to-noise ratio (SNR) increases quadratically with the number of reflecting elements, being more cost-efficient than massive MIMO and multi-antenna amplify-and-forward (AF) relaying. Focusing on interference suppression, the authors in [34] studied the use of an RIS to spatially-multiplex many transmitter-receiver pairs communicating over the same resource block. Specifically, they proposed an *alternating projection* algorithm that, from CSI, approximates an RIS phase shift configuration that nullifies the mutual interference generated by all devices pairs. In general, the methods to optimize the RIS configuration assume knowledge of CSI, which can be acquired in advance by channel estimation based on UL pilot signals. However, the CSI acquisition can be a resource-intensive procedure in systems aided by passive RISs, since the absence of RF chains and signal processing capacity at the RIS moves the estimation to the BS. On this matter, prior information about the channel fading distribution and spatio-temporal sparsity can be exploited to develop more efficient channel estimation schemes [35]. Still, channel estimation remains an open problem that needs further research. Tackling the issue of CSI availability, in [36], the authors proposed a phase hopping scheme where the RIS randomly reconfigures the individual phase shift of each element at the symbol time-scale. Exceptionally, the scheme substantially improves the communication outage performance regardless of CSI at the RIS and transmitter, being a promising solution for URLLC applications. Furthermore, addressing the integration of RIS into communication protocols, the authors in [37] proposed a *multiple access* framework and a *medium access control (MAC)* protocol design for systems aided by many

RISs. In particular, they assume the practical scenario of static UEs operating together with mobile ones. The proposed framework and protocol encompass key steps prior to UL data transmission, accounting for a pilot transmissions phase along with a computing and feedback phase for UE classification, channel estimation, optimization of the RISs phase shift configuration and other resource allocation tasks.

The extremely low power consumption and the potential of improving the system EE are aspects that have been attracted much attention to the development of RISs. Indeed, with passive beamforming, an RIS is capable of amplifying and forwarding radio signals without needing any power amplifier, motivating an obvious comparison with AF relaying. In this regard, the authors in [38] demonstrated that RIS-aided systems can achieve up to 300% higher EE compared to systems aided by a multi-antenna AF relay. To attain such remarkable result, the authors proposed resource allocation schemes to maximize the EE via power allocation and optimization of the phase shift configuration at the RIS, constrained by individual link budget guarantees for the UEs. The proposed schemes rely on optimization tools such as *alternating optimization*, *gradient descent search*, and *sequential fractional programming*. Lastly, the authors proposed a realistic RIS power consumption model, which have been extensively adopted in the literature to study the EE of RIS-aided systems. Alternatively, the authors in [39] compared the EE performance of RIS with decode-and-forward (DF) relaying, a superior but more complex scheme in comparison to AF relaying. Their main findings concluded that, to compete with a DF relay, an RIS needs hundreds of reconfigurable elements. Still, when high rates are needed, the RIS surpasses the EE obtained by DF relaying.

To achieve practical implementations of RIS-aided systems, it is necessary to develop solutions to several technical challenges. As mentioned previously, the acquisition of channel knowledge to optimize the RIS phase shift configuration is still an open problem that requires efficient solutions. Moreover, the operation of RISs in scenarios where the direct channels between the UEs and the BS are present must be further investigated [38]. Another research direction consists on developing strategies for determining the RIS deployment location. In this regard, efficient strategies must account for aspects such as the spatial density of UEs and the presence of neighboring communication cells [40]. Due to the many variables involved in such a problem, AI techniques are promising candidates to tackled it. In conclusion, efforts are required on the study of the fundamental limits of RIS-aided systems, in addition to the development of robust optimization and resource allocation schemes based on electromagnetic-based realistic models for the behavior of the RISs [18].

5 Summaries of Contributions

This chapter provides the summaries of the contributions that make up this thesis. Specifically, Section 5.1 presents the list of published and submitted scientific papers, accompanied by a diagram relating them to the thesis objectives. Then, Section 5.2 summarizes the findings of the contribution associated to XL-MIMO, while Section 5.3 does the same for the contributions associated to RISs.

5.1 List of Papers

The following published and submitted scientific papers present the contributions of this thesis, and are fully available in Part II:

[A] J. H. Inacio de Souza, J. C. Marinello F., A. Amiri and T. Abrão, "QoS-Aware User Scheduling in Crowded XL-MIMO Systems Under Non-Stationary Multi-State LoS/NLoS Channels," *IEEE Transactions on Vehicular Technology* (IF: 6.8, Qualis: A1-Eng. IV), vol. 72, no. 6, pp. 7639-7652, June 2023, doi: `https://doi.org/10.1109/TVT.2023.3243488`.

[B] J. H. Inacio de Souza, J. C. Marinello F., T. Abrão and C. Panazio, "Energy Efficiency and Throughput of Random Access Protocols for RIS-Aided IoT Networks," *2022 IEEE 8th World Forum on Internet of Things (WF-IoT),* Yokohama, Japan, 26 Oct.-11 Nov. 2022, pp. 1-6, doi: `https://doi.org/10.1109/WF-IoT54382.2022.10152283`.

[C] J. H. Inacio de Souza, J. C. Marinello F., T. Abrão and C. Panazio, "Reconfigurable Intelligent Surfaces to Enable Energy-Efficient IoT Networks," *2022 Symposium on Internet of Things (SIoT),* São Paulo, Brazil, 24-28 Oct. 2022, pp. 1-4, doi: `https://doi.org/10.1109/SIoT56383.2022.10070317`.

[D] J. H. Inacio de Souza, V. Croisfelt, F. Saggese, T. Abrão and P. Popovski, "Randomized Control of Wireless Temporal Coherence via Reconfigurable Intelligent Surface," *2023 IEEE International Conference on Communications Workshops (ICC Workshops),* Rome, Italy, 28 May-01 June 2023, pp. 1535-1540, doi: `https://doi.org/10.1109/ICCWorkshops57953.2023.10283583`.

[E] J. H. Inacio de Souza, V. Croisfelt, R. Kotaba, T. Abrão and P. Popovski, "Uplink Multiplexing of eMBB/URLLC Services Assisted by Reconfigurable Intelligent Surfaces," *submitted to IEEE Communications Letters* (IF: 4.1, Qualis: A1-Eng. IV), Jan. 2024, pp. 1-5. Pre-print available at `https://arxiv.org/abs/2305.04629`.

In addition, the following selected publications, which are *not included* in the thesis, were developed in parallel with the main contributions:

[F] L. M. Taniguchi, J. H. Inacio de Souza, D. W. M. Guerra and T. Abrão, "Resource Efficiency and Pilot Decontamination in XL-MIMO Double-Scattering Correlated Channels," *Transactions on Emerging Telecommunications Technologies* (IF: 3.6, Qualis: A3-Eng. IV), Dec. 2021, vol. 32, no. 12, pp. e4365, doi: `https://doi.org/10.1002/ett.4365`.

[G] H. L. dos Santos, J. H. Inacio de Souza, J. C. Marinello F. and T. Abrão, "LSTM-ACB-Based Random Access for Mixed Traffic IoT Networks," *2022 IEEE 8th World Forum on Internet of Things (WF-IoT),* Yokohama, Japan, 26 Oct.-11 Nov. 2022, pp. 1-6, doi: `https://doi.org/10.1109/WF-IoT54382.2022.10152141`.

[H] W. Souza Jr, J. H. Inacio de Souza, J. C. Marinello and T. Abrão, "Energy Efficiency in RIS-Aided Massive MIMO and XL-MIMO Communication Systems" in *Massive MIMO for Future Wireless Communication Systems: Technology and Applications*. A. L. Imoize and W. Montlouis, Wiley/IEEE Press, 2024, ch. 7. Accepted on Jan. 2024.

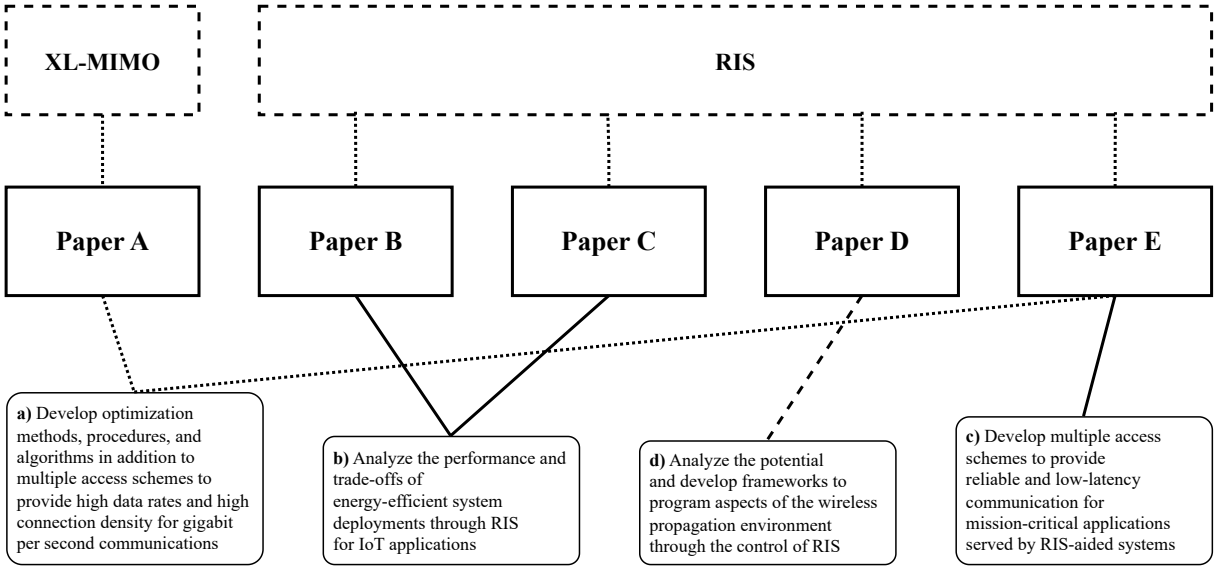
The contributions of the thesis can be divided into two main groups. On the one hand, one group addresses XL-MIMO communications systems, including only paper [A]. On the other hand, the other group focuses on communications systems assisted by RISs, comprising papers [B], [C], [D] and [E]. Guided by these two groups, Fig. 5 presents a diagram relating each paper to the thesis objectives, providing a comprehensive view of the contributions.

5.2 Extra-Large Scale Massive MIMO

Paper A: QoS-Aware User Scheduling in Crowded XL-MIMO Systems Under Non-Stationary Multi-State LoS/NLoS Channels

This paper investigates the problem of scheduling and power allocation in XL-MIMO systems with high user density. Recalling Chapter 3, physically large arrays with many antennas provide high-resolution spatial multiplexing capability and high array gains. These characteristics make XL-MIMO a promising setup to address crowded areas, serving simultaneously large numbers of users with moderate to high data rates. In this scenario, when the number of users is in the same order of the number of antennas at the BS, a user scheduling strategy must be implemented to exploit the spatial DoFs offered by the XL-MIMO channel. Importantly, to leverage the system resources, the scheduling strategy must be guided by an optimization goal and limited by the physical constraints of the

Figure 5 – Diagram of the contributions and their relation to the investigated physical layer technologies and the thesis objectives. The full version of all scientific papers is available in Part II.



Source: The author.

network infrastructure, which includes the available frequency channels, time slots, and power budget. On this basis, this paper focuses on optimizing the system sum-rate, ensuring that the users are provided with data rates lower-bounded by a minimum operating value. Such a goal is achieved by developing a joint user scheduling and power allocation method that balances performance and computational complexity in realistic crowded XL-MIMO environments.

The introduced system model comprises the DL of a single-cell XL-MIMO system. Particularly, the multi-state channel model enables the representation of complex urban environments, where users with LoS and users with non-line-of-sight (NLoS) receive antenna may coexist in the communication cell. This aspect is controlled by a parameter that models the probability of the user antenna experiencing a LoS channel. Additionally, the channel model incorporates the spherical wavefront model, accounting for the effects of near-field propagation observed in the XL-MIMO wireless channel. Before data transmission, the BS schedules the users based on CSI, described by their respective channel vectors. For signal transmission, the BS employs the ZF precoder and distributes its transmit power budget among the data streams intended for each scheduled user. The allocated powers are determined such that the rate of all users is lower-bounded by a minimum operating value. The joint user scheduling and power allocation problem is defined to maximize the system sum-rate, subject to the BS transmit power budget and the minimum operating rate. The original problem formulation is intricate, since the optimization variables, *i.e.*, the *set of scheduled users* and the *allocated powers*, are highly coupled. Thus, the decision of scheduling any user affects the powers assigned to all the other ones and, consequently,

may prevent fulfillment with the problem constraints. Besides, the formulated problem can be reduced to the *knapsack* problem, classified as nondeterministic polynomial-time complete (NP-complete), which makes it computationally expensive to find an optimal solution to it [41]. Given the challenging problem formulation, the elected approach is to decouple it into two sub-problems, one related to scheduling and the other to power allocation, and propose a low-complexity method for solving the entangled user scheduling.

Describing the elected approach, the devised scheduling sub-problem aims to maximize the number of scheduled users, considering that each user data stream is assigned with the power to yield the minimum operating rate. Then, the power allocation sub-problem aims to maximize the system sum-rate, assuming the set of users scheduled in the previous sub-problem and guaranteeing that their rates are lower-bounded by the minimum operating value. Importantly, both sub-problems are constrained by the BS transmit power budget, ensuring that the QoS requirements of the scheduled users are met, given the BS capacity. Analyzing the scheduling sub-problem, it is still a decision problem and therefore NP-complete. For this reason, a low-complexity *graph search* algorithm is proposed to approximate a suboptimal solution to it. The algorithm is based on the system representation by an undirected weighted graph (UWG), where each user is described by a *node*. Each node receives a *weight* equal to the required transmit power for the single-user capacity to reach the minimum operating rate. In addition, *edges* are drawn between pairs of nodes according to an orthogonality metric involving the associated channel vectors. Accordingly, the set of scheduled users is found by running a greedy clique search over the graph, identifying the users that have mutually favorable channel properties regarding orthogonality, while requiring a total transmit power that meets the BS limitation. Given the set of scheduled users, the power allocation sub-problem is solved optimally using the *water-filling* algorithm [42]. Exceptionally, in case the obtained scheduling solution results in an unfeasible power allocation sub-problem, a simple user removal strategy is proposed to obtain a feasible one. In summary, the proposed method comprises a user scheduling algorithm based on greedy clique search, followed by a user removal stage and then optimal power allocation.

The proposed method is numerically compared with the state of the art regarding sum-rate, number of scheduled users, and computational complexity. Specifically, the methods from [43, 44] are adopted for comparison, in addition to a benchmark based exclusively on the users channel powers. Aiming to cover several realistic scenarios, different LoS/NLoS setups are inspected, assuming the cases with perfect and imperfect CSI at the BS. In short, the proposed method consistently presents superior performance in comparison with the benchmarks. Particularly, it reaches remarkable number of scheduled users and sum-rate, scheduling up to 5 times more users than a method from [44]. Also, it achieves a good performance-complexity trade-off, having a running time up to 3.7 and 79 times lower than the methods from [44] and [43], respectively. One of the key improvements observed in

the proposed method is the enhancement of the cell coverage. While the methods from [44] covers only half of the cell radius, the proposed method can schedule users in the whole area, serving the users with unfavorable channel condition located at the cell edge. To conclude, the proposed method has steady performance even in the presence of imperfect CSI, working reliably with channel vectors subject to bounded estimation errors.

This paper proposes a QoS-aware joint user scheduling and power allocation method to address crowded XL-MIMO scenarios. Specially, using the proposed method, the spatial DoFs provided by the XL-MIMO channel can be exploited to deliver high-rate communications for areas presenting elevated user density. Attractively, the method yields a competitive performance-complexity trade-off, outperforming the state of the art in terms of number of scheduled users and cell coverage. This unveils the potential of XL-MIMO to operate in densely-populated urban areas or in venues used to host large-scale cultural, sporting, and business events.

5.3 Reconfigurable Intelligent Surfaces

Paper B: Energy Efficiency and Throughput of Random Access Protocols for RIS-Aided IoT Networks

This paper investigates the design of energy-efficient wireless networks for IoT devices by deploying RISs to enhance the radio links from the devices to the BS. As discussed in Chapter 4, since its operation essentially relies on passive reflecting elements, an RIS has the capability of extending the network coverage consuming low power compared to the equipment at the cell site. For this reason, this technology has been seen as a building block to achieve ubiquitous and energy-efficient IoT networks, improving the channel of devices located in shadowed areas at the cost of a marginal increase in the power consumed by the system.

To study the EE of an RIS-aided IoT network, this paper concentrates on the UL, assuming the transmissions follow the two time-scale RA protocol proposed in [45]. Such a protocol tries to select appropriate RIS configurations to assist the transmissions of the contending devices based on a DL training phase, followed by a UL access phase. In the DL phase, the devices measure the channel quality yielded by different RIS configurations from a training signal transmitted by the BS. Then, from the measured qualities, the devices decide the set of time slots to transmit according to the access policy adopted by the network. In short, the *access policy* defines the number of packet replicas sent by each device and how the time slots to transmit them are selected based on the measured channel qualities. Lastly, in the UL phase, the devices send packet replicas within the selected time slots while the BS decodes them, applying SIC to resolve potential packet collisions.

The main contribution of this paper is the derivation of expressions for the throughput and EE of the described RIS-aided IoT network. Specifically, a realistic power consumption model for the system is developed, accounting for the RF power spent in the protocol phases, in addition to the static power spent by hardware at the RIS, devices, and BS. Moreover, an expression for the system throughput is derived considering the structure of the RA protocol. In summary, the throughput consists in the ratio between the number of devices that had their packets successfully decoded at the BS and the duration of the RA frame, which includes the duration of both DL and UL phases. Subsequently, the expression for the overall system EE is derived from the throughput expression and the power consumption model. Such an expression helps in developing a comparison of the EE achieved by the network when it adopts different access policies. Meanwhile, it permits examining the impact on EE caused by the key system parameters, namely, the *device transmit power*, the *number of elements* at the RIS, the *number of contending devices*, and the *number of time slots* in the RA frame.

The numerical results demonstrate the trends of the throughput and EE according to the mentioned parameters. They are obtained in the case of four access policies, two proposed in [45], and the other two adapted from the RA protocols in [46] and [47], which originally cover the scenario of communication without RIS. Importantly, the two latter access policies do not rely on the devices measuring the channel qualities, so the DL phase is absent in them. Given this, half of the access policies exploit information on the channel qualities, while the other half do not. As a first conclusion, the results reveal that the contribution of the RIS for the system power consumption is negligible, in a way that large numbers of reflecting elements cause marginal EE loss. Moreover, the results show that the access policies that exploit information of the channel qualities attain competitive throughput and EE, simultaneously. Remarkably, one of these access policies achieves its EE peak at a low value of device transmit power. Therefore, this access policy contributes significantly to reducing the power consumption at the devices-side, a key aspect in IoT networks, since, in general, the devices are powered by limited power sources such as low-capacity batteries.

In summary, the discussion on the results evidences the gains attained by RISs to deploy ubiquitous and energy-efficient IoT networks, as they can increase the system throughput with marginal impact on the EE. Notably, these gains are multiplied with the adoption of the adequate access policy. Hence, from the EE perspective, the access policies that exploit information of the channel qualities are promising, potentially reducing the power consumption at the devices-side.

Paper C: Reconfigurable Intelligent Surfaces to Enable Energy-Efficient IoT Networks

This paper also addresses the design of energy-efficient networks for IoT devices through the use of RISs. However, in this case, the EE study focuses exclusively on the device-side and on physical layer aspects, embracing the transmission protocol proposed in [45] and the adaptions introduced in paper [B]. This investigation is motivated by the outcome of paper [B] showing that the RIS has potential of reducing the device transmit power. Closely to what was discussed in Chapter 2, reducing the power consumed by the communication functions in IoT devices is crucial to implement ubiquitous networks with low maintenance cost and low carbon footprint. On this basis, the objective of this paper is to investigate how the IoT device transmit power and, consequently, its EBL, behave according to RIS parameters controlled by the system designer, in particular the *number of reflecting elements* and the *size of the RIS configurations codebook*.

Although it is based on the system models in [45] and paper [B], the considered UL scenario presents simplifications to keep focus only on the physical layer aspects. One of the main simplifications is the assumption of collision-free packet transmission, so that single-device transmission takes place. This covers the case where multiple contending devices transmit over orthogonal time slots. Moreover, it is assumed that the RIS is configured based on a predefined codebook with a limited number of configurations, where each configuration sets the same reflecting angle to all RIS elements. Each configuration is identified by a phase shift taken from equally-spaced values in the interval $[0, \pi/2]$. The UL transmission follows an RA frame comprised by the sequential *training*, *access*, and *acknowledgement* phases, described in the following from the device perspective. In the training phase, the device receives signals from the BS over several time slots to measure the channel quality provided by each RIS configuration in the codebook. Based on these measurements, the device identifies the time slot that faced the best channel quality or, equivalently, the time slot subject to the RIS configuration that maximized the SNR at the BS. Then, in the access phase, the device transmits its packet in the time slot identified previously. Importantly, to achieve successful packet decoding at the BS, the device uses a fixed rate code, adapting its transmit power according to the target SNR of the data detection scheme at the BS, along with the channel condition. Lastly, the device receives an acknowledgement signal from the BS reporting the successful or failed packet decoding. Notice that the SNR improvement leveraged by the RIS allows decreasing the device transmit power without compromising the packet decoding at the BS. On this basis, an expression is derived to compute the transmit power necessary to reach the target SNR at the BS. Obtaining such a value is essential to study the device EE, since, as seen in paper [B], the RF power represents a significant fraction of the power consumed by the device for communication.

The contribution of this paper is the derivation of an expression for the device EBL,

which depends on the device *battery capacity*, its *energy consumption*, and the time span between consecutive *reporting periods*, namely, the device activation events. In particular, the energy consumption is strongly linked to the transmit power necessary to reach the target SNR at the BS. Therefore, the obtained expression for the EBL embodies the variation of the channel condition across the service area and the impact of the RIS parameters, comprising the number of elements and the size of the configurations codebook. To study how the EBL behaves with these aspects, simulation results are provided, showing the distribution of the EBL in the service area assuming several RIS setups. In short, the results demonstrate that improvements on the EBL and cell coverage are achieved with a larger RIS and a larger codebook. Detailing, increasing the number of RIS elements enhances the *array factor*, improving the channel condition and, consequently, demanding a lower transmit power at the device. Furthermore, as the codebook size increases, the RIS can produce strong beams in more directions, which enables providing improved channel conditions to a larger portion of the service area. Thus, since the device has a constrained transmit power, the channel improvement provided by the RIS can help the device, when located in shadowed areas, to reach the target SNR at the BS, increasing the network coverage. It is worth noting that increasing the codebook size implies that the channel quality produced by more RIS configurations must be measured by the device during the training phase. As a conclusion, the codebook size impacts the overhead generated by the RA protocol, which is not observed in the number of RIS elements. Hence, the two RIS parameters yield different system trade-offs in terms of EBL and network overhead.

In conclusion, this paper studies the EE of an RIS-aided IoT network by deriving the device EBL. Specially, the paper focuses on the EE problem at the devices-side, which deserves more attention as IoT devices are powered by limited-capacity batteries with potential expensive and difficult replacement process. The EBL behavior is analyzed from numerical results generated by considering different codebook sizes and number of RIS elements. Such an analysis reveals that the RIS contributes to increase significantly the EBL through a reduction on the device transmit power. At the same time, the RIS extends the network coverage by improving the channel condition across the service area. However, to attain the gains on EBL and coverage while keeping a low network overhead, it is crucial to carefully optimize the number of reflecting elements and the size of the configurations codebook at the RIS.

Paper D: Randomized Control of Wireless Temporal Coherence via Reconfigurable Intelligent Surface

This paper focuses on the potential of RIS to shape the temporal statistics of the wireless propagation channel. Predominantly, the literature on RIS concentrates on system performance enhancement through algorithmic optimization of the RIS configuration from,

e.g., CSI, prior environment information, or predefined configuration codebooks. Indeed, relatively few works explore the potential of stochastic, time-varying, RIS configurations to enable efficient deployments, that don't rely on costly tasks arising from the procedure of configuring the RIS, such as network coordination and full CSI acquisition. To help to fill this research gap, this paper aims to characterize the temporal correlation of the wireless channel in a system aided by an RIS, which phase shift configurations are sampled from a random distribution. Such an analysis allows understanding how the RIS parameters, specifically its *number of elements* and the *random distribution* that generates the configurations, affect the behavior of slow- and fast-fading channels. Subsequently, a framework is proposed to arbitrarily control the channel correlation and, consequently, its coherence time, by properly selecting the mentioned RIS parameters.

Describing the studied scenario, it comprises a point-to-point narrowband communication system with one RIS, one transmitter, and one receiver, both equipped with a single antenna. Consistent with Chapter 4, the scenario presents two distinct radio paths connecting the transmitter and receiver, the direct path and the reflected path, particularly controlled by the RIS. Moreover, the channel response is modeled by a discrete-time-varying model based on the Rician-fading, where each channel coefficient has a deterministic LoS component in addition to a time-variant NLoS component. Specifically, such an NLoS component is modeled as a first-order autoregressive (AR(1)) random process, accounting for the natural temporal statistics of the channel. The RIS configuration is described by the phase shifts applied by each reflecting element that comprises the surface. Importantly, it is assumed that these phase shifts change at every time index, while the values are obtained by sampling a random distribution defined by the system designer. As the channel and the RIS configuration change with the time index, the effective channel experienced by the transmitter-receiver pair also changes. Hence, the objective is to assess the impact of the physical channel and the RIS parameters on the channel temporal statistics.

As an initial analysis, the paper derives the autocorrelation function (ACF) of the effective channel. Notably, the obtained ACF depends not only on channel parameters, but also on RIS parameters, which can be controlled by the system designer. Specifically, the channel parameters are the AR(1) parameters and the powers of the LoS and NLoS components. Meanwhile, the RIS parameters are the number of reflecting elements and the random distribution that generates the RIS configurations. Even though the general ACF expression unveils the dependence of the channel temporal statistics on these parameters, the intertwined relationship among them restricts further analysis. Hence, to get a tractable expression, some simplifying assumptions are taken. Importantly, the RIS phase shifts are considered to be obtained, at each time index, by sampling a *uniform* random distribution over a parameterized interval. As a result, the simplified ACF presents explicitly dependence on the number of RIS elements and the introduced uniform distribution parameter. Then, from the simplified ACF, a framework is developed to design the RIS parameters according

to the desired shape for the channel ACF. Therefore, using the framework, slow-fading channels can be converted into fast-fading by tailoring the uniform distribution parameter, whereas the opposite can be achieved by controlling the number of RIS elements. This is possible because the channel ACF is the inverse Fourier transform of the *Doppler spectrum*, from which the channel coherence time can be evaluated [48].

Summarizing, this paper analyzes the temporal statistics of the wireless channel subject to an RIS set up with stochastic configurations that change at each time index. Subsequently, exploiting the dependence of the channel statistics on the RIS parameters, a framework to shape the channel ACF by tailoring such parameters is proposed. Therefore, the RIS can exert control over the channel temporal statistics, converting slow-fading channels into fast-fading and vice versa.

Paper E: Uplink Multiplexing of eMBB/URLLC Services Assisted by Reconfigurable Intelligent Surfaces

This paper investigates the coexistence of eMBB and URLLC traffic in the UL of RIS-assisted communications systems. The 3rd Generation Partnership Project (3GPP) 5G new radio (5G NR) specifications address such an issue with the *preemptive puncturing* scheme, which, based on BS coordination, interrupts temporarily the eMBB traffic in favor of the URLLC transmissions. However, when it comes to the UL of RIS-assisted systems, this scheme imposes critical algorithmic restrictions for the RIS operation, in addition to an increased overhead that potentially violates the URLLC latency constraint. Illustrating, due to the unpredictability and low-latency constraint of the URLLC traffic, when a URLLC packet arrives, the BS has no feasible time to estimate the device CSI for control purposes. This prohibits the use of algorithms that optimize the RIS configuration based on URLLC CSI to assist the URLLC traffic. Meanwhile, the task of controlling the RIS from the BS adds extra overhead caused by the exchange of control messages. Hence, there is a need for new strategies to address the UL coexistence of eMBB and URLLC services, with major focus in RIS-assisted systems. To contribute to this matter, this paper proposes a UL traffic multiplexing scheme deployed by the RIS, which attractively doesn't require BS coordination and relies only on eMBB CSI.

The system model comprises one single-antenna BS, one hybrid RIS, and two single-antenna UEs, where one UE uses the eMBB service while the other uses the URLLC one. Concentrating on the UL, the communication frame is organized into the sequential *CSI estimation*, *computation*, and *payload transmission* phases. In short, in the first phase, the BS estimates the eMBB CSI. Then, in the computation phase, the BS optimizes the RIS configurations based on the obtained CSI, sending the configurations to be stored at the RIS. Finally, in the last phase, the UEs transmit their payload data. Importantly, the eMBB traffic is scheduled before the frame start, while the URLLC bursty traffic is

scheduled at any time during the payload transmission phase. The proposed scheme uses the RIS to multiplex the services through two different configurations, the *eMBB-oriented* configuration to multiplex the eMBB traffic, and the *URLLC-oriented* configuration for the URLLC one. During the payload transmission phase, the RIS starts configured with the eMBB-oriented configuration, enabling the reception of the scheduled eMBB traffic. Eventually, if the URLLC traffic start is detected by the RIS, the URLLC-oriented configuration is set, silencing the eMBB channel to suppress interference. In particular, the detection of the URLLC traffic start is done locally at the RIS, based on a dedicated preamble transmitted by the UE and simple processing of the signal received by the RIS antenna. Given the proposed multiplexing scheme, the challenge addressed by this paper is the design of the two RIS configurations relying only on eMBB CSI, obligatorily meeting the eMBB high SE requirement and the URLLC reliability and low-latency requirements.

As the main contribution, this paper proposes methods to compute the eMBB- and URLLC-oriented configurations based solely on eMBB CSI. Specifically, the eMBB-oriented configuration is designed aiming to maximize the UE SNR so as to provide the high SE demanded by the eMBB service. Thus, since full eMBB CSI is available at the BS, the configuration that satisfies this optimization goal is the well-known *coherent passive beamformer*, computed by a simple closed-form expression [35]. Differently, exploiting the available eMBB CSI to compensate for the absence of the URLLC one, the URLLC-oriented configuration is designed to silence the eMBB channel, mitigating the eMBB interference on the URLLC traffic. On this basis, two different algorithms are proposed to compute such a configuration. One of them is the *phasors rotation*, a heuristic algorithm that tries to cancel out the eMBB channel gain by compensating the phase shifts of the RIS elements via subtraction. The other is the *interference nulling*, an iterative alternating projection algorithm that presents high probability of convergence. Specifically, it approximates a solution to the optimization problem of finding a RIS configuration that nullifies the eMBB channel. These two configurations are capable of suppressing the interference caused by the eMBB UE without BS coordination, providing the reliable and low-latency communication required by the URLLC service. With the establishment of proper designs for the eMBB- and URLLC-oriented configurations, the description of the proposed RIS-aided traffic multiplexing scheme is complete. Subsequently, a performance and latency analyses are carried out to evaluate the proposed scheme. Focusing on the particular requirements of each service, the performance analysis generates expressions for the URLLC outage probability and the eMBB SE. Then, the latency analysis measures the delay introduced by the scheme to the URLLC latency, accounting for the time necessary to detect the traffic at the RIS, switch over the RIS configurations, and transmit the URLLC data symbols.

A set of numerical results is deployed to verify the performance of the proposed multiplexing scheme and compare it to the state of the art. In particular, the benchmarks

include the preemptive puncturing scheme and the RIS configurations proposed to assist the URLLC and eMBB services in [49]. Meanwhile, the numerical analysis evaluates the outage probability for the URLLC service and the SE for the eMBB one, as functions of system parameters such as the *transmit power*, *number of RIS elements*, and the *URLLC traffic miss detection rate*. The results reveal that the proposed scheme achieves an outage probability comparable to preemptive puncturing, introducing a lower latency due to its independence from BS coordination. In this regard, remarkably, the URLLC-oriented RIS configuration computed by the interference nulling algorithm outperforms the outage probability of preemptive puncturing by up to 4.9 times. Also, when compared to a configuration introduced in [49], the proposed interference nulling decreases the outage probability by up to 3 orders of magnitude. In respect of the eMBB service, the proposed multiplexing scheme provides high SE even for a UE located at the cell edge, proving that this essential service requirement is satisfied over the whole cell area. Lastly, the results show that the scheme for URLLC traffic detection at the RIS is not considered to be a major limiting factor for the performance of the multiplexing scheme. This is explained by the existence of schemes to reach the needed miss detection rate that rely on efficient implementations in terms of computational burden and use of the radio resources.

In summary, this paper proposes a service multiplexing scheme to support the coexistence of URLLC and eMBB services in the UL of RIS-aided systems. Attractively, to multiplex both services, the scheme relies only on eMBB CSI and introduces lower latency than the 5G NR preemptive puncturing. All this is achieved without BS coordination and by light signal processing at the RIS, comprising an efficient implementation. Hence, the proposed scheme conciliates the traffic multiplexing strategy defined by the current 3GPP specifications with the operational constraints of the RISs, a technology still under development, revealing paths to exploit their benefits for the next-generation cellular networks.

6 Conclusions and Remarks

This chapter concludes the Part I of the thesis, summarizing the key results obtained by each contribution and providing directions for motivating future work.

6.1 Conclusions

The conclusions of the thesis contributions are listed in the following.

- Addressing the problem of user scheduling and power allocation in XL-MIMO systems with high user density, paper [A] has proposed an efficient method to maximize the sum-rate and enhance the number of scheduled users. Satisfying individual user QoS constraints, the proposed method combines a graph search algorithm and optimal power allocation, attaining competitive sum-rate even when accounting for uncertainties in the CSI available at the BS. Presenting a good performance-complexity trade-off, the method can schedule up to 5 times more users than the benchmarks, while having a running time up to 79 times lower. Finally, the proposed method provides fair coverage as it schedules users over the entire service area, improving the communication performance of users located in the cell edge.

- To develop ubiquitous IoT networks for low-power machine-type devices, paper [B] have studied the EE of an RIS-aided RA protocol. The main contributions of the paper are a general power consumption model and a throughput formula that can be applied to evaluate the performance of different access policies. Using them to verify the system EE, it has been demonstrated that RISs are promising components to deploy networks for low-power devices, providing high throughput gains with marginal impact on EE, even for surfaces with large numbers of reflecting elements. Furthermore, specifically on the RA protocol, access policies that exploit the channel qualities achieved superior EE performance than the ones that don't. Finally, the results revealed that, to improve EE, the studied scheme permits that the devices transmit at extremely-low power levels, while the RIS with many elements compensates for the throughput reduction.

- With emphasis on battery-powered IoT devices, paper [C] have studied the EE at the devices-side in an RIS-aided communication system. Considering the same RA protocol and a scenario similar to paper [B], this paper focused exclusively on the study of physical layer aspects. Specifically, the paper introduced an expression for the device EBL which, among other variables, depends on the number of RIS elements and the RIS configuration. Numerical evaluations have shown that, in

addition to coverage extension, an RIS can provide substantial gains on the EBL by reducing the required device transmit power. Importantly, the results also evidenced that, to fully exploit the potentials of the RIS, the number of RIS elements and the size of the configurations codebook should be jointly optimized.

- Paper [D] have explored the capability of an RIS to exert control over the channel temporal statistics. Specifically, it proposed a randomized framework to control the channel coherence time by properly setting the number of RIS elements and designing the random distribution that generates the time-varying reflecting coefficients. Such a framework was developed based on the introduced discrete-time-varying channel model where the NLoS channel components are modelled as stationary AR(1) random processes. Then, the derivation of the channel ACF revealed the dependence of the channel temporal statistics on the parameters of the RIS, controlled by the system designer. Numerical results corroborated the effectiveness of the proposed analytical framework, enabling, via optimization of the RIS parameters, the conversion of slow-fading channels into fast-fading and vice versa.

- In paper [E], an RIS-aided service multiplexing scheme to support the coexistence of URLLC and eMBB UL traffic has been proposed. Satisfying the URLLC latency requirements, the proposed scheme relies only on CSI of the eMBB UE scheduled in advance. Inspired by the preemptive puncturing scheme, but without BS coordination, the multiplexing scheme works by using the RIS for temporarily silencing the eMBB traffic during the URLLC transmission, suppressing any interference to provide low-latency and reliable communication. Numerical results demonstrated that the proposed scheme achieves outage probability up to 4.9 times lower than preemptive puncturing. Furthermore, the multiplexing scheme obtains significant gains on the average URLLC latency, reaching 0.30 ms against the 0.17 ms achieved by preemptive puncturing.

6.2 Future Work Directions

The contributions discussed in Sections 5.2 and 5.3 have provided solutions to relevant technical problems involving XL-MIMO and RIS-aided communications systems. In particular, such solutions with their respective benefits and limitations, in addition to the results obtained via numerical and analytical analyzes, introduced novel research questions that need to be addressed. These open questions motivated the future work directions summarized in the following.

- In paper [A], providing connectivity for the nonscheduled users remains as a relevant open problem. In this regard, it is necessary to investigate the impact of allocating such users in subsequent frames, analyzing performance in terms of latency and

long-term average user rate. Furthermore, the integration of the XL-MIMO systems with other network technologies, such as non-terrestrial networks, may be a promising direction to address the connectivity issue related to the nonscheduled users.

- Regarding papers [B] and [C], the communication overhead and its dependence on the size of the RIS configurations codebook must be further investigated. Moreover, the problem of codebook design must be considered assuming complex scenario setups, including the case with several RISs and with existence of LoS links between the UEs and the BS. Finally, one potential direction to improve the RA protocol performance is to develop access policies based on AI techniques, such as reinforcement learning.

- Based on the contribution of paper [D], it is necessary to investigate applications of the proposed framework to improve the end-user performance, enhancing the availability and reliability of the communication service. Furthermore, there is a need to develop schemes to obtain practical realizations of the suggested flexible block-fading model and identify potential applications to it.

- To verify the feasibility of the proposed multiplexing scheme of paper [E] in realistic network environments, the scenario with a multiple-antenna BS, multicarrier modulation, and considering the coexistence of several UEs of each service type must be investigated. Besides, the multiplexing scheme must be evaluated accounting for the resource slicing strategy and considering more challenging practical scenarios, such as the scenario where part of the UEs necessarily need the RIS assistance to communicate, while the remaining UEs don't.

6.3 References

[1] "Ericsson mobility report," Ericsson, June 2023.

[2] "Next G Alliance report: Evolution of sustainability indicators for next-generation radio network technologies," Next G Alliance, Alliance for Telecommunications Industry Solutions, Dec. 2023.

[3] "Transforming our world: the 2030 agenda for sustainable development," *Resolution 70/1 of the UN General Assembly*, Oct. 2015.

[4] "Framework and overall objectives of the future development of IMT for 2030 and beyond," *Recommendation ITU-R M.2160-0*, Nov. 2023.

[5] "Going green: benchmarking the energy efficiency of mobile networks (second edition)," GSMA Intelligence, Feb. 2023.

[6] "Pesquisa nacional por amostra de domicílios contínua: Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal 2022," IBGE, Coordenação de Pesquisas por Amostra de Domicílios, Nov. 2023.

[7] "Closing the usage gap in Brazil: Key barriers to mobile internet adoption and use," GSMA Intelligence, Feb. 2023.

[8] "Hexa-X architecture for B5G/6G networks – final release," *Deliverable Hexa-X D1.4*, July 2023.

[9] "Next G Alliance report: Roadmap to 6G," Next G Alliance, Alliance for Telecommunications Industry Solutions, Feb. 2022.

[10] "Framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation ITU-R M.2083-0*, Sept. 2023.

[11] "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," *Recommendation ITU-R M.2410-0*, Nov. 2017.

[12] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, July 2021.

[13] Y. Han, S. Jin, M. Matthaiou, T. Q. S. Quek, and C.-K. Wen, "Toward extra large-scale MIMO: New channel properties and low-cost designs," *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14 569–14 594, Aug. 2023.

[14] V. Croisfelt, A. Amiri, T. Abrao, E. de Carvalho, and P. Popovski, "Accelerated randomized methods for receiver design in extra-large scale MIMO arrays," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6788–6799, July 2021.

[15] A. Amiri, C. N. Manchón, and E. de Carvalho, "Uncoordinated and decentralized processing in extra-large MIMO arrays," *IEEE Wireless Communications Letters*, vol. 11, no. 1, pp. 81–85, Jan. 2022.

[16] A. Amiri, S. Rezaie, C. N. Manchón, and E. de Carvalho, "Distributed receiver processing for extra-large MIMO arrays: A message passing approach," *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2654–2667, Apr 2022.

[17] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[18] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent

surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[19] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[20] A. Ali, E. D. Carvalho, and R. W. Heath, "Linear receivers in non-stationary massive MIMO channels with visibility regions," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 885–888, June 2019.

[21] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[22] P. Ramezani and E. Björnson, *Near-Field Beamforming and Multiplexing Using Extremely Large Aperture Arrays.* Cham: Springer International Publishing, 2024, pp. 317–349.

[23] A. Amiri, M. Angjelichinoski, E. de Carvalho, and R. W. Heath, "Extremely large aperture massive MIMO: Low complexity receiver architectures," in *2018 IEEE Globecom Workshops (GC Wkshps)*, 9-13 Dec. 2018, pp. 1–6.

[24] J. C. Marinello, G. Brante, R. D. Souza, and T. Abrão, "Exploring the non-overlapping visibility regions in XL-MIMO random access and scheduling," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6597–6610, Aug. 2022.

[25] J. Rodríguez Sánchez, F. Rusek, O. Edfors, M. Sarajlić, and L. Liu, "Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 687–700, Jan. 2020.

[26] Z. Wang, J. Zhang, H. Du, D. Niyato, S. Cui, B. Ai, M. Debbah, K. B. Letaief, and H. V. Poor, "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Communications Surveys & Tutorials*, pp. 1–46, Jan. 2024, early access.

[27] J. P. González-Coma, F. J. López-Martínez, and L. Castedo, "Low-complexity distance-based scheduling for multi-user XL-MIMO systems," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2407–2411, Nov. 2021.

[28] J. C. Marinello, T. Abrão, A. Amiri, E. de Carvalho, and P. Popovski, "Antenna selection for improving energy efficiency in XL-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 305–13 318, Nov. 2020.

[29] J. H. Inacio de Souza, A. Amiri, T. Abrão, E. de Carvalho, and P. Popovski, "Quasi-distributed antenna selection for spectral efficiency maximization in subarray switching XL-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6713–6725, July 2021.

[30] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[31] A. Albanese, F. Devoti, V. Sciancalepore, M. Di Renzo, and X. Costa-Pérez, "MARISA: A self-configuring metasurfaces absorption and reflection solution towards 6G," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, 2-5 May 2022, pp. 250–259.

[32] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3083–3098, May 2022.

[33] L. Zhang, X. Q. Chen, Q. Liu, Shuoand Zhang, J. Zhao, J. Y. Dai, G. D. Bai, X. Wan, Q. Cheng, G. Castaldi, V. Galdi, and T. J. Cui, "Space-time-coding digital metasurfaces," *Nature Communications*, vol. 9, no. 4334, Oct. 2018.

[34] T. Jiang and W. Yu, "Interference nulling using reconfigurable intelligent surface," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1392–1406, May 2022.

[35] E. Björnson *et al.*, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 135–158, Mar. 2022.

[36] K.-L. Besser and E. A. Jorswieck, "Reconfigurable intelligent surface phase hopping for ultra-reliable communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9082–9095, Nov. 2022.

[37] X. Cao, B. Yang, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Han, H. V. Poor, and L. Hanzo, "Massive access of static and mobile users via reconfigurable intelligent surfaces: Protocol design and performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1253–1269, Apr. 2022.

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

[39] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, Feb. 2020.

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

[41] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. Cambridge, MA, USA: The MIT Press, 2009.

[42] D. Palomar and J. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 686–695, Feb. 2005.

[43] T. Yoo and A. Goldsmith, "Sum-rate optimal multi-antenna downlink beamforming strategy based on clique search," in *GLOBECOM '05. IEEE Global Telecommunications Conference, 2005.*, vol. 3, 28 Nov.-2 Dec. 2005, pp. 1510–1514.

[44] J. P. González-Coma, F. J. López-Martínez, and L. Castedo, "Low-complexity distance-based scheduling for multi-user XL-MIMO systems," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2407–2411, Aug. 2021.

[45] V. Croisfelt, F. Saggese, I. Leyva-Mayorga, R. Kotaba, G. Gradoni, and P. Popovski, "A random access protocol for RIS-aided wireless communications," Mar. 2022, arXiv:2203.03377.

[46] E. Casini, R. De Gaudenzi, and O. Del Rio Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1408–1419, Apr. 2007.

[47] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 477–487, Feb. 2011.

[48] A. Goldsmith, *Wireless Communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2005.

[49] V. D. P. Souto *et al.*, "IRS-aided physical layer network slicing for URLLC and eMBB," *IEEE Access*, vol. 9, pp. 163 086–163 098, Dec. 2021.

Part II

Papers

7 Paper A: QoS-Aware User Scheduling in Crowded XL-MIMO Systems Under Non-Stationary Multi-State LoS/NLoS Channels

João Henrique Inacio de Souza, José Carlos Marinello Filho, Abolfazl Amiri, and Taufik Abrão

This paper has been published in the *IEEE Transactions on Vehicular Technology* (IF: 6.8, Qualis: A1-Eng. IV), vol. 72, no. 6, pp. 7639-7652, June 2023, doi: `https://doi.org/10.1109/TVT.2023.3243488`.

Abstract

Ensuring that the quality-of-service (QoS) requirements are satisfied in wireless communications systems with high user density is challenging due to the limitations on the transmit power budget and the number of resource blocks. In this paper, we propose a QoS-aware joint user scheduling and power allocation technique to enhance the number of served users in the downlink of crowded extra-large scale massive multiple-input multiple-output (XL-MIMO) with minimum QoS requirements guarantee. The proposed technique is constituted by two sequential procedures: the clique search-based scheduling (CBS) algorithm for user scheduling followed by optimal power allocation with transmit power budget and minimum achievable rate per user constraints.We propose a generalized non-stationary multi-state channel model based on spherical wave propagation assuming that users under line-of-sight (LoS) and non-LoS (NLoS) transmission coexist in the same communication cell. This is done to accurately evaluate the proposed technique in realistic XL-MIMO scenarios. Numerical results reveal that the proposed CBS algorithm provides fair coverage over the whole cell area, achieving remarkable numbers of scheduled users with satisfied QoS requirements when users under the LoS and NLoS channel states coexist in the communication cell.

Index Terms: XL-MIMO, user scheduling, resource allocation, channel non-stationarities.

7.1 Introduction

Extra-large scale massive multiple-input multiple-output (XL-MIMO) are the deployments of massive MIMO base stations (BSs) made of arrays of antennas with extreme physical dimensions, often with the size of thousands of wavelengths [1]. Such deployments are promising designs to address crowded communication scenarios, integrating the antenna elements with architectural structures of the environment, *e.g.*, the ceiling, walls, and columns of a stadium, warehouse, or shopping mall [2]. For this reason, the distances between the users and the antenna elements are small compared with the co-located BS design, typically adopted in cellular systems.

Such small distances between the users in conjunction with the very large extent of the antenna array create spatial non-stationarities on the wireless channel, which drastically changes the signal propagation aspects compared with the conventional massive MIMO scenarios. We investigate two of these aspects. First, each antenna element experiences different average received power and phase from each user, suggesting the operation under the near-field propagation regime. Hence, the wireless channel is well-modeled by the *spherical wavefront* (SW) propagation model rather than the conventional plane wavefront propagation model [3]. Second, the closeness between the users and the antenna elements results in a predominantly line-of-sight (LoS) situation. However, due to the relief and presence of scatterers and obstacles in the environment, it is not accurate to assume that all the radio links experience LoS transmission. Hence, as is discussed in [4], it is reasonable to assume that part of the radio links are under the LoS regime, while the remaining links are under the non-line-of-sight (NLoS) one.

Under the near-field propagation regime considering the SW model, the array gain is limited [3]. This suggests that asymptotic favorable propagation, which results in the orthogonality between the channel vectors of different users, may also be compromised by the near-field propagation condition. However, further investigation is needed to support this claim. The looseness of the favorable propagation with the SW model increases the necessity of scheduling spatially compatible users to achieve reasonable DL performance while optimizing the expenditure of scarce radio resources, *e.g.*, transmit power and resource blocks. In wireless systems with limited resources, resource allocation (RA) techniques are essential to assure that the target quality-of-service (QoS) levels are met for a higher number of served users. RA techniques for multi-user MIMO systems are surveyed in [5]. Despite that, further investigation is needed to assess the effectiveness of the conventional resource allocation strategies in crowded multi-user MIMO systems with physically large antenna arrays.

In this sense, in [6, 7] the authors propose methods for antenna selection aiming to maximize the energy efficiency and the spectral efficiency, respectively. The proposed methods use different approaches to compute the set of active antennas, including meta-heuristic optimization, greedy strategies, and heuristic procedures deploying approximate

expressions for the performance metrics. However, proposing solutions for different wireless propagation conditions while optimizing the expenditure of scarce radio resources, is still a need.

Generally, accurate *channel state information* (CSI) is required to perform efficient RA, precoding, and detection. Due to the near-field propagation and spatial non-stationarities of the XL-MIMO channel, new methods for CSI acquisition must be developed. Specifically, [8] proposed an efficient subarray-wise channel estimation method based on the orthogonal matching pursuit (OMP) algorithm. Such an algorithm is applied under the consideration that the channel is spatially stationary in the subarrays, so the subarray channel becomes sparse in the angular domain. Differently, in [9] is demonstrated that the SW channel is no longer sparse in the angular domain since the energy of each radio path spread towards multiple angles. In this sense, the authors demonstrate that the XL-MIMO channel is sparse in the polar domain, and exploit such a property in the polar-domain simultaneous OMP algorithm, an accurate and low-complexity approach for XL-MIMO channel estimation.

In [10, 11], user scheduling techniques for MIMO systems that employ zero-forcing beamforming as the spatial multiplexing strategy are proposed. Specifically, the semiorthogonal user selection (SUS) algorithm proposed in [11] can achieve the same asymptotic achievable sum-rate as that of dirty paper coding. However, the analysis is restricted to up to 4 antennas at the BS and Rayleigh fading channel, which not applies to physically large arrays. On the other hand, the problem of user scheduling for the XL-MIMO systems with the SW model is addressed in [12]. The authors propose a strategy based on the equivalent distance, a measure that combines the distance from the users to the BS and the interference level produced by the other scheduled users. The developed algorithm can outperform the performance of SUS with less computational complexity. Despite that, analyses of the individual achievable rates and the coverage reached by the algorithm are not provided. In [13], authors take advantage of the non-overlapping visibility regions (VR) across the array in a crowded XL-MIMO system to propose a joint random access and user scheduling protocol. Such a protocol explores the different VRs of the users to improve the access performance, besides seeking users with non-overlapping VRs to be scheduled in the same payload data pilot resource.

In [14], the problem of RA in the DL of user-centric cell-free MIMO networks is studied. The authors propose an iterative method to promote max-min fairness in the system by scheduling users, allocating power, and selecting the pilot length. In [15], the problem of DL precoding in millimeter-wave (mm-Wave) systems is examined. Exploiting the sparsity of the mm-Wave channel in the angular domain, three different algorithms for beam selection and user scheduling are proposed. Compared with the state-of-the-art, the proposed methods allow for increasing the number of users to be served, especially when there are fewer available DL beams.

The problem of user scheduling, via optimizing the configurations of a reconfigurable

intelligent surface (RIS), and BS beamforming in multi-user networks assisted by RIS is analyzed in [16]. Specifically, a scheme with two graph neural networks is developed to schedule users and optimize the RIS configurations. The proposed scheme promotes fairness among the users, achieving better network utility.

In [17], the problem of maximizing energy efficiency under minimum throughput constraints in the DL channel of multi-cell massive MIMO schemes is studied. To increase the number of served users by each cell, a mini-slot-based transmit beamforming method is proposed, by scheduling the farther and nearer users into different time mini-slots. As a result, the BS can double the number of served users and reduce significantly the inter-cell interference, improving the performance of the border users.

Contribution. The contribution of this work is threefold.

i) We propose a QoS-aware joint user scheduling and power allocation algorithm based on the search in a graph for the DL channel of crowded XL-MIMO systems. Such an algorithm is designed to increase the number of served users with satisfied QoS requirements in XL-MIMO systems with high user density. The proposed graph-based technique can increase significantly not only the number of scheduled users, but the system achievable sum-rate, while providing fair coverage across the whole cell area.

ii) We propose a non-stationary multi-state channel model based on the SW propagation considering that users under LoS and NLoS transmission coexist in the same communication cell. In the proposed model, it is assumed that users under the LoS and NLoS states experience different propagation characteristics both in the multi-path fading and in the path loss rule. This is done aiming to capture the complexity of the propagation environment with physically large arrays of antennas.

iii) We extensively evaluate the performance of the proposed QoS-aware joint user scheduling and power allocation technique with numerical simulations. The performance of the proposed graph-based technique is evaluated in crowded XL-MIMO scenarios under different channel conditions and also compared with state-of-the-art techniques. As an outcome, a comprehensive analysis is obtained, including results on achievable sum-rate, number of scheduled users, computational complexity, and distribution of the scheduled users.

Notation. Boldface lowercase $\mathbf{a}$ and uppercase $\mathbf{A}$ letters represent vectors and matrices, respectively. Calligraphic letters $\mathcal{A}$ represent finite sets. $\mathbf{I}_n$ denotes the identity matrix of size n. $\mathbf{0}_n$ denotes the zero column vector of length n. $\{\cdot\}^{\mathsf{T}}$ and $\{\cdot\}^{\mathsf{H}}$ denote, respectively, the transpose and the complex conjugate transpose operators. $\wp(\cdot)$ denotes the power set operator.

7.2 System Model

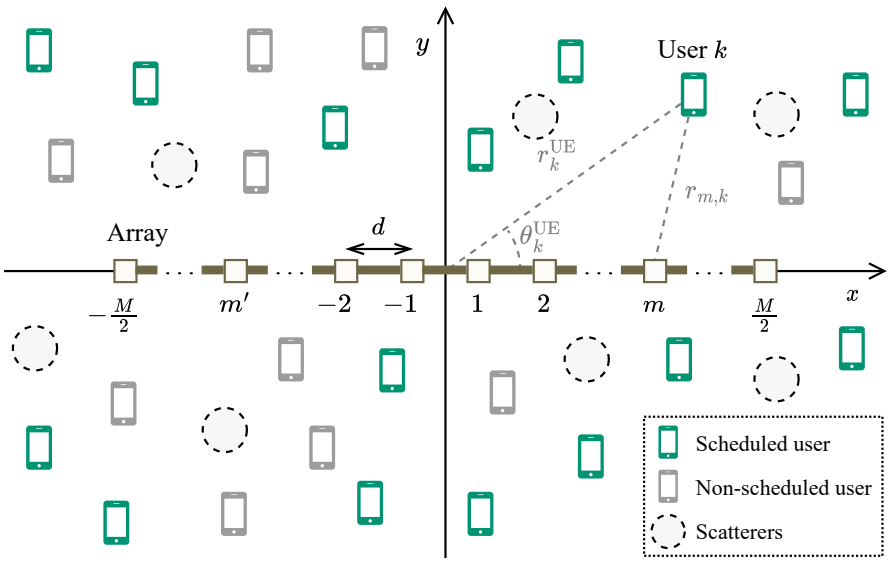
In the following, we describe the model of the communication system analyzed in this work. We consider the DL transmission of a narrowband XL-MIMO system with $K \in \mathbb{Z}_+$ users operating in the time-division duplexing (TDD) mode. The BS is equipped with $M \in \mathbb{Z}_+$ antennas organized as a uniform linear array (ULA) with elements spaced by the distance $d > 0$ meters. Therefore, the array has an aperture of $D = (M-1)d$ meters. A sketch of the crowded multi-user XL-MIMO communication scenario is depicted in Fig. 6. The BS antennas and the users are located at the xy-plane. Indeed, the ULA is oriented along the x-axis, and its center is at the origin of the coordinate system.

Adopting the representation in polar coordinates, the position of the user $k \in \{1, \ldots, K\}$ is described by the ordered pair $(r_k^{\mathrm{UE}}, \theta_k^{\mathrm{UE}})$, where $r_k^{\mathrm{UE}} > 0$ is the distance from the user k to the origin and $\theta_k^{\mathrm{UE}} \in [0, 2\pi)$ is the angle between the direction of user k and the x-axis direction (see Fig. 6). Similarly, the position of the antenna m is described by the pair $(r_m^{\mathrm{BS}}, \theta_m^{\mathrm{BS}})$, whose coordinates are calculated by:

$$r_m^{\mathrm{BS}} = \left(|m| - \frac{1}{2}\right) d \text{ and } \theta_m^{\mathrm{BS}} = \begin{cases} 0, & \text{if } m > 0 \\ \pi, & \text{if } m < 0 \end{cases}, \tag{7.1}$$

$\forall m \in [-M/2, M/2],\ m \neq 0.$

Figure 6 – Diagram of the crowded multi-user XL-MIMO communication scenario, indicating the scheduled and non-scheduled users, as well as the scatterers.



Source: The authors.

7.2.1 Channel Model

In this subsection, we formulate the channel model based on the SW model considering that users under LoS and NLoS channel states coexist in the same communication cell. For this reason, we define two channel vectors, one for the LoS channel model, and the other for the NLoS one. Then, we define a unified model capturing the multi-state aspect of the proposed channel model.

The LoS channel follows the SW model. Hence, the channel response between the antenna m and the user k can be expressed as [12]:

$$a_{m,k} = \sqrt{\frac{\beta_0^{\text{LoS}}}{r_{m,k}^{\gamma^{\text{LoS}}}}} e^{-j\frac{2\pi}{\lambda} r_{m,k}}, \tag{7.2}$$

where $\beta_0^{\text{LoS}} > 0$ is the path loss attenuation at a reference distance, $\gamma^{\text{LoS}} > 0$ is the path loss exponent, $\lambda > 0$ is the carrier wavelength, and $r_{m,k} > 0$ is the distance from the antenna m to the user k, calculated by:

$$r_{m,k} = \sqrt{(r_k^{\text{UE}})^2 + (r_m^{\text{BS}})^2 - 2 r_k^{\text{UE}} r_m^{\text{BS}} \cos(\theta_k^{\text{UE}} - \theta_m^{\text{BS}})}. \tag{7.3}$$

Considering (7.2), the channel vector $\mathbf{a}_k^{\text{LoS}} \in \mathbb{C}^M$ for the LoS channel is equal to:

$$\mathbf{a}_k^{\text{LoS}} = [a_{1,k} \ \cdots \ a_{M,k}]^{\intercal}. \tag{7.4}$$

Differently, the NLoS channel follows the i.i.d. Rayleigh-fading model with the path loss computed independently for each antenna due to the variation of the average received power across the large-aperture XL-MIMO array [6]. The path loss of the NLoS radio link between the antenna m and user k is equal to:

$$\beta_{m,k} = \frac{\beta_0^{\text{NLoS}}}{r_{m,k}^{\gamma^{\text{NLoS}}}}, \tag{7.5}$$

where $\beta_0^{\text{NLoS}} > 0$ is the path loss attenuation at a reference distance and $\gamma^{\text{NLoS}} > 0$ is the path loss exponent, and $r_{m,k}$ is as in (7.3). Hence, the channel vector $\mathbf{a}_k^{\text{NLoS}} \in \mathbb{C}^M$ for the NLoS channel is defined such that:

$$\mathbf{a}_k^{\text{NLoS}} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{\Sigma}_k), \tag{7.6}$$

with the diagonal covariance matrix $\mathbf{\Sigma}_k \in \mathbb{R}^{M \times M}$ containing the path loss coefficients w.r.t. all the antenna elements, *i.e.*,

$$\mathbf{\Sigma}_k = \text{diag}([\beta_{1,k} \ \cdots \ \beta_{M,k}]^{\intercal}). \tag{7.7}$$

Note that the uncorrelated channel assumption in (7.7) is justified since the antenna elements are separated by the distance $d \geq \lambda/2$.

Definition 7.1. Let $x_k \in \{0, 1\}$ be the *channel state indicator* associated with the user k, equal to 1 if the channel is under the LoS state, or 0 if it is under the NLoS state. To capture the influence of topographic features related to the communication cell on the channel state, *e.g.*, the effect of relief, as well as the spatial configuration of scatterers and obstacles, the indicator is modeled as a random variable. Therefore, $x_k \sim f_{x_k|(r_k^{\mathrm{UE}},\theta_k^{\mathrm{UE}})}(x \mid (r_k^{\mathrm{UE}}, \theta_k^{\mathrm{UE}}))$, where $f_{x_k|(r_k^{\mathrm{UE}},\theta_k^{\mathrm{UE}})} : \{0, 1\} \to \mathbb{R}_+^*$ is the conditional probability mass function (pmf) that depends on the position of the user k in the cell.

Given the definition of the channel state indicator, as well as the LoS and NLoS channel vectors, the *multi-state channel vector* $\mathbf{a}_k \in \mathbb{C}^M$ can be defined as:

$$\mathbf{a}_k = x_k \mathbf{a}_k^{\mathrm{LoS}} + (1 - x_k)\mathbf{a}_k^{\mathrm{NLoS}}. \tag{7.8}$$

Notice that, when $x_k = 1$, eq. (7.8) is equal to the LoS channel vector. On the other hand, when $x_k = 0$, the channel vector of user k is equal to the NLoS channel vector. Hence, users with different channel states may coexist in the same communication cell, depending on the definition of the state indicator pmf, $f_{x_k|(r_k^{\mathrm{UE}},\theta_k^{\mathrm{UE}})}$.

In this sense, for the sake of simplicity and to enable the evaluation of the proposed techniques in a variety of channel scenarios, in the remainder of this work we consider that the channel state indicators follow a Bernoulli random distribution with parameter $0 \leq \rho \leq 1$, namely the *LoS probability*. Therefore, the conditional pmf results:

$$f_{x_k|(r_k^{\mathrm{UE}},\theta_k^{\mathrm{UE}})}(x \mid (r_k^{\mathrm{UE}}, \theta_k^{\mathrm{UE}})) = \rho^x (1 - \rho)^{1-x}, \tag{7.9}$$

where $x \in \{0, 1\}$ and $\forall k \in \{1, \ldots, K\}$.

7.2.2 Signal Model

Now, we define the model for the signal received by the users. Let $\mathcal{K} \subseteq \{1, \ldots, K\}$ be the set of scheduled users. The transmitted signal by the BS is equal to:

$$\mathbf{z} = \textstyle\sum_{k \in \mathcal{K}} \sqrt{p_k} s_k \mathbf{f}_k, \tag{7.10}$$

where $p_k > 0$ is the power allocated for the user k, $s_k \in \mathbb{C}$ such that $\mathbb{E}\{|s_k|^2\} = 1$ is the signal intended for the user k, and $\mathbf{f}_k \in \mathbb{C}^M$ such that $\|\mathbf{f}_k\|_2^2 = 1$ is the precoding vector computed for the user k. The received signal by the user $k \in \mathcal{K}$ is equal to:

$$y_k = \sqrt{p_k} s_k \mathbf{a}_k^{\mathsf{H}} \mathbf{f}_k + \textstyle\sum_{i \in \mathcal{K} \setminus k} \sqrt{p_i} s_i \mathbf{a}_k^{\mathsf{H}} \mathbf{f}_i + w_k, \tag{7.11}$$

where $w_k \sim \mathcal{CN}(0, \sigma_w^2)$ is the additive white Gaussian noise. Given the received signal in (7.11), the signal-to-interference-plus-noise ratio (SINR) calculated for the user $k \in \mathcal{K}$ is equal to [12]:

$$\mathrm{SINR}_k = \frac{p_k |\mathbf{a}_k^{\mathsf{H}} \mathbf{f}_k|^2}{\sum_{i \in \mathcal{K} \setminus k} p_i |\mathbf{a}_k^{\mathsf{H}} \mathbf{f}_i|^2 + \sigma_w^2}. \tag{7.12}$$

Without loss of generality, let $k \in \{1, \ldots, |\mathcal{K}|\}$ be the indices of the scheduled users. Then, the channel matrix with the channel vectors of all the scheduled users is defined as $\mathbf{A} \in \mathbb{C}^{M \times |\mathcal{K}|}$ such that $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_{|\mathcal{K}|}]$. For the sake of simplicity, we consider perfect CSI available at the transmitter[1], and that $|\mathcal{K}| \leq M$ and $\text{rank}(\mathbf{A}) = |\mathcal{K}|$, then the BS can transmit the DL signal using the zero-forcing (ZF) precoder. Hence, the precoding vector for each user $k \in \mathcal{K}$ is equal to:

$$\mathbf{f}_k^{\text{ZF}} = \frac{\mathbf{A}(\mathbf{A}^{\text{H}}\mathbf{A})^{-1}\mathbf{e}_k}{[(\mathbf{A}^{\text{H}}\mathbf{A})^{-1}]_{k,k}^{\frac{1}{2}}}, \tag{7.13}$$

where $\mathbf{e}_k \in \{0,1\}^{|\mathcal{K}|}$ is the k-th vector of the standard basis of the $|\mathcal{K}|$-dimensional Euclidean space.

Considering that the ZF precoder mitigates the inter-user interference (IUI), *i.e.*, $\mathbf{a}_i^{\text{H}}\mathbf{f}_j = 0, \forall i, j \in \mathcal{K}, i \neq j$, substituting (7.13) in eq. (7.12) results in the SINR calculated for user k using the ZF precoder:

$$\text{SINR}_k^{\text{ZF}} = \frac{p_k}{\sigma_w^2[(\mathbf{A}^{\text{H}}\mathbf{A})^{-1}]_{k,k}}. \tag{7.14}$$

Using Shannon's equation, the achievable rate of the user $k \in \mathcal{K}$ using the ZF precoder is equal to:

$$R_k^{\text{ZF}} = \log_2\left(1 + \frac{p_k}{\sigma_w^2[(\mathbf{A}^{\text{H}}\mathbf{A})^{-1}]_{k,k}}\right). \tag{7.15}$$

Observing (7.15), we note that the achievable rate of the user k not only depends on its respective allocated power and the noise power, but also on the overall set of scheduled users and their respective channel vectors. For this reason, the user scheduling process is crucial for attaining reasonable performance levels.

Let $\overline{R}_k > 0, \forall k \in \mathcal{K}$ be the minimum achievable rate that the BS must serve to user k. Given the set of scheduled users $\mathcal{K}$, the maximum DL achievable sum-rate of the XL-MIMO system using the ZF precoder is succeeded with the allocated powers that solve the following optimization problem:

$$\{p_k^*\}_{k\in\mathcal{K}} = \underset{\{p_k\}_{k\in\mathcal{K}}}{\arg\max} \quad \sum_{k\in\mathcal{K}} R_k^{\text{ZF}}, \tag{7.16a}$$

$$\text{subject to} \quad R_k^{\text{ZF}} \geq \overline{R}_k, \forall k \in \mathcal{K}, \tag{7.16b}$$

$$\sum_{k\in\mathcal{K}} p_k \leq P_{\max}, \tag{7.16c}$$

$$p_k > 0, \forall k \in \mathcal{K}, \tag{7.16d}$$

where $P_{\max} > 0$ is the maximum power available for DL transmission, and the achievable rates $R_k^{\text{ZF}}, \forall k \in \mathcal{K}$ are given by eq. (7.15). Since eq. (7.16) is equivalent to the optimization problem of allocating power on independent parallel Gaussian channels, the set of powers $\{p_k^*\}_{k\in\mathcal{K}}$ that solve it follows the water-filling distribution [18].

[1] Further, in Subsection 7.5.4, the impact of inaccurate CSI at the transmitter on the performance of the user scheduling techniques is numerically evaluated.

7.3 User Scheduling: Problem Formulation

In this section, we introduce the formulation of the studied user scheduling problem. The optimization problem of joint DL user scheduling and power allocation with individual minimum achievable rate constraints and transmit power budget can be defined as:

$$\mathcal{P}_0 : \underset{\mathcal{K},\{p_k\}_{k\in\mathcal{K}}}{\text{maximize}} \quad \sum_{k\in\mathcal{K}} R_k^{\text{ZF}}, \tag{7.17a}$$

$$\text{subject to} \quad R_k^{\text{ZF}} \geq \overline{R}_k, \forall k \in \mathcal{K}, \tag{7.17b}$$

$$\sum_{k\in\mathcal{K}} p_k \leq P_{\max}, \tag{7.17c}$$

$$\mathcal{K} \subseteq \{1,\ldots,K\}, \tag{7.17d}$$

$$p_k > 0, \forall k \in \mathcal{K}. \tag{7.17e}$$

The constraints (7.17b) ensure that all the scheduled users are served with a minimum achievable rate. Moreover, the constraint (7.17c) ensures that the DL transmitted power does not exceed $P_{\max}$. Finally, the constraints (7.17e) and (7.17d) define the domain of the optimization variables.

The optimization problem $\mathcal{P}_0$ is concave in the variables $\{p_k\}_{k\in\mathcal{K}}$, but not in the variable $\mathcal{K}$. For this reason, it isn't possible to solve $\mathcal{P}_0$ optimally with standard convex optimization tools. An alternative path to reach a sub-optimal solution is to split $\mathcal{P}_0$ into two sub-problems in which each variable is optimized independently. We discuss this strategy in the sequel.

Let $g : \wp(\{\mathbf{a}_k\}_{k=1}^K) \rightarrow \mathbb{R}_+$ be a function that measures the **spatial compatibility** between users from their channel vectors. The spatial compatibility quantifies how efficiently these channel vectors can be separated in space. Examples of spatial compatibility metrics are the *condition number* and the *null-space projection* of the channel matrix. Since there exists a correspondence between spatial compatible users and the precoding performance, optimizing a spatial compatibility metric is a promising path to obtain a good set of scheduled users [5]. The generic user scheduling problem solved by maximizing a spatial compatibility metric is called *user grouping* and can be formulated as:

$$\mathcal{P}_1 : \ \mathcal{K}^* = \arg\max_{\mathcal{K}} \quad g(\{\mathbf{a}_k\}_{k\in\mathcal{K}}), \tag{7.18a}$$

$$\text{subject to} \quad \mathcal{K} \subseteq \{1,\ldots,K\}. \tag{7.18b}$$

The optimization problem $\mathcal{P}_1$ is an NP-complete combinatorial problem solved only by exhaustive search. Since, in crowded XL-MIMO systems, the number of users in the communication cell is high, the size of the solution space of $\mathcal{P}_1$ scales quickly. Hence, in such a case, it is impractical to solve the user grouping problem in a feasible time. For this reason, Section 7.4 develops an effective, quasi-optimal, and computationally efficient method to carry out user scheduling in crowded XL-MIMO scenarios.

Given the set of scheduled users $\mathcal{K}^*$, the optimal set of allocated powers $\{p_k^*\}_{k\in\mathcal{K}^*}$ can be calculated by solving the following optimization sub-problem:

$$\mathcal{P}_2: \ \{p_k^*\}_{k\in\mathcal{K}^*} = \underset{\{p_k\}_{k\in\mathcal{K}^*}}{\arg\max} \quad \sum_{k\in\mathcal{K}^*} R_k^{\mathrm{ZF}}, \tag{7.19a}$$

$$\text{subject to} \quad R_k^{\mathrm{ZF}} \geq \overline{R}_k, \forall k \in \mathcal{K}^*, \tag{7.19b}$$

$$\sum_{k\in\mathcal{K}^*} p_k \leq P_{\max}, \tag{7.19c}$$

$$p_k > 0, \forall k \in \mathcal{K}^*. \tag{7.19d}$$

The optimization problem $\mathcal{P}_2$ is identical to (7.16) and, if feasible, it can be solved optimally by the water-filling solution [18]. A simple way to check the feasibility of $\mathcal{P}_2$ is presented in the following.

Remark 7.1. The optimization problem $\mathcal{P}_2$ is feasible if and only if the sum of the minimum allocated powers necessary to serve each user with its respective minimum achievable rate does not exceed $P_{\max}$, *i.e.*,

$$\sum_{k\in\mathcal{K}^*} \overline{p}_k = \sigma_w^2 \sum_{k\in\mathcal{K}^*} (2^{\overline{R}_k} - 1)[(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}]_{k,k} \leq P_{\max}, \tag{7.20}$$

where $\overline{p}_k > 0$ is the minimum power required to serve user k with the achievable rate in eq. (7.15) equal to the minimum achievable rate $\overline{R}_k$.

Considering that the feasibility criterion in eq. (7.20) is satisfied, the solution of $\mathcal{P}_2$ is given by:

$$p_k^* = \max(\overline{p}_k, \ \mu - \sigma_w^2[(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}]_{k,k}), \tag{7.21}$$

$\forall k \in \mathcal{K}$, where $\mu \in \mathbb{R}$ is a constant called *water-level*. Moreover, to meet the constraint (7.19c) with equality, the optimal water-level can be obtained by satisfying:

$$\sum_{k\in\mathcal{K}^*} \max(\overline{p}_k, \ \mu - \sigma_w^2[(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}]_{k,k}) - P_{\max} = 0, \tag{7.22}$$

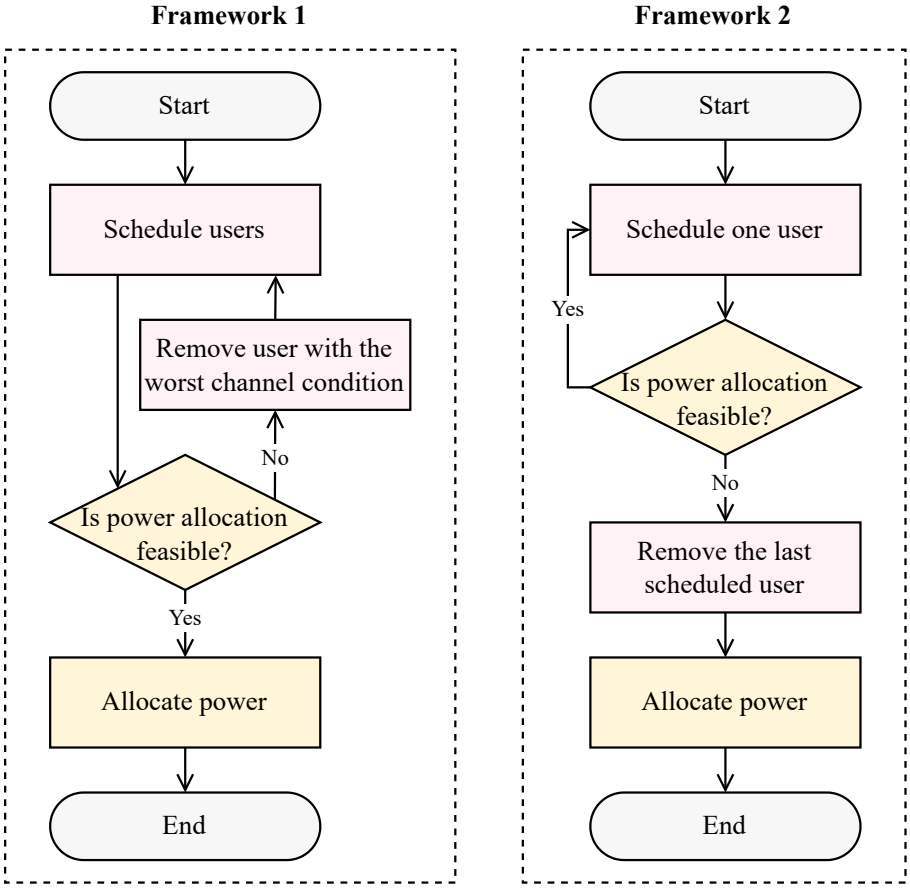
which can be easily solved by a root-finding algorithm [18].

A sub-optimal solution of the original problem $\mathcal{P}_0$ can be obtained by sequentially solving $\mathcal{P}_1$ and $\mathcal{P}_2$. Since the two optimization variables are decoupled in the formulated sub-problems, the set of scheduled users $\mathcal{K}^*$ may result in an infeasible power allocation policy. In such a case, the set $\mathcal{K}^*$ must be altered to enable power allocation.

To solve the user scheduling problem by sequentially solving the sub-problems $\mathcal{P}_1$ and $\mathcal{P}_2$, we propose the two distinct frameworks presented in Fig. 7. In *Framework 1*, firstly the set of scheduled users is computed. Next, if the power allocation with the calculated set of users is infeasible, the users with the worst channel condition are removed until feasibility is reached. Finally, the power allocation is carried out. Differently, in *Framework 2*, users are scheduled iteratively until the power allocation problem becomes infeasible. When an infeasible set of users is reached, the last scheduled user is removed, then a power allocation procedure is carried out. From an implementation perspective, in general,

solutions that fit into Framework 2 demand higher computational complexity than those fitting into Framework 1. This is due to the power allocation feasibility test carried out at every iteration, which commonly requires the calculation of the precoding vectors. Although in Framework 1 this feasibility test is carried out as well, if the user scheduling procedure is carefully designed, the number of tests can be drastically reduced, increasing its computational advantage w.r.t. Framework 2.

Figure 7 – Flowcharts of the two distinct frameworks adopted for joint user scheduling and power allocation procedure by sequentially solving $\mathcal{P}_1$ and $\mathcal{P}_2$. Note that, in Framework 2, the feasibility of the power allocation problem is checked at every iteration of the user scheduling procedure. On the other hand, in Framework 1, the power allocation feasibility check is carried out more than once only if the set of scheduled users results in an infeasible power allocation problem.



Source: The authors.

7.3.1 $\mathcal{P}_2$ Infeasibility Test

In this subsection, we present an efficient method to check the infeasibility of the optimization problem (7.19) without needing to calculate the inverse matrix $(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}$. This method motivates the development of the graph representation used in the proposed

scheduling algorithm and reduces significantly the number of operations required to test the feasibility of the set of scheduled users.

Let $R_k^{\mathrm{SU}} > 0$ be the single-user capacity of the user k calculated by:

$$R_k^{\mathrm{SU}} = \log_2\left(1 + \frac{p_k \|\mathbf{a}_k\|_2^2}{\sigma_w^2}\right). \tag{7.23}$$

The single-user capacity is the achievable rate if, during the DL, the BS transmits only the signal of user k. Following this definition, the minimum required power $\overline{p}_k^{\mathrm{SU}} > 0$ for the user k to experience its minimum achievable rate $\overline{R}_k$ is equal to:

$$\overline{p}_k^{\mathrm{SU}} = \frac{\sigma_w^2}{\|\mathbf{a}_k\|_2^2}(2^{\overline{R}_k} - 1). \tag{7.24}$$

Lemma 7.1. *For any set of scheduled users $\mathcal{K} \subseteq \{1, \ldots, K\}$, if the sum of the minimum powers required to equal the single-user capacity of each scheduled user to its minimum achievable rate is equal to or greater than $P_{\max}$, i.e., $\sum_{k\in\mathcal{K}} \overline{p}_k^{SU} \geq P_{\max}$, the optimization problem* (7.19) *is infeasible. Such a condition is sufficient but not necessary to confirm the infeasibility of $\mathcal{P}_2$. If $\sum_{k\in\mathcal{K}} \overline{p}_k^{SU} < P_{\max}$, the feasibility or infeasibility of $\mathcal{P}_2$ can only be proved by checking whether eq.* (7.20) *holds.*

Proof. The effective channel gain obtained by user k with the ZF precoder is upper-bounded by [19]:

$$\|\mathbf{a}_k\|_2^2 \geq [(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}]_{k,k}^{-1} = \|\mathbf{a}_k\|_2^2 - \mathbf{a}_k^{\mathsf{H}}\breve{\mathbf{A}}_k(\breve{\mathbf{A}}_k^{\mathsf{H}}\breve{\mathbf{A}}_k)^{-1}\breve{\mathbf{A}}_k^{\mathsf{H}}\mathbf{a}_k, \tag{7.25}$$

where $\breve{\mathbf{A}}_k \in \mathbb{C}^{M\times|\mathcal{K}|-1}$ is the channel matrix with the channel vectors of all the scheduled users, except for k, *i.e.*,

$$\breve{\mathbf{A}}_k = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_{k-1} \ \mathbf{a}_{k+1} \ \cdots \ \mathbf{a}_{|\mathcal{K}|}]. \tag{7.26}$$

Since $\mathbf{A}$ has full rank, $(\breve{\mathbf{A}}_k^{\mathsf{H}}\breve{\mathbf{A}}_k)^{-1}$ is positive definite and, consequently, the equality in eq. (7.25) is obtained if and only if $\breve{\mathbf{A}}_k^{\mathsf{H}}\mathbf{a}_k = \mathbf{0}_{|\mathcal{K}|-1}$, *i.e.*, the channel vector of user k is orthogonal to the channel vectors of all the other users. Therefore, we obtain the following relationship between the sum of the minimum allocated powers required to attain the minimum achievable rates of the scheduled users:

$$\sum_{k\in\mathcal{K}} \overline{p}_k \geq \sum_{k\in\mathcal{K}} \overline{p}_k^{\mathrm{SU}}. \tag{7.27}$$

Accordingly, noticing that it is impossible to get perfectly orthogonal channel vectors in practice, if $\sum_{k\in\mathcal{K}} \overline{p}_k^{\mathrm{SU}} \geq P_{\max}$ we have that $\sum_{k\in\mathcal{K}} \overline{p}_k > P_{\max}$, indicating that $\mathcal{K}$ is an infeasible set of scheduled users for the optimization problem $\mathcal{P}_2$. Conversely, in the ideal case where the scheduled users have all mutually orthogonal channel vectors, eq. (7.27) attains equality. □

7.4 User Scheduling Based on Clique Search

In this section, first, we introduce the concept of the undirected vertex-weighted graph (UWG) model for modeling the interference between the users in the proposed user scheduling XL-MIMO operating under non-stationary multi-state LoS and NLoS channels. Then, we formulate a clique search problem on the UWG to solve the user scheduling and power allocation sub-problems $\mathcal{P}_1$ and $\mathcal{P}_2$ proposed in Section 7.3.

7.4.1 Undirected Vertex-Weighted Graph Model

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a UWG, where $\mathcal{V} = \{v_1, \ldots, v_V\}$ is the set with the graph vertices such that $|\mathcal{V}| = V$, and $\mathcal{E} \subseteq \{\{v_i, v_j\} \mid v_i, v_j \in \mathcal{V}, v_i \neq v_j\}$ is the set with the graph edges. Let $\mathbf{E} \in \{0,1\}^{V \times V}$ be the adjacency matrix of the graph $\mathcal{G}$ such that:

$$[\mathbf{E}]_{i,j} = \begin{cases} 1, & \text{if } \{v_i, v_j\} \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases}. \tag{7.28}$$

The vertex weight function $\omega : \mathcal{V} \to \mathbb{R}$ characterizes the weight of each vertex $v_i \in \mathcal{V}$.

In our work, the UWG $\mathcal{G}$ represents the users in the communication cell and the orthogonality relationship between their respective channel vectors. Each user $k \in \{1, \ldots, K\}$ is represented by a vertex v_k. Moreover, the edges $\mathcal{E}$ are described by the adjacency matrix constructed from the channel vectors according to the ϵ-orthogonal rule,

$$[\mathbf{E}]_{i,j} = \begin{cases} 0, & \text{if } i = j \\ I_{\{\epsilon' \in \mathbb{R} \mid \epsilon' < \epsilon\}} \left(\frac{|\mathbf{a}_i^{\mathrm{H}} \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2} \right), & \text{otherwise} \end{cases}, \tag{7.29}$$

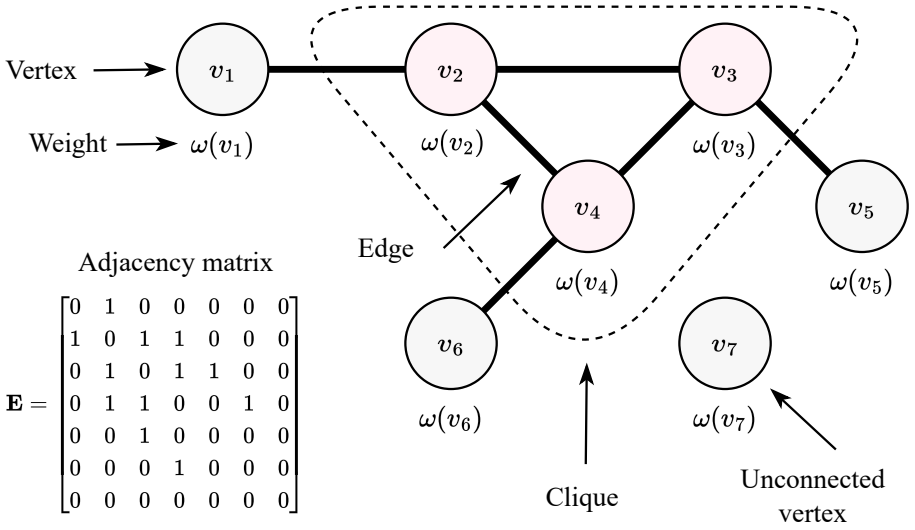
$\forall i, j \in \{1, \ldots, K\}$, where $I_{\mathcal{A}}(\cdot)$ denotes the indicator function and $0 \leq \epsilon \leq 1$ represents the *admissibility for channel orthogonality*. Such a method for defining the graph edges will make only those vertices representing users with quasi-orthogonal channel vectors to be connected. Finally, the weight of vertex k is defined as the minimum power required for user k to achieve a single-user capacity equal to its minimum achievable rate, *i.e.*,

$$\omega(v_k) = \overline{p}_k^{\mathrm{SU}} = \frac{\sigma_w^2}{\|a_k\|_2^2} (2^{\overline{R}_k} - 1), \ \forall k \in \{1, \ldots, K\}. \tag{7.30}$$

Fig. 8 depicts a diagram of a UWG representation of a hypothetical XL-MIMO system with $K = 7$ users and its equivalent adjacency matrix.

Definition 7.2. The subgraph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ is called a *clique* if the vertices in $\mathcal{V}'$ are mutually adjacent to one another, *i.e.*, $\mathcal{E}' = \{\{v_i, v_j\} \mid v_i, v_j \in \mathcal{V}', v_i \neq v_j\}$. Moreover, the number of vertices in the clique is called the *clique number* [20].

Figure 8 – UWG representation of a hypothetical XL-MIMO system with $K = 7$ users and its equivalent adjacency matrix. The vertices v_2, v_3, and v_4 form a clique in this graph.



Source: The authors.

7.4.2 User Scheduling Based on Clique Search

Now, we formulate the user scheduling process as a clique search problem in the UWG $\mathcal{G}$ that represents the XL-MIMO system. Let $\mathcal{C}_{\mathcal{G}}$ be the set of all the cliques of the graph $\mathcal{G}$. We aim to find the clique with the largest clique number and with the sum of the weights less than $P_{\max}$. This clique search problem can be written in the form:

$$\mathcal{P}_3: \ \mathcal{G}^* = \arg\max_{\mathcal{G}'} \quad |\mathcal{V}'|, \tag{7.31a}$$

$$\text{subject to} \quad \textstyle\sum_{v_k \in \mathcal{V}'} \omega(v_k) < P_{\max}, \tag{7.31b}$$

$$\mathcal{G}' = (\mathcal{V}', \mathcal{E}') \in \mathcal{C}_{\mathcal{G}}. \tag{7.31c}$$

The constraint (7.31b), derived from Lemma 7.1, ensures via the sum of the weights of the clique vertices that the set of scheduled users represented by the vertices results in a possibly feasible $\mathcal{P}_2$. Moreover, the constraint (7.31c) ensures that $\mathcal{G}^*$ is a clique of graph $\mathcal{G}$. Typically, clique problems are NP-complete, requiring high computational effort to be solved in large graphs. Specifically, problem $\mathcal{P}_3$ can be solved by clique enumeration [20], which consists of listing all the possible cliques in the graph $\mathcal{G}$, evaluating which one simultaneously satisfies (7.31a) and (7.31b). The search space of this procedure can be reduced by evaluating the constraint (7.31b) during each iteration of the clique enumeration. However, the worst-time complexity of this procedure is still the same as the original one. For this reason, we propose a low-complexity procedure to calculate a near-optimal solution of $\mathcal{P}_3$.

Algorithm 1 presents a method to find a near-optimal solution for the clique search problem $\mathcal{P}_3$. We call this algorithm as *clique search-based scheduling* (CBS). In this

pseudocode, the operator neighbors : $\mathcal{V} \to \wp(\mathcal{V})$ returns the set of vertices that have edges with the input vertex v_k, *i.e.*, neighbors$(v_k) = \{v_i \mid \{v_k, v_i\} \in \mathcal{E}\}$.

Algorithm 1 CBS – Greedy algorithm to solve the clique search problem $\mathcal{P}_3$.

input: The set of graph vertices, $\mathcal{V}$
output: The set of clique vertices, $\mathcal{V}'$

1: $v_k \leftarrow \arg\min \{\omega(v_i) \mid v_i \in \mathcal{V}\}$
2: $\mathcal{V}'^{(0)} \leftarrow \{v_k\}$
3: $\mathcal{N}^{(0)} \leftarrow \text{neighbors}(v_k)$
4: $\Omega^{(0)} \leftarrow \omega(v_i)$
5: $n \leftarrow 0$
6: **repeat**
7: $\quad v_k \leftarrow \arg\min \{\omega(v_i) \mid v_i \in \mathcal{N}^{(n)}\}$
8: $\quad \mathcal{V}'^{(n+1)} \leftarrow \mathcal{V}'^{(n)} \cup \{v_k\}$
9: $\quad \mathcal{N}^{(n+1)} \leftarrow \mathcal{N}^{(n)} \cap \text{neighbors}(v_k)$
10: $\quad \Omega^{(n+1)} \leftarrow \Omega^{(n)} + \omega(v_i)$
11: $\quad$ **if** $\Omega^{(n+1)} \geq P_{\max}$ **then**
12: $\quad\quad$ exit loop
13: $\quad$ **end if**
14: $\quad n \leftarrow n + 1$
15: **until** $\mathcal{N}^{(n)} = \emptyset$
16: **return** $\mathcal{V}' \leftarrow \mathcal{V}'^{(n)}$

The greedy algorithm starts by finding the vertex that requires the least weight and adding it to the clique. The neighboring vertices of the first vertex constitute the clique neighborhood, computed in line 3. In the loop beginning at line 6, the vertex of the clique neighborhood with the least weight is added to the clique. Next, in line 9, the clique neighborhood is updated with the vertices that are simultaneously neighbors of all the clique vertices. This loop repeats until the sum of the weights in the clique is less than or equal to $P_{\max}$, or if there are no more vertices in the clique neighborhood. When the first stop criterion is met, the algorithm outputs the set of vertices obtained during the previous iteration, satisfying the constraint (7.31b). Hence, the set of scheduled users is calculated from the output of the Algorithm 1 by $\mathcal{K} = \{k \in \{1, \ldots, K\} \mid v_k \in \mathcal{V}'\}$. It is worth mentioning that the scheduling criterion used in CBS results in a set of users *possibly feasible* w.r.t. $\mathcal{P}_2$, without guaranteeing that the users in $\mathcal{K}$ can be scheduled satisfying the power budget[2] and minimum achievable rate constraints simultaneously. For this reason, we need a procedure to verify the feasibility of $\mathcal{K}$ w.r.t. the problem $\mathcal{P}_2$ and remove users from the set if the power allocation is infeasible. We describe the adopted approach in the following.

[2] Constraint (7.31b) is a necessary condition for ensuring the power feasibility of the solution, but not a sufficient condition. This occurs since the simplified feasibility test in constraint (7.31b) assumes perfect orthogonality between scheduled users. If there exists a certain interference level between them, more transmit power is required, and therefore a more accurate, time-consuming procedure to verify the feasibility of the solution has to be performed.

7.4.3 Obtaining a Feasible Set of Scheduled Users

Algorithm 2 User removal – Obtaining a feasible set of scheduled users.

input: The infeasible set of scheduled users, $\mathcal{K}$, the minimum achievable rates, $\{\overline{R}_k\}_{k\in\mathcal{K}}$, and the channel vectors, $\{\mathbf{a}_k\}_{k\in\mathcal{K}}$

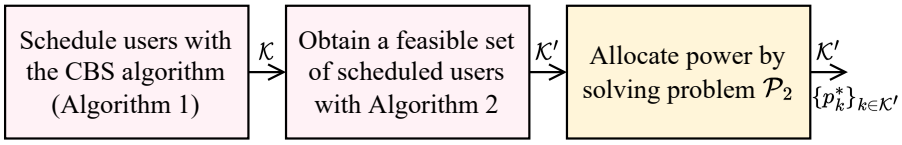
output: The feasible set of scheduled users, $\mathcal{K}'$

1: $\mathcal{K}' \leftarrow \mathcal{K}$
2: $\mathbf{A} \leftarrow [\mathbf{a}_1 \ \cdots \ \mathbf{a}_{|\mathcal{K}'|}]$
3: **repeat**
4: $\quad k^* \leftarrow \arg\min \{\|\mathbf{a}_k\|_2^2 \mid k \in \mathcal{K}'\}$
5: $\quad \mathcal{K}' \leftarrow \mathcal{K}' \backslash k^*$
6: $\quad \mathbf{A} \leftarrow [\mathbf{a}_1 \ \cdots \ \mathbf{a}_{k^*-1} \ \mathbf{a}_{k^*+1} \ \cdots \ \mathbf{a}_{|\mathcal{K}'|}]$
7: $\quad \overline{p}_k = \sigma_w^2(2^{\overline{R}_k} - 1)[(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}]_{k,k}, \forall k \in \mathcal{K}'$
8: **until** $\sum_{k\in\mathcal{K}'} \overline{p}_k \leq P_{\max}$
9: **return** $\mathcal{K}'$

When the set of scheduled users produced by the CBS algorithm is infeasible, a user removal procedure must be adopted to get a feasible one. The proposed user removal algorithm is based on the $\mathcal{P}_2$ feasibility criterion in eq. (7.20) and described in Algorithm 2. In this procedure, the user with the lowest channel power, namely the user with the worst channel condition, is removed from the set of scheduled users until the power allocation problem $\mathcal{P}_2$ becomes feasible. Accordingly, in lines 4-7, the user with the lowest channel power is removed, then the minimum powers necessary to serve the users in the reduced set are calculated. These steps repeat until the power allocation feasibility is confirmed according to the condition in eq. (7.20), as evaluated in line 8.

Finally, with a feasible set of scheduled users, power allocation can be carried out satisfying both the *transmit power budget* and the *minimum achievable rate* constraints. Fig. 9 sketches out the whole proposed technique for joint user scheduling and power allocation problems in crowded XL-MIMO systems.

Figure 9 – Flowchart for the proposed joint user scheduling and power allocation techniques, constituted by the CBS algorithm, a user removal algorithm, and optimal power allocation.



Source: The authors.

7.4.4 Scheduling Users Analyzing Their Channel Powers

Now, we describe a simple but effective approach to schedule users based on the powers of the channel vectors. Let $P_k > 0$ be the power of the channel vector of user k, calculated

from the multi-state channel vector in eq. (7.8) by:

$$P_k = \|\mathbf{a}_k\|_2^2. \tag{7.32}$$

The power of the channel vector is a suitable measure of the channel quality for a given user, resulting from its distance w.r.t. the BS array and its channel state. Moreover, by inspecting eqs. (7.20) and (7.25), one can see that the minimum allocated power necessary to serve user $k \in \mathcal{K}$ with its respective minimum achievable rate is inversely proportional to P_k,

$$\overline{p}_k = \frac{\sigma_w^2(2^{\overline{R}_k} - 1)}{P_k - \mathbf{a}_k^{\mathsf{H}} \breve{\mathbf{A}}_k (\breve{\mathbf{A}}_k^{\mathsf{H}} \breve{\mathbf{A}}_k)^{-1} \breve{\mathbf{A}}_k^{\mathsf{H}} \mathbf{a}_k}. \tag{7.33}$$

Therefore, we use the channel powers to develop a user scheduling technique named *channel power-based scheduling* (CPBS).

Let $n \in \mathbb{Z}_+^*$ be the number of the iteration of the CPBS algorithm. During each iteration, the CPBS schedules the user with the largest channel power, solving the following optimization problem until a stop criterion is met:

$$k^* = \arg\max_k \quad P_k, \tag{7.34a}$$

$$\text{subject to} \quad k \in \{1, \ldots, K\} \backslash \mathcal{K}^{(n-1)}. \tag{7.34b}$$

The pseudocode of the CPBS algorithms is given in Algorithm 3. The CPBS algorithm operates with the procedure described in the following. In lines 5 and 6, the algorithm schedules the user with the largest channel power by solving problem (7.34). The algorithm repeats this procedure until the set of scheduled users results in an infeasible power allocation problem according to the criterion described in Remark 7.1, or if all the users in the communication cell are scheduled. Similarly to the CBS algorithm, the set of scheduled users calculated by the CPBS generates an infeasible power allocation problem. Therefore, the procedure derived in Section 7.4.3 must be applied to $\mathcal{K}$ to generate the final set of scheduled users.

7.5 Numerical Results

In this section, we present numerical results to demonstrate the effective performance of the introduced user scheduling methods operating in crowded XL-MIMO systems. In the Monte-Carlo simulations, we consider $K = 10^3$ users located inside a cell such that $r_k^{\mathrm{UE}} \in [0.03, 1]$ km and $\theta_k^{\mathrm{UE}} \in [-\pi, \pi]$, $\forall k \in \{1, \ldots, K\}$. The users are uniformly distributed in the cell area. Hence, the angles θ_k^{UE} follow a uniform distribution, while the distances r_k^{UE} follow the probability density function [21]:

$$f_{r_k^{\mathrm{UE}}}(r) = \begin{cases} 2r(r_{\max}^2 - r_{\min}^2)^{-1}, & \text{if } r_{\min} \le r \le r_{\max} \\ 0, & \text{otherwise} \end{cases}, \tag{7.35}$$

Algorithm 3 CPBS – Channel power-based scheduling.

input: The number of users, K, and the channel vectors, $\{\mathbf{a}_k\}_{k=1}^{K}$
output: The set of scheduled users, $\mathcal{K}$

$P_k \leftarrow \|\mathbf{a}_k\|_2^2, \forall k \in \{1,\ldots,K\}$
$\mathcal{K}^{(0)} \leftarrow \emptyset$
$\Omega^{(0)}, n \leftarrow 0$
repeat
  $k^* \leftarrow \arg\max \{P_k \mid k \in \{1,\ldots,K\} \backslash \mathcal{K}^{(n)}\}$
  $\mathcal{K}^{(n+1)} \leftarrow \mathcal{K}^{(n)} \cup \{k^*\}$
  $\Omega^{(n+1)} \leftarrow \Omega^{(n)} + \overline{p}_{k^*}^{\mathrm{SU}}$
  if $\Omega^{(n+1)} \geq P_{\max}$ **then**
    exit loop
  end if
  $n \leftarrow n+1$
until $|\mathcal{K}^{(n)}| = K$
return $\mathcal{K} \leftarrow \mathcal{K}^{(n)}$

where $0 < r_{\min} < r_{\max}$ are respectively the possible minimum and maximum distances from the array center to any user in the communication cell. The BS is equipped with $M = 10^3$ antennas. The minimum achievable rate per user is set to $\overline{R}_k \in [5, 15]$ bps/Hz, $\forall k \in \{1,\ldots,K\}$. We choose the minimum value of 5 bps/Hz to meet the ITU-R experienced data rate requirement of 100 Mbps for the dense urban eMBB scenario [22]. On the matter of the channel model, the path loss attenuation and coefficients are defined according to the ITU-R urban micro-cell environment [23]. The complete list of the simulation parameters is organized in Table 3. The evaluation metrics are calculated by averaging the results obtained from $S = 10^3$ realizations. During each realization, the users' positions and the channels are generated by sampling random distributions following the definitions provided in Section 7.2.

7.5.1 Evaluation Metrics

The metrics used to evaluate the user scheduling techniques are *a*) the achievable sum-rate; *b*) the average achievable rate; and *c*) the number of scheduled users. Moreover, we have defined metrics to analyze the *d*) distribution of the scheduled users across the cell, and *e*) the probability of a user being scheduled given its channel state.

From the achievable rate of the user k defined in eq. (7.15), the system *achievable sum-rate* is calculated by:

$$R = \sum_{k\in\mathcal{K}} \log_2\left(1 + \frac{p_k}{\sigma_w^2[(\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}]_{k,k}}\right). \tag{7.36}$$

Using eq. (7.36), the *average achievable rate* can be expressed by dividing the sum-rate by the number of scheduled users:

$$\overline{R} = \frac{R}{|\mathcal{K}|}. \tag{7.37}$$

Table 3 – Simulation parameters.

Parameter	**Value**
Number of antennas	$M = 10^3$
Carrier frequency	$f_c = 4$ GHz
Rayleigh distance	$r_{\text{Rayl.}} \approx 37.4$ km
System bandwidth	$B = 20$ MHz
Antennas spacing	$d = \lambda/2 = 3.75$ cm
Number of users	$K = 10^3$
Minimum achievable rate	$\overline{R}_k \in [5, 15]$ bps/Hz, $\forall k$
User distance range	$r_k^{\text{UE}} \in [0.03, 1]$ km
User angle range	$\theta_k^{\text{UE}} \in [-\pi, \pi]$
Transmit power budget	$P_{\max} \in [0, 30]$ dBm
LoS probability	$\rho \in \{0, 0.25, 0.75, 1\}$
LoS channel path loss exponent	$\gamma^{\text{LoS}} = 2.20$
NLoS channel path loss exponent	$\gamma^{\text{NLoS}} = 3.67$
LoS channel path loss attenuation	$\beta_0^{\text{LoS}} = 10^{-4.00}$
NLoS channel path loss attenuation	$\beta_0^{\text{NLoS}} = 10^{-3.85}$
Noise power spectral density	-174 dBm/Hz
Admissibility for channel orthogonality	$\epsilon = 0.4$
Number of realizations	$S = 10^3$

Source: The authors.

To measure the *distribution of the scheduled users* across the cell, we determine the complementary cumulative distribution function (CCDF) of the 2D distance between the scheduled users and the array center. The CCDF for a distance $r \geq 0$ is calculated by:

$$\bar{F}(r) = \Pr(k \in \mathcal{K} \mid r_k^{\text{UE}} > r). \tag{7.38}$$

From a numerical perspective, the CCDF in eq. (7.38) can be approximated by deploying the result of a Monte-Carlo simulation as:

$$\widehat{\bar{F}}(r) = \frac{\sum_{\mathcal{K} \in \mathcal{S}} \sum_{k \in \mathcal{K}} I_{\{r' \in \mathbb{R} | r' > r\}}(r_k^{\text{UE}})}{\sum_{\mathcal{K} \in \mathcal{S}} |\mathcal{K}|}, \tag{7.39}$$

where $\mathcal{S} = \{\mathcal{K}_1, \ldots, \mathcal{K}_S\}$ is the set containing all the sets of scheduled users obtained in each of the $S = |\mathcal{S}|$ Monte-Carlo realizations.

A complementary metric to evaluate the distribution of the scheduled users is the probability of a user being scheduled given its channel state. These probabilities w.r.t. a user under the LoS or NLoS channel state are respectively given by:

$$P_{\text{LoS}} = \Pr(k \in \mathcal{K} \mid x_k = 1), \tag{7.40}$$

$$P_{\text{NLoS}} = \Pr(k \in \mathcal{K} \mid x_k = 0). \tag{7.41}$$

Similarly to eq. (7.38), these two probabilities can be estimated from the result of a Monte-Carlo simulation by calculating:

$$\hat{P}_{\mathrm{LoS}} = \frac{\sum_{\mathcal{K}\in\mathcal{S}} \sum_{k\in\mathcal{K}} I_{\{1\}}(x_k)}{\sum_{\mathcal{K}\in\mathcal{S}} |\mathcal{K}|}, \tag{7.42}$$

$$\hat{P}_{\mathrm{NLoS}} = \frac{\sum_{\mathcal{K}\in\mathcal{S}} \sum_{k\in\mathcal{K}} I_{\{0\}}(x_k)}{\sum_{\mathcal{K}\in\mathcal{S}} |\mathcal{K}|}. \tag{7.43}$$

7.5.2 Baseline Techniques

The baseline user scheduling techniques for XL-MIMO systems adopted for comparison with the proposed CBS and CPBS algorithms include the *greedy weighted clique* (GWC) search algorithm [10], the *distance-based scheduling* (DBS), and the *simplified DBS* (DBS-s), both latter proposed in [12].

In GWC, the users in the communication cell are represented by a UWG. The edges are drawn according to the ϵ-orthogonal rule as described in Section 7.4, while the vertices weights are the single-user capacities considering uniform power allocation. The GWC algorithm of [10] implements a greedy algorithm to search the maximum weighted clique in the graph aiming to obtain a set of scheduled users that have simultaneously high channel powers and quasi-orthogonal channel vectors.

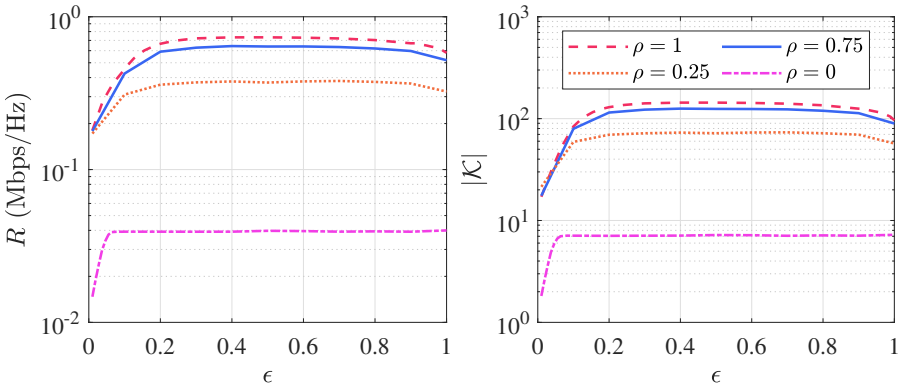
Differently, the DBS algorithm performs user scheduling using a metric named *equivalent distance*, defined in eq. (7) of [12]. The equivalent distance of a given user during an iteration of the DBS algorithm essentially depends on its distance to the center of the BS array and the sum of the inner products between its channel vector and the precoding vectors of the currently scheduled users. Hence, users with lower equivalent distance values have higher scheduling priority. The algorithm proceeds by selecting the users with the lowest equivalent distance until there is a reduction in the achievable sum-rate. Alternatively, also in [12] is proposed the DBS-s algorithm, a version of the DBS with lower computational complexity. In this algorithm, the equivalent distance metric is substituted by the distance between the user and the center of the BS array, reducing the complexity of the algorithm at the cost of performance degradation.

For a fair comparison between the proposed and baseline scheduling techniques, we have included one additional stop criterion on the GWC, DBS, and DBS-s algorithms. During the end of each iteration of the baseline algorithms, the feasibility of the power allocation problem is evaluated by applying eq. (7.20). If this criterion is violated, the last scheduled user is removed from the set and the algorithm stops. After the user scheduling procedure, the power allocation is carried out by calculating the solution of optimization problem $\mathcal{P}_2$.

7.5.3 User Scheduling Performance

Fig. 10 depicts the achievable sum-rate and the number of scheduled users obtained by the proposed CBS algorithm depending on the parameter of admissibility for channel orthogonality. This analysis is paramount to tune the CBS ϵ parameter for the performance comparison carried out in the following. From Fig. 10, one can see that both the sum-rate and the number of scheduled users are almost constant for $\rho = 0$ and $\epsilon > 0.1$. This occurs because, due to the law of large numbers, the NLoS channel state benefits from the favorable propagation offered by the XL-MIMO array. Differently, for $\rho > 0$, the peaks of sum-rate and number of scheduled users are achieved in the range $\epsilon \in [0.3, 0.7]$. In this case, similarly to the result obtained in [11], if ϵ is close to 1, the performance degrades due to the reduction in the effective channel gains paid to obtain the IUI suppression provided by the ZF precoder. On the other hand, if ϵ is close to 0, the multi-user diversity gain decreases. Considering this result, we choose $\epsilon = 0.4$ to generate the remaining numerical results.

Figure 10 – Achievable sum-rate and number of scheduled users of the CBS algorithm *vs.* the admissibility channel orthogonality (ϵ) under scenarios with different LoS probabilities. $M = 10^3$ antennas, $K = 10^3$ users, $P_{\max} = 0$ dBm.

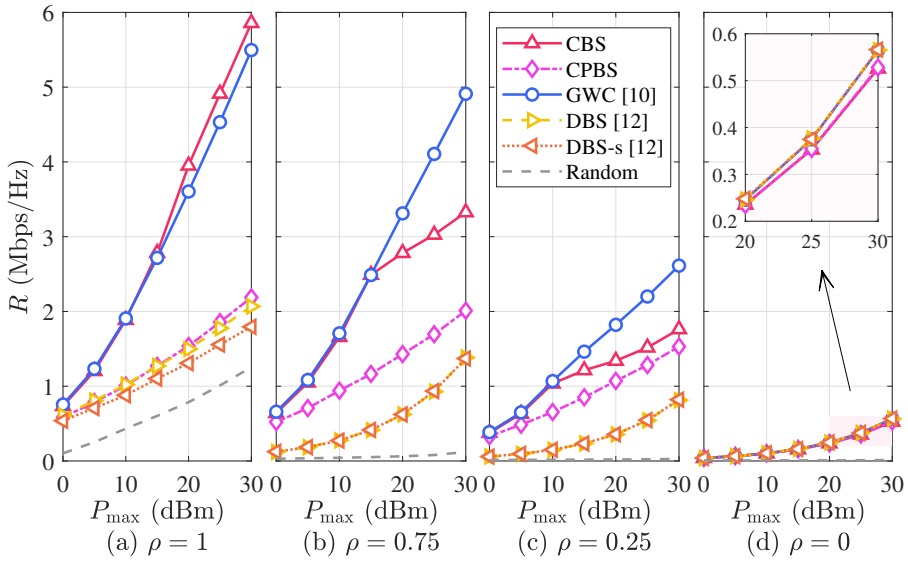


Source: The authors.

Firstly, we evaluate the user scheduling techniques in terms of the **achievable sum-rate**. Fig. 11 depicts the achievable sum-rate obtained by the techniques depending on the transmit power and considering different LoS probability values. At first glance, we see that the LoS probability value changes drastically the performance of the user scheduling algorithms since the users' channel qualities depend directly on the channel state. We note that decreasing the LoS probability reduces the achievable sum-rate obtained by all the algorithms. Moreover, as expected, the sum-rate increases with the transmit power. It is worth noticing that the graph-based techniques achieve the best sum-rate performance among the evaluated ones. The CBS and GWC algorithms have similar performance for low transmit power. However, for $P_{\max} > 15$ dBm and $\rho \leq 0.75$, the GWC algorithm outperforms significantly the CBS one. After the graph-based techniques,

the CBS algorithm is the one that achieves the best performance. The DBS and DBS-s algorithms achieve almost the same performance for all the evaluated cases, outperforming only the random scheduling. Indeed, it is worth mentioning that, except for the random scheduling, all the evaluated techniques achieve similar performance for $\rho = 0$. The random scheduling achieves the poorest performance because its scheduling criterion does not take into account the quality of the users' channel vectors. Such a behavior repeats for all the numerical results in the sequel.

Figure 11 – Achievable sum-rate *vs.* the transmit power under scenarios with different LoS probabilities. Simulation parameters are detailed in Table 3.

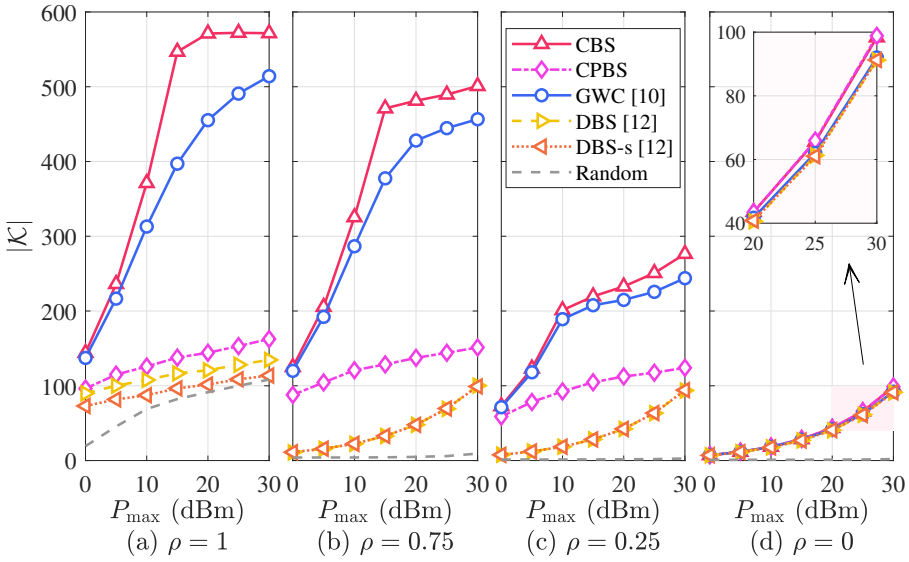


Source: The authors.

Fig. 12 depicts the **number of scheduled users** as a function of the transmit power considering different values of LoS probability. One can see that the number of scheduled users increases with the transmit power and LoS probability. Specifically for $\rho = 0$, all the evaluated techniques, except for the random scheduling, schedule almost the same number of users. For $\rho > 0$, the CBS algorithm achieves the best performance in terms of the number of scheduled users, followed by GWC. It is worth mentioning that, despite the GWC attaining a higher sum-rate than the CBS for $P_{\max} > 15$ dBm and $\rho \in (0, 0.75]$, the CBS consistently schedules a higher number of users. Finally, similarly to what occurs with the sum-rate metric, the number of scheduled users achieved by all the scheduling techniques treated herein, except for the random scheduling, are nearly the same for $\rho = 0$.

Fig. 13 depicts the **average achievable rate** depending on the transmit power considering different values of LoS probability. For all the evaluated techniques, except for the random scheduling, the average rate is inversely proportional to the number of scheduled users. Specifically, for $\rho < 1$, the CBS algorithm obtains achievable rate values close to the minimum achievable rate of 5 bps/Hz. For $\rho \geq 0.75$, the best techniques in terms of the average rate are the DBS and DBS-s. On the other hand, for $\rho \leq 0.25$, the

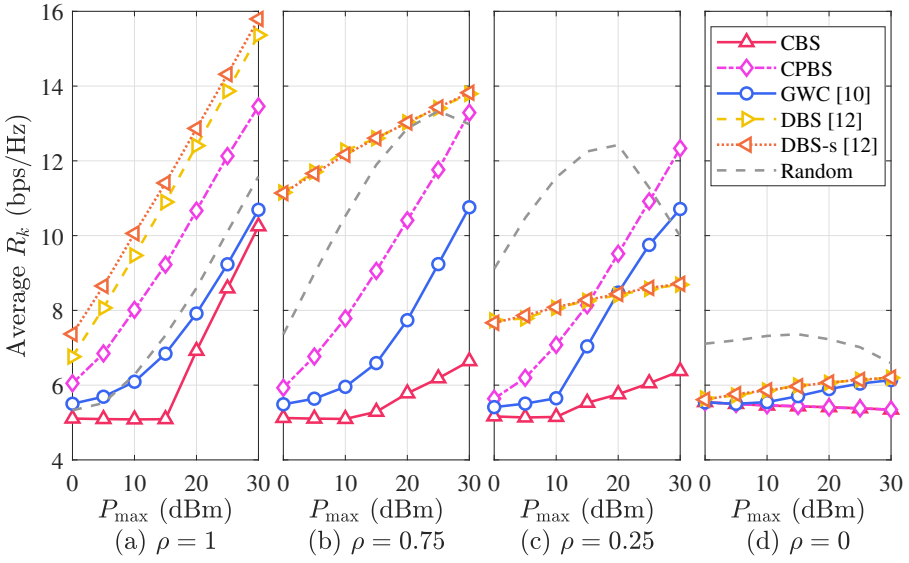
Figure 12 – Number of scheduled users *vs.* the transmit power under scenarios with different LoS probabilities. Simulation parameters are detailed in Table 3.



Source: The authors.

random scheduling achieves the best performance in terms of average rate. However, it is important to mention that this high average rate is obtained at the cost of scheduling an extremely low number of users, as demonstrated in Fig. 12.

Figure 13 – Average rate of the scheduled users *vs.* the transmit power under scenarios with different LoS probabilities. Simulation parameters are detailed in Table 3.

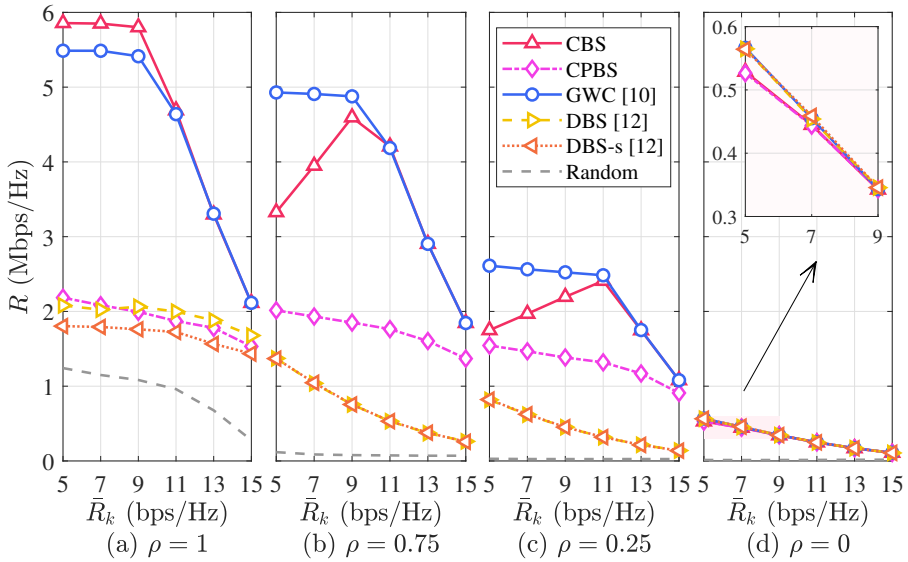


Source: The authors.

Fig. 14 depicts the achievable sum-rate as a function of the minimum achievable rate considering different values of LoS probability. For all the techniques, except for the CBS, the achievable sum-rate decreases by increasing the minimum achievable rate. This behavior is expected since increasing the minimum achievable rate constraint implies allocating

more power per user to satisfy this requirement. Especially, the GWC for $\rho \in \{0.25, 0.75\}$ reaches achievable sum-rate values that increase with the minimum achievable rate, up to a point where this behavior reverses. As we will see in the result in the sequel, this maximum point of achievable sum-rate occurs due to a reduction in the number of scheduled users slower than the increase in the minimum achievable rate.

Figure 14 – Achievable sum-rate *vs.* minimum achievable rate per user under scenarios with different LoS probabilities. $P_{\max} = 30$ dBm.



Source: The authors.

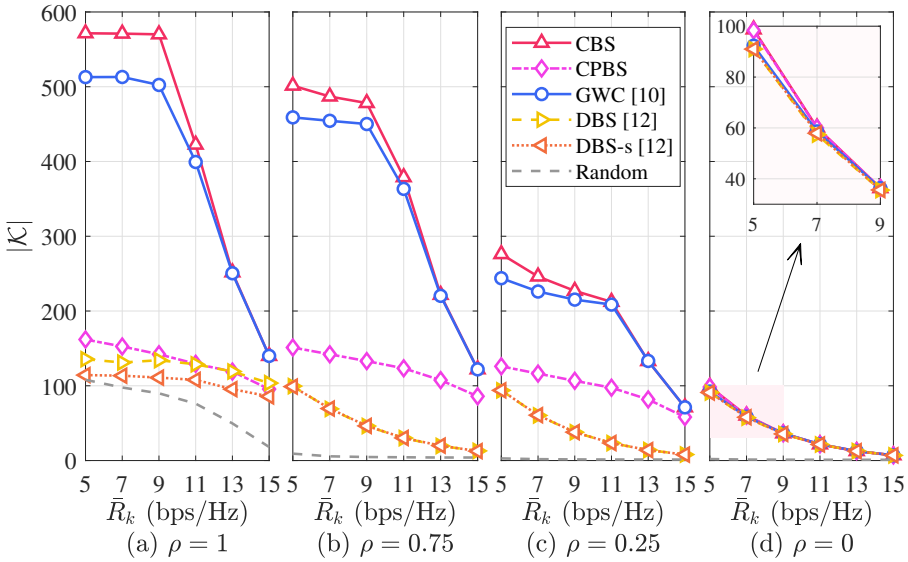
Fig. 15 depicts the number of scheduled users depending on the minimum achievable rate considering different values of LoS probability. As expected, the stricter minimum achievable rate constraints with a fixed transmit power budget results in a reduction on the number of scheduled users. Specifically, we see that the CBS and GWC algorithms present a slow rate of decrease in the number of scheduled users for $\rho = 0.75$ and $\overline{R}_k \leq 9$ bps/Hz, $\forall k$, and for $\rho = 0.25$ and $\overline{R}_k \leq 11$ bps/Hz, $\forall k$. This is the cause of the partially increasing behavior of the achievable sum-rate obtained by these algorithms identified in Figs. 14 for $\rho \in \{0.25, 0.75\}$.

7.5.4 User Scheduling Performance with Inaccurate CSI

Since the proposed and benchmark scheduling algorithms rely on CSI, it is important to evaluate their performance in the case when inaccurately estimated channel vectors are available at the BS. Let $\hat{\mathbf{a}}_k \in \mathbb{C}^M$ denotes the estimated channel vectors that follow the model below:

$$\hat{\mathbf{a}}_k = \sqrt{1-\alpha^2}\mathbf{a}_k + \alpha\boldsymbol{\nu}_k, \forall k \in \{1, \ldots, K\}, \tag{7.44}$$

Figure 15 – Number of scheduled users *vs.* minimum achievable rate under scenarios with different LoS probabilities. $P_{\max} = 30$ dBm.



Source: The authors.

where $0 \leq \alpha < 1$, and $\boldsymbol{\nu}_k \sim \mathcal{CN}(\mathbf{0}_M, \|\mathbf{a}_k\|_2^2 \mathbf{I}_M)$ denotes the estimation error. Therefore, the NMSE of the estimate is:

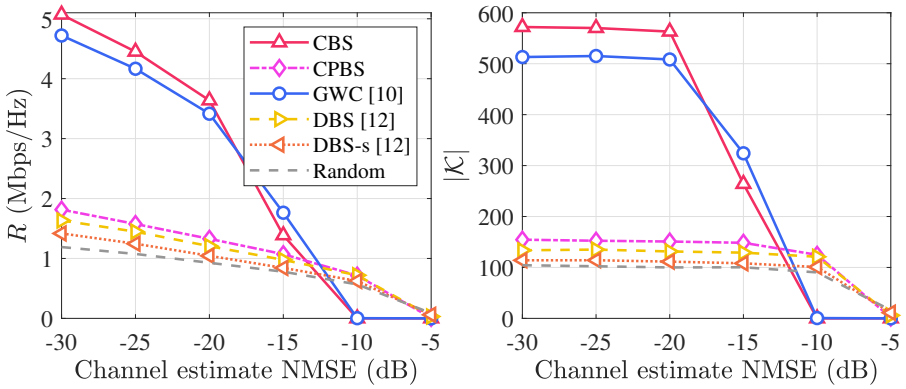
$$\mathbb{E}\left\{\frac{\|\hat{\mathbf{a}}_k - \mathbf{a}_k\|_2^2}{\|\mathbf{a}_k\|_2^2}\right\} = 2(1 - \sqrt{1-\alpha^2}), \tag{7.45}$$

where the expectation is taken over $\boldsymbol{\nu}_k$. In this sense, Fig. 16 depicts the achievable sum-rate and the number of scheduled users *vs.* the normalized mean-squared error (NMSE) of the estimated channel vectors. The NMSE values are in accordance with recent results for channel estimation in XL-MIMO systems [9]. Analyzing Fig. 16, one can notice that the achievable sum-rate decreases with the channel estimate NMSE, once the mismatch between the channel vectors and their estimates produces IUI, preventing the users from reaching their required QoS levels. Moreover, the channel estimation error has a significant impact on the number of scheduled users by the CBS and GWC algorithms when the NMSE is above -20 dB, and when it is above -10 dB for the others. In the worst case, the methods cannot carry out user scheduling when the NMSE is greater or equal to -10 dB for the graph-based techniques and -5 dB for the others. Therefore, even though the CBS algorithm is affected by inaccurate CSI as the other scheduling techniques, it can achieve remarkable performance in crowded XL-MIMO scenarios, scheduling around 560 users with a channel estimate NMSE of -20 dB.

7.5.5 Computational Complexity

Table 4 shows the **running time** of the evaluated scheduling algorithms for $\rho \in \{0, 0.25, 0.75, 1\}$. With this result, the performance-complexity trade-off obtained by the

Figure 16 – Achievable sum-rate and number of scheduled users *vs.* channel estimate NMSE. $\rho = 1$, $P_{\max} = 30$ dBm, $\overline{R}_k = 5$ bps/Hz.



Source: The authors.

algorithms can be compared by the ratio between the running time and the number of scheduled users (Fig. 12). For $\rho = 1$, the CBS algorithm has the best trade-off, achieving the highest number of scheduled users at the expense of a running time 3.7 times lower than DBS and 79 times lower than GWC. On the other hand, for $\rho = 0$, the CPBS algorithm achieves the best trade-off, since all the evaluated techniques schedule approximately the same number of users, except for the random scheduling. For $\rho \in \{0.25, 0.75\}$, the DBS-s algorithm achieves the best trade-off. Whereas, it is worth noticing that the CBS algorithm can schedule up to 5 times more users than DBS-s with a similar trade-off. Therefore, the benefits of the proposed CBS algorithm are achieved in realistic scenarios where users under the LoS and NLoS channel states coexist, *i.e.*, $\rho > 0$.

Table 4 – Running time (s) of the scheduling algorithms.

ρ	**CBS**	**CPBS**	**GWC** [10]	**DBS** [12]	**DBS-s** [12]	**Random**
1.00	**0.43**	0.41	32.40	1.61	0.26	0.15
0.75	1.19	0.37	20.88	0.60	**0.16**	$1 \cdot 10^{-3}$
0.25	0.50	0.23	2.94	0.51	**0.14**	$4 \cdot 10^{-4}$
0.00	0.26	**0.12**	0.40	0.46	0.12	$3 \cdot 10^{-4}$

$P_{\max} = 30$ dBm, $\overline{R}_k = 5$ bps/Hz

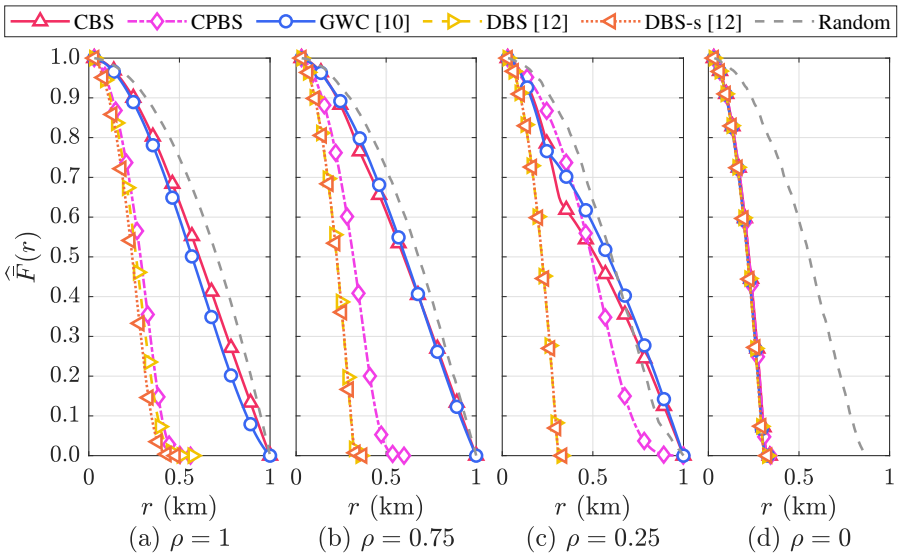
Source: The authors.

7.5.6 Distribution of the Scheduled Users

Fig. 17 depicts the CCDF of the 2D distance between the scheduled users and the array center considering different values of LoS probability. For $\rho \leq 0.75$, both graph-based user scheduling techniques, our proposed CBS and the GWC provide a much higher cell coverage when compared with the DBS and DBS-s algorithms, allowing scheduling users located at cell border ($0.8 < r_k^{\text{UE}} \leq 1.0$ km). In fact, the distance-based techniques can

schedule only users that are at the maximum distance of 0.5 km apart from the array, half the cell radius. The scheduled users' distribution in Fig. 17 demonstrates the superiority of the proposed CBS algorithm in providing a more uniform communication experience for users positioned throughout the communication cell, including border users. On the other hand, for $\rho = 0$, one can see that, except for the random scheduling, all the evaluated techniques achieve the same poor coverage, scheduling users only around 0.3 km apart from the array center. Specifically, this result occurs due to the high path loss associated with the NLoS channel state and the limited transmit power. Particularly, we note that the random scheduling attains the best coverage in all the evaluated LoS channel probability cases. However, such good coverage comes at the price of low performance in terms of achievable sum-rate and number of scheduled users, as demonstrated in Figs. 11 and 12.

Figure 17 – CCDF of the 2D distance between the scheduled users and the array center under scenarios with different LoS probabilities. $P_{\max} = 30$ dBm, $\overline{R}_k = 5$ bps/Hz.

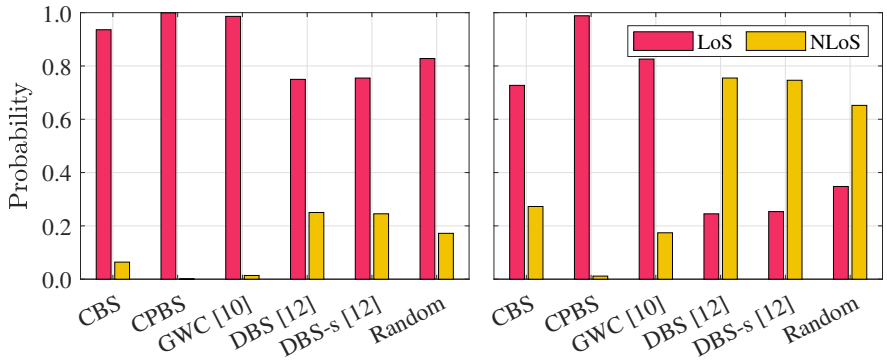


Source: The authors.

Fig. 18 depicts the probability of scheduling users under LoS and NLoS channel states considering two different values of LoS probability. The techniques that consider the minimum required power to attain the minimum achievable rate and the channel quality tend to schedule users in the most favorable channel state, namely the LoS state. On the other hand, the techniques that consider the distance tend to schedule users in the same proportion as that they appear in the communication cell.

From the numerical results presented in Figs. 11-18 and Table 4 one can conclude that the CBS algorithm achieves high numbers of scheduled users and provides fair coverage with good performance-complexity trade-off when $0 < \rho \leq 1$. This includes the cases when users under both LoS and NLoS channel states coexist in the communication cell. On the matter of the CPBS algorithm, it demonstrates a better trade-off than DBS in all

Figure 18 – Probability of a user being scheduled given its channel state under scenarios with different LoS probabilities. $P_{\max} = 30$ dBm, $\overline{R}_k = 5$ bps/Hz.



(a) $\rho = 0.75$ (b) $\rho = 0.25$

Source: The authors.

the evaluated scenarios and improved coverage for $\rho = 0.25$. As a result, in the context of XL-MIMO systems operating in crowded scenarios, the CBS algorithm is a promising technique to schedule high numbers of users ensuring the QoS levels constraints, while providing fair coverage over the whole cell area.

7.6 Conclusions

In this paper, we study the problem of user scheduling in crowded XL-MIMO systems with per-user QoS requirements. To increase the number of served users in the DL channel of XL-MIMO systems with high user density, we propose the CBS algorithm, a QoS-aware joint user scheduling and power allocation algorithm based on the search in a graph. Moreover, to capture the complexity of the propagation environment with physically large arrays, we propose a non-stationary multi-state channel model that accounts for the co-existence of users under LoS and NLoS transmission in the same communication cell. The user scheduling performance of the developed CBS algorithm is evaluated and compared with the state-of-the-art, considering the proposed channel model under different conditions. As a result, the CBS algorithm demonstrates superiority over state-of-the-art techniques in crowded XL-MIMO scenarios, scheduling up to 5 times more users with satisfied QoS requirements than the DBS algorithm [12] when users under both LoS and NLoS channel states coexist ($0 < \rho < 1$). Still, this improved performance is reached even when inaccurate CSI is available at the BS. Besides, the CBS algorithm achieves fair coverage over the whole cell area, providing a more uniform communication experience for all users, including the ones located at the cell border. Finally, the computational complexity analysis demonstrates that the CBS algorithm presents a good performance-complexity trade-off, scheduling more users than the graph-based GWC algorithm [10] in a running time up to 79 times lower. Future research directions include incorporating

strategies to schedule the non-scheduled users on subsequent frames, improving the users' average achievable rates, and thereby promoting fairness among them.

7.7 References

[1] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[2] A. Amiri, C. N. Manchón, and E. de Carvalho, "A message passing based receiver for extra-large scale MIMO," in *2019 IEEE 8th CAMSAP*, 15-18 Dec. 2019, pp. 564–568.

[3] H. Lu and Y. Zeng, "How does performance scale with antenna number for extremely large-scale MIMO?" in *ICC 2021 - IEEE International Conference on Communications*, 14-23 June 2021, pp. 1–6.

[4] J. Liu, Y. Ma, J. Wang, N. Yi, R. Tafazolli, S. Xue, and F. Wang, "A non-stationary channel model with correlated NLoS/LoS states for ELAA-mMIMO," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 7-11 Dec. 2021, pp. 1–6.

[5] E. Castañeda, A. Silva, A. Gameiro, and M. Kountouris, "An overview on resource allocation techniques for multi-user MIMO systems," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 239–284, Oct. 2017.

[6] J. C. Marinello, T. Abrão, A. Amiri, E. de Carvalho, and P. Popovski, "Antenna selection for improving energy efficiency in XL-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 305–13 318, Sep. 2020.

[7] J. H. I. de Souza, A. Amiri, T. Abrão, E. de Carvalho, and P. Popovski, "Quasi-distributed antenna selection for spectral efficiency maximization in subarray switching XL-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6713–6725, May 2021.

[8] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 633–637, May 2020.

[9] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2663–2677, Apr. 2022.

[10] T. Yoo and A. Goldsmith, "Sum-rate optimal multi-antenna downlink beamforming strategy based on clique search," in *GLOBECOM '05. IEEE Global Telecommunications Conference, 2005.*, vol. 3, 28 Nov.-2 Dec. 2005, pp. 1510–1514.

[11] ——, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[12] J. P. González-Coma, F. J. López-Martínez, and L. Castedo, "Low-complexity distance-based scheduling for multi-user XL-MIMO systems," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2407–2411, Aug. 2021.

[13] J. C. M. Filho, G. Brante, R. D. Souza, and T. Abrão, "Exploring the non-overlapping visibility regions in XL-MIMO random access and scheduling," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6597–6610, Aug. 2022.

[14] Y. Ming, Z. Sha, Y. Dong, and Z. Wang, "Downlink resource allocation with pilot length optimization for user-centric cell-free MIMO networks," *IEEE Communications Letters*, pp. 1–5, Aug. 2022, early access.

[15] N. Qu, B. Li, R. Shafin, M. Liu, F. Gong, and L. Liu, "Angle-based downlink beam selection and user scheduling for massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 10, pp. 8077–8089, Oct. 2022.

[16] Z. Zhang, T. Jiang, and W. Yu, "Learning based user scheduling in reconfigurable intelligent surface assisted multiuser downlink," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 1026–1039, Aug. 2022.

[17] L. D. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Energy-efficient multi-cell massive MIMO subject to minimum user-rate constraints," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 914–928, Feb. 2021.

[18] D. Palomar and J. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 686–695, Feb. 2005.

[19] H. Lu and Y. Zeng, "Near-field modeling and performance analysis for multi-user extremely large-scale MIMO communication," *IEEE Communications Letters*, vol. 26, no. 2, pp. 277–281, Feb. 2022.

[20] J. L. Gross, J. Yellen, and P. Zhang, *Handbook of Graph Theory*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2014.

[21] J. C. Marinello and T. Abrão, "Collision resolution protocol via soft decision retransmission criterion," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4094–4097, Apr. 2019.

[22] *Minimum requirements related to technical performance for IMT-2020 radio interface(s)*, International Telecommunication Union, Nov. 2017, Rep. ITU-R M.2410-0.

[23] *Guidelines for evaluation of radio interface technologies for IMT-2020*, International Telecommunication Union, Oct. 2017, Rep. ITU-R M.2412-0.

8 Paper B: Energy Efficiency and Throughput of Random Access Protocols for RIS-Aided IoT Networks

João Henrique Inacio de Souza, José Carlos Marinello Filho, Taufik Abrão and Cristiano Panazio

This paper has been published in the *proceedings of the 2022 IEEE 8th World Forum on Internet of Things (WF-IoT),* Yokohama, Japan, 26 Oct.-11 Nov. 2022, pp. 1-6, doi: https://doi.org/10.1109/WF-IoT54382.2022.10152283.

Abstract

Green Internet of Things (IoT) aims to enable a sustainable smart world by making energy efficiency (EE) the main performance indicator for IoT hardware and software. With respect to network design, this implies in developing energy-efficient communication protocols and network architectures adapted to the ubiquitousness of the IoT machine-type devices (MTDs) and the sporadic traffic generated by them, keeping a low power consumption at the MTDs-side. In this sense, reconfigurable intelligent surfaces (RISs) have presented the capacity of significantly improving the network coverage using mostly passive reflecting elements, drastically reducing the power expenditure. In this paper, we develop a realistic power consumption model and an expression for the overall system EE for RIS-aided IoT networks that adopt a two time-scale random access (RA) protocol to handle the uplink transmissions. Specifically, during each time slot of the RA protocol, the RIS covers a specific area of interest in the communication cell with a predefined set of phase-shift configurations, changing the channel qualities of the contending MTDs. Numerical results comparing the RA protocol performance reveal that access policies that exploit information of the channel qualities are suitable for green IoT networks, simultaneously attaining competitive EE and throughput combined with low power consumption at the MTDs-side.

Index Terms: Green Internet of Things (IoT); random access protocols; reconfigurable intelligent surface (RIS); energy efficiency (EE).

8.1 Introduction

The concept of green IoT brings the challenge of developing a sustainable IoT environment by paths such as designing energy-efficient systems and devices, and prioritizing the use of renewable power sources [1]. From the communication protocol standpoint, this objective leads to optimized schemes that make an efficient use of radio resources, including transmit power and time-frequency resource blocks. In this sense, [2] has recently proposed a framework to improve the devices' EE by optimizing the transmit power and the maximum number of packet replicas in the *irregular repetition slotted* ALOHA (IRSA) protocol. Although optimized protocols are able to offer increased EE, combining these designs with energy-efficient technologies can further enhance these gains.

RIS is a promising low-cost and low-power technology to enable sustainable wireless networks. This technology can produce anomalous reflection of the incoming signal to create hot spots in areas of interest in the communication cell without any power amplifier (PA), consuming much less energy than common active transceiver technologies, such as arrays with massive numbers of antennas. Specifically, in the analysis of [3], it is shown that RIS-aided networks can achieve up to a threefold increase in EE when compared with networks assisted by multiple-antenna amplify-and-forward relays.

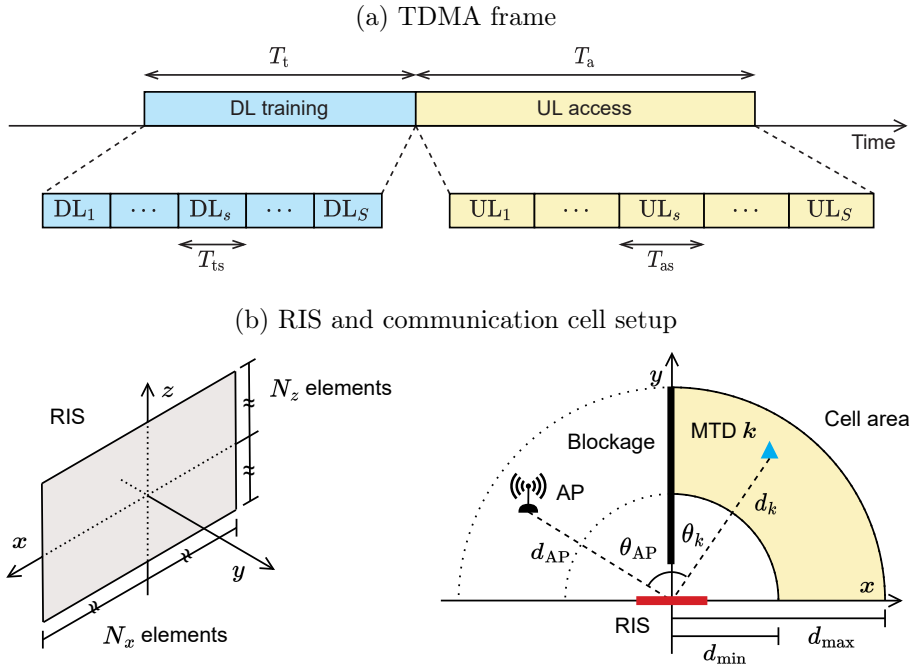
Recently, [4] has proposed a two time-scale RA protocol for RIS-aided networks in which the RIS has a limited set of phase-shift configurations, designed to cover specific areas of interest in the communication cell. During the first phase of the protocol, the contending MTDs measure their channel qualities over a sequence of training time slots, where in each slot the RIS induces a different phase-shift configuration. Then, at the second phase, the RIS repeats the configuration change pattern over a sequence of access time slots, where in each slot the MTDs independently decide to transmit according to the measured channel qualities. This protocol increases the average number of *successful access attempts* at the expense of increased access delay, which is a minor issue in delay-tolerant IoT applications.

In this paper, we develop a realistic power consumption model and the throughput formula for RIS-aided IoT networks that adopt the RA protocol proposed in [4]. With this model, we derive an expression for the overall system EE, evaluating the protocol performance with four different access policies. Moreover, we extend the *strongest-configuration access policy* (SCP) given in [4] to fully exploit the successive interference cancellation (SIC) receiver implemented at the access point (AP)-side. Finally, we adapt the *contention resolution diversity slotted ALOHA* (CRDSA) [5] and the state-of-the-art IRSA [6] protocols to create two novel access policies suitable for IoT networks.

8.2 System Model

In this section, we describe the time-division multiple access (TDMA) frame, the RIS-aided communication scheme, and the two time-scale RA protocol. The TDMA frame, depicted in Fig. 19a, is comprised by a downlink (DL) training block of duration $T_\mathrm{t} > 0$ and an uplink (UL) access block of duration $T_\mathrm{a} > 0$. The training and access blocks are divided into $S \in \mathbb{Z}_+$ time slots with lengths respectively equal to $T_\mathrm{ts} > 0$ and $T_\mathrm{as} > 0$, *i.e.*, $T_\mathrm{t} = ST_\mathrm{ts}$ and $T_\mathrm{a} = ST_\mathrm{as}$. The two blocks of the TDMA frame are used by the RA protocol to handle the UL transmissions of the contending MTDs, namely the MTDs that want to transmit a data burst during an RA frame.

Figure 19 – (a) Diagram of the TDMA frame, and (b) setup of the RIS-aided communication scheme.



Source: The authors.

Fig. 19b depicts the RIS and communication cell setup. The RIS is a thin surface located at the xz-plane with center at the origin. The surface is constituted by $N_x \in \mathbb{Z}_+$ reflecting elements in the x-axis direction and $N_z \in \mathbb{Z}_+$ in the z-axis direction, totaling $N = N_x N_z$ elements. Each reflecting element has a rectangular surface of sides $d_x, d_z \leq \lambda$, where λ denotes the carrier wavelength. Similarly to [4], we assume that the signal reflection produced by the RIS is independent of the z-dimension. We consider that the RIS has a fixed number of phase-shift configurations equal to the number of time slots, S. Specifically, the set of phase-shift configurations of the RIS as a function of the number of time slots is denoted by:

$$\Theta(S) = \left\{ \frac{\pi(s-1)}{2(S-1)} \middle| \, s \in \{1, \ldots, S\} \right\}. \tag{8.1}$$

Therefore, the s-th phase-shift configuration induced by the RIS in the impinging signal is denoted by $\theta_s \in \Theta(S)$. Since in the RA protocol we assume that the RIS has a dedicated phase-shift configuration for each time slot, throughout the text we use s to denote the s-th RIS configuration and the s-th time slot interchangeably.

In the proposed setup, the AP and MTDs are in the xy-plane, located respectively in the fourth and first quadrants. Specifically, the AP is located at a distance $d_{\text{AP}} \in [d_{\min}, d_{\max}]$ from the origin, with $0 < d_{\min} < d_{\max}$, forming an angle $\theta_{\text{AP}} \in [0, \pi/2]$ with the RIS boresight. Similarly, the MTD k is located at a distance $d_k \in [d_{\min}, d_{\max}]$ from the origin, forming an angle $\theta_k \in [0, \pi/2]$ with the RIS boresight. The contending MTDs communicate with the AP via a virtual line-of-sight path created by the RIS, since the first and fourth quadrants are separated by a blockage that prevents direct radio links between them.

Let $\alpha_{\text{AP}} > 0$ and $\alpha_k > 0$ be the AP and the k-th MTD antenna gains, respectively. The channel coefficient of the MTD k during the access time slot s, denoted by $h_k(s) \in \mathbb{C}$, is equal to [7]:

$$h_k(s) = \sqrt{\beta_k} e^{j\psi_k} \Omega_k(s), \tag{8.2}$$

where $\beta_k \in \mathbb{R}$ is the path loss, $\psi_k \in \mathbb{R}$ is the total phase-shift, and $\Omega_k(s) \in \mathbb{C}$ is the RIS array factor [4], respectively defined by:

$$\beta_k = \frac{\alpha_{\text{AP}} \alpha_k}{(4\pi)^2} \left(\frac{d_x d_z}{d_{\text{AP}} d_k} \right)^2 \cos^2 \theta_k, \tag{8.3}$$

$$\psi_k = \omega \left(d_{\text{AP}} + d_k - (\sin\theta_{\text{AP}} - \sin\theta_k) \frac{N_x + 1}{2} d_x \right), \tag{8.4}$$

$$\Omega_k(s) = N_z \sum_{n=1}^{N_x} e^{j\omega(\sin\theta_k - \sin\theta_s) n d_x}, \tag{8.5}$$

where $\omega = 2\pi/\lambda$ denotes the wavenumber.

Let $\mathcal{K}$ be the set with the indices of the contending MTDs during an RA frame, and such that its cardinality is $|\mathcal{K}| = K$ with $K \in \mathbb{Z}_+$. In addition, let $\mathcal{K}(s) \subseteq \mathcal{K}$ be the set with the indices of the contending MTDs that transmit during the access time slot s. Therefore, the UL received signal at the AP during the access time slot s is equal to:

$$\mathbf{y}(s) = \sum_{k \in \mathcal{K}(s)} \sqrt{\rho_{\text{MTD}}} h_k(s) \mathbf{v}_k + \mathbf{w}(s), \tag{8.6}$$

where ρ_{MTD} is the MTD transmit power, $\mathbf{v}_k \in \mathbb{C}^L$ such that $\|\mathbf{v}_k\|_2^2 = 1$ is the data burst of the MTD k with $L \in \mathbb{Z}_+$ symbols, and $\mathbf{w}(s) \sim \mathcal{CN}(\mathbf{0}_L, \sigma_w^2 \mathbf{I}_L)$ is the additive noise. From eq. (8.6), one can derive the signal-to-noise ratio (SNR) of the MTD k during the access time slot s as:

$$\gamma_k(s) = \frac{\rho_{\text{MTD}} |h_k(s)|^2}{\sigma_w^2}. \tag{8.7}$$

During the DL training, the MTDs measure their respective channel qualities within each time slot by indirectly estimating their SNRs, $\{\gamma_k(s) \mid s \in \{1, \ldots, S\}\}$ for each $k \in \mathcal{K}$. For the sake of simplicity but without loss of generality, we consider that the channel qualities

measured by the MTDs can be defined as:

$$q_k(s) = c_k \gamma_k(s) + \varepsilon_k(s) \tag{8.8}$$

for $k \in \mathcal{K}$, and $s \in \{1, \ldots, S\}$, where $c_k \in \mathbb{R}$ is a constant to account for the DL channel quality, and $\varepsilon_k(s) \in \mathbb{R}$ is the estimation error such that $\mathbb{E}\{\varepsilon_k(s)\} = 0$ and $\mathrm{Var}\{\varepsilon_k(s)\} = \sigma^2_{\varepsilon_k}$. Further details on the measurement of the channel qualities during the DL training can be found in [4].

8.2.1 Random Access Protocol

In this subsection, we describe the *two phases* that constitute the RA protocol, the DL training phase and the UL access phase.

Phase I: Downlink training

1. AP transmits a training pilot over S consecutive time slots, varying the RIS phase-shift configuration to one of the S available ones during each slot.

2. The contending MTDs listen to the channel during each time slot, estimating and buffering their instantaneous channel qualities as defined in eq. (8.8) for each RIS phase-shift configuration.

Phase II: Uplink access

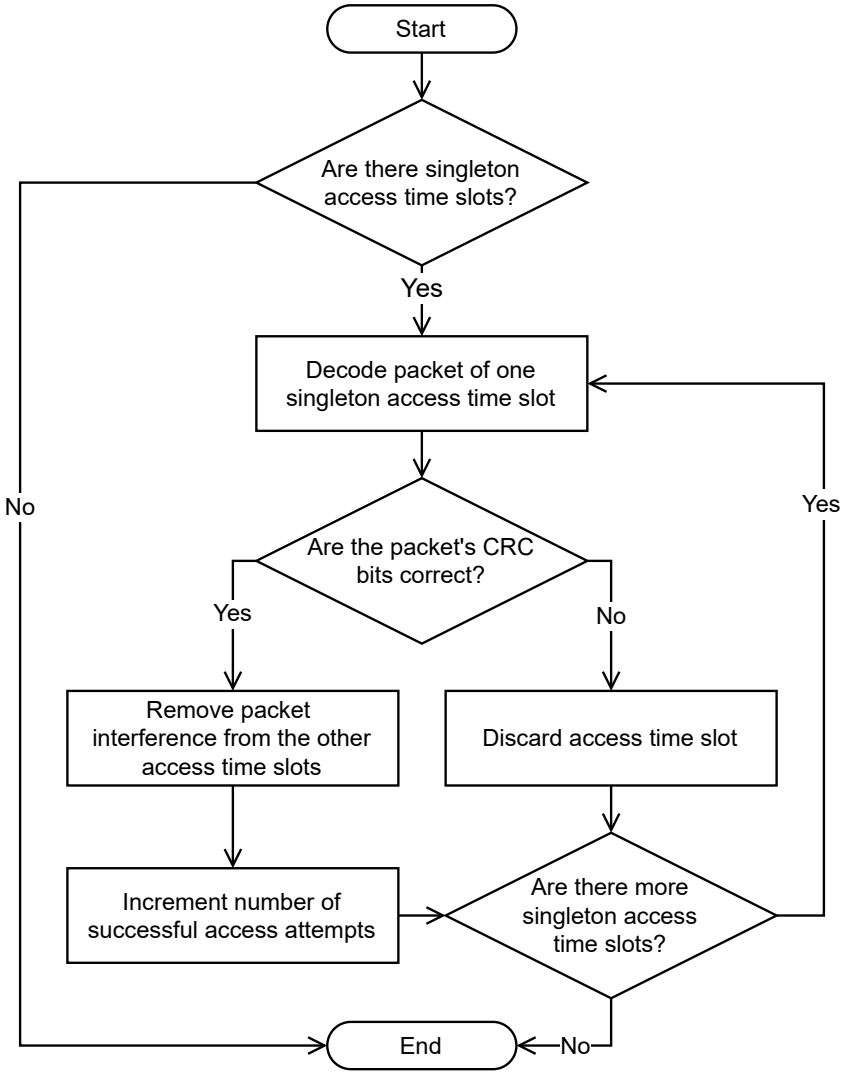
1. AP listens to the channel over S consecutive time slots, using the same change pattern of the RIS phase-shift configuration adopted in the training phase.

2. The contending MTDs send packet replicas during a selected set of time slots, chosen independently by each MTD following a predefined access policy. Specifically, four different policies, described in Subsection 8.2.2, are evaluated in this work. Notice that the packet replicas have cyclic redundancy check (CRC) bits appended to enable detection of decoding errors.

3. AP receiver starts by decoding the *singleton time slots*, namely the time slots in which a single packet was transmitted, or, later in the receiver operation, the time slots that have only one remaining packet after interference cancellation (IC). In practice, preamble recognition can be applied to detect the singleton time slots [5]. After that, a collision resolution procedure based on SIC is carried out to try to decode the colliding packets. The operation of the AP SIC receiver is illustrated in Fig. 20. One should note that a packet is successfully decoded if the CRC bits are correct. In this sense, it is considered that the AP receiver can decode a packet if it is within a singleton time slot and its SNR is above the threshold $\gamma_{\mathrm{th}} \in \mathbb{R}$ in dB.

Therefore, the collision resolution between n colliding packets is possible when $n-1$ of those packets are decoded in prior iterations.

Figure 20 – Flowchart of the SIC receiver deployed in the AP to decode the signals transmitted by the contending MTDs during the access phase.



Source: The authors.

8.2.2 Access Policies

Four different access policies have been deployed in the UL access phase of the analyzed RA protocol. Specifically, the *access policy* determines the *access time slots* in which a contending MTD will transmit. Hence, carefully designing the access policies is crucial to enhance the protocol performance by jointly exploiting the spatial diversity among the contending MTDs and the time diversity created by switching between the different RIS configurations.

Configuration-Aware Random Access Policy (CARP)

In CARP [4], a contending MTD k transmits during each access time slot s with the probability:

$$p_k(s) = \frac{q_k(s)}{\sum_{s=1}^{S} q_k(s)}, \tag{8.9}$$

where the channel qualities $q_k(s)$ for $k \in \mathcal{K}$ and $s \in \{1, \ldots, S\}$ are defined in (8.8). With these probabilities, the contending MTDs are more likely to transmit within the access time slots that provided the best channel qualities during the DL training phase. In CARP, the decision of transmitting during a time slot is made from Bernoulli trials following the probabilities defined by (8.9). If no time slot is selected, the contending MTD will transmit during the time slot that provided the best channel quality. Therefore, each contending MTD can transmit a single packet or up to S packet replicas.

s-Strongest-Configurations Access Policy (s-SCP)

This policy is a generalization of the SCP, originally introduced in [4], to fully exploit the benefits of the SIC procedure at the AP's receiver. In s-SCP, a contending MTD selects the $s \in \mathbb{Z}_+$ access time slots that provided the best channel qualities during the DL training phase. Hence, in this access policy, each contending MTD transmits s packet replicas, totaling sK packet replicas during the UL access phase. It is straightforward to notice that s-SCP is equivalent to the original SCP if $s = 1$, and SIC cannot be applied. Conversely, SIC can be conveniently applied for $s \geq 2$.

Contention Resolution Diversity Slotted ALOHA Access Policy (CRDSAP)

This policy is an adaption of the CRDSA protocol originally proposed in [5] as an extension of the *slotted ALOHA* with support to IC techniques. In CRDSAP, a contending MTD transmits during $s = 2$ access time slots chosen randomly with the same probability, totaling $2K$ transmitted packet replicas during the UL access phase. As the access time slots are chosen randomly without any information about the channel qualities, CRDSAP does not require the DL training phase of the RA protocol. The essential difference between the CRDSAP and the CRDSA protocol is that, during each time slot of the former, the contending MTDs experience different channel conditions due to the changing RIS phase-shift configuration. This creates spatial and time diversities that, as it will be demonstrated further, can increase the system EE and throughput.

Irregular Repetition Slotted ALOHA Access Policy (IRSAP)

This policy is an adaption of the IRSA protocol proposed in [6], introducing variable packet repetition rates to increase the throughput attained by CRDSA. In IRSAP, a contending MTD transmits s packet replicas following a statistical distribution defined by

the probability generating function [2]:

$$\Lambda(x) = \left(1 + \frac{1}{S-1}\right) \sum_{s=2}^{S} \frac{x^s}{(s-1)s}. \tag{8.10}$$

The s access time slots selected to transmit the replicas are chosen with the same probability. Since the access time slots are chosen randomly without any information about the channel qualities, IRSAP does not require the DL training phase. The difference between IRSAP and the original IRSA protocol is the spatial and time diversities created within the UL access phase by changing the RIS configuration between the time slots. In this policy, each MTD can transmit 2 up to S packet replicas, resulting in the average total number of transmitted packet replicas:

$$K\left(1 + \frac{1}{S-1}\right) \sum_{s=2}^{S} \frac{1}{s-1}, \tag{8.11}$$

which can be obtained straightforwardly from eq. (8.10).

8.2.3 System Power Consumption Model

In this subsection, we present the power consumption model, an extension of the model proposed in [3] adapted to account for the power consumed during the RA protocol phases.

Let $\bar{P}_{\mathrm{AP}} > 0$ be the static power dissipated by the AP hardware, $\rho_{\mathrm{AP}} > 0$ be the AP transmit power during each training time slot, and $\xi_{\mathrm{AP}} > 1$ be the inverse of the AP transmit PA efficiency. The power consumed by the AP is equal to:

$$P_{\mathrm{AP}} = S\xi_{\mathrm{AP}}\rho_{\mathrm{AP}} + \bar{P}_{\mathrm{AP}}. \tag{8.12}$$

To derive the power consumption of the RIS, we consider that each of the N reflecting elements is constituted by a phase-shifter with limited resolution [3]. Let $P_n(b) > 0$ be the power consumption of a phase-shifter having b-bit resolution. The power consumed by the RIS is equal to:

$$P_{\mathrm{RIS}} = NP_n(b). \tag{8.13}$$

Finally, let $\bar{P}_{\mathrm{MTD}} > 0$ be the static power consumed by the hardware of each MTD. In addition, let $s_k \in \{1, \ldots, S\}$ be the number of packet replicas sent by the contending MTD k during an RA frame, and $\xi_{\mathrm{MTD}} > 1$ be the inverse of the transmit PA efficiency of each MTD. The power consumed by the contending MTD k is equal to:

$$P_{\mathrm{MTD},k} = s_k\xi_{\mathrm{MTD}}\rho_{\mathrm{MTD}} + \bar{P}_{\mathrm{MTD}}. \tag{8.14}$$

With the power consumed by the AP, RIS, and contending MTDs, we can write the power consumed by the system during an RA frame as:

$$P = P_{\mathrm{AP}} + P_{\mathrm{RIS}} + \textstyle\sum_{k\in\mathcal{K}} P_{\mathrm{MTD},k}. \tag{8.15}$$

8.3 System Energy Efficiency

Considering the RA protocol described in Section 8.2, a *successful access attempt* (SA) occurs when, during the access phase, one packet replica transmitted by a contending MTD can be successfully decoded at the SIC receiver. Therefore, let $A \leq K$ be the number of SAs during the RA frame. From Fig. 19a, the system throughput can be calculated by:

$$G = \frac{A}{T_{\mathrm{t}} + T_{\mathrm{a}}} = \frac{A}{ST_{\mathrm{ts}} + ST_{\mathrm{as}}} = \frac{A}{(1+r)ST_{\mathrm{as}}} \text{ [packet/s]}, \tag{8.16}$$

where, in the last equality, $r = T_{\mathrm{ts}}/T_{\mathrm{as}}$ is the ratio between the duration of the training and the access time slots, namely the *training time slot ratio*. One should notice that, for access policies that do not require the DL training phase, such as the CRDSAP and IRSAP, the throughput can be calculated from (8.16) by setting $r = 0$.

Considering the system throughput in eq. (8.16) and the power consumption model in eq. (8.15), the overall system EE is:

$$\eta_{\mathrm{EE}} = \frac{G}{P_{\mathrm{AP}} + P_{\mathrm{RIS}} + \sum_{k\in\mathcal{K}} P_{\mathrm{MTD},k}} \text{ [packet/s/W]}, \tag{8.17}$$

where the denominator summands are given respectively by eqs. (8.12), (8.13), and (8.14).

8.4 Numerical Results

To evaluate the applicability of the RA protocol proposed in [4] to RIS-aided IoT networks, we present numerical results on the *throughput* (G) and *EE* (η_{EE}) considering the four access policies described in Subsection 8.2.2. The parameter values used to carry out the simulations and calculate the results are given in Table 5. In addition, for simplicity, we assume perfect estimation of the contending MTDs channel qualities, *i.e.*, in (8.8), $c_k = 1$ and $\varepsilon_k(s) = 0$ for all $k \in \mathcal{K}$ and $s \in \{1, \ldots, S\}$.

Remark 8.1. Figs. 24–26 reveal the *optimal throughput* and *optimal EE* in terms of the number of time slots, S. The optimal metric values are calculated knowing that $A \leq K$, and that the denominators of eqs. (8.16) and (8.17) are strictly increasing functions of S. Hence, the optimal throughput and EE w.r.t. S can be obtained by evaluating the metrics with the number of time slots in the set $\{S \in \mathbb{Z}_+ \mid A(S) \leq K\}$.

Fig. 21 presents the throughput and EE *vs.* the number of contending MTDs, K. Two values of number of time slots, S, are considered. CARP outperforms significantly the other policies both in throughput and EE for higher numbers of contending MTDs, being suitable for crowded networks. In the sequence, s-SCP presents an advantage over CRDSAP and IRSAP for $S = 20$ and $K > 11$. It is worth mentioning that CARP and s-SCP present slow throughput and EE decaying when compared to CRDSP and IRSAP. On the matter of the policies that do not require the DL training phase (CRDSAP and IRSAP), CRDSAP performance degrades with K later than IRSAP.

Table 5 – Parameter values used to generate the numerical results.

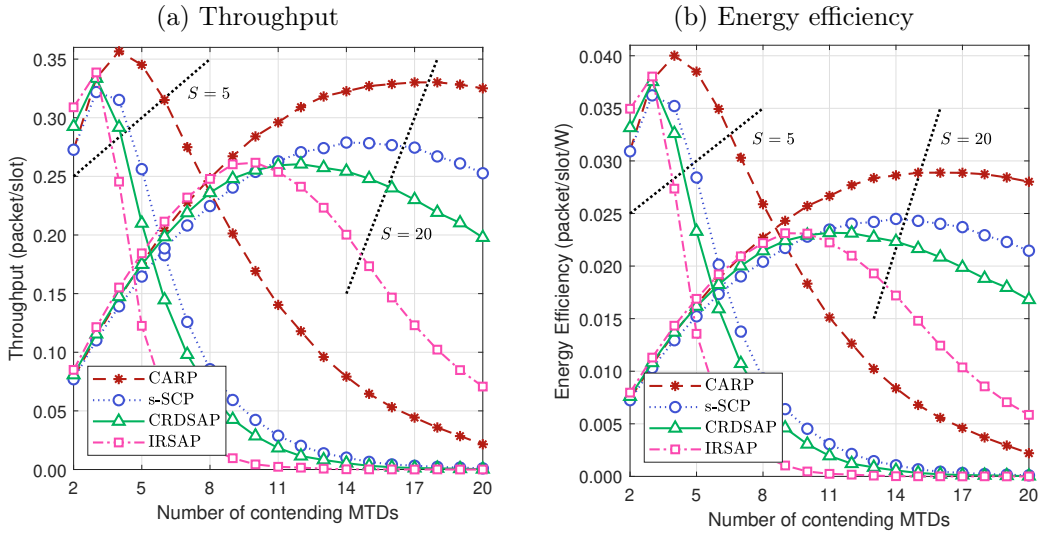
Parameter	**Description**
$\xi_{\mathrm{AP}} = 1.2$	Inverse of the efficiency of transmit PA
$\rho_{\mathrm{AP}} = 100$ mW	Transmit power
$\bar{P}_{\mathrm{AP}} = 9$ dBW	Static power dissipated by the hardware
$(d_{\mathrm{AP}}, \theta_{\mathrm{AP}}) = (20\mathrm{m}, \pi/4)$	AP position
$\alpha_{\mathrm{AP}} = 5$ dB	Antenna gain
$\xi_{\mathrm{MTD}} = 1.2$	Inverse of the efficiency of transmit PA
$\rho_{\mathrm{MTD}} \in [1, 100]$ mW	Transmit power
$\bar{P}_{\mathrm{MTD}} = 40$ mW	Static power dissipated by the hardware
$d_k \in [25, 100]$ m	Distance from the origin
$\theta_k \in [0, \pi/2]$	Angle with the RIS boresight
$\alpha_k = 5$ dB	MTD antenna gain
$N \in [4, 400]$ $(N_x = N_z)$	Number of RIS elements
$P_n(3) = 1.5$ mW	Power consumption of a phase-shifter having 3-bit resolution
$d_x = d_z = 10$ cm	RIS elements size
$S \in [1, 40]$	Number of time slots
$K \in [2, 20]$	Number of contending MTDs
$T_{\mathrm{as}} = 1$ s	Access time slot duration
$r = 0.2$	Training time slot ratio
$\gamma_{\mathrm{th}} = 0$ dB	SNR threshold of the SIC receiver
$\lambda = 10$ cm	Carrier wavelength
$\sigma_w^2 = -94$ dBm	Noise power
$s = 2$	s-SCP parameter

Source: The authors.

Fig. 22 presents the throughput and EE *vs.* the MTD transmit power, ρ_{MTD}. For $S = 5$, the CARP outperforms significantly all the other policies both in throughput and EE. Still, s-SCP attains performance close to CRDSAP, while IRSAP achieves poor throughput and EE values. Differently, for $S = 20$, CARP and s-SCP outperforms the other policies for low values of transmit power ($\rho_{\mathrm{MTD}} < 20$ mW). Finally, in the region $S = 20$, $\rho_{\mathrm{MTD}} > 20$ mW, IRSAP attains the best throughput, while CRDSAP attains the best EE due to the reduced number of transmitted packet replicas.

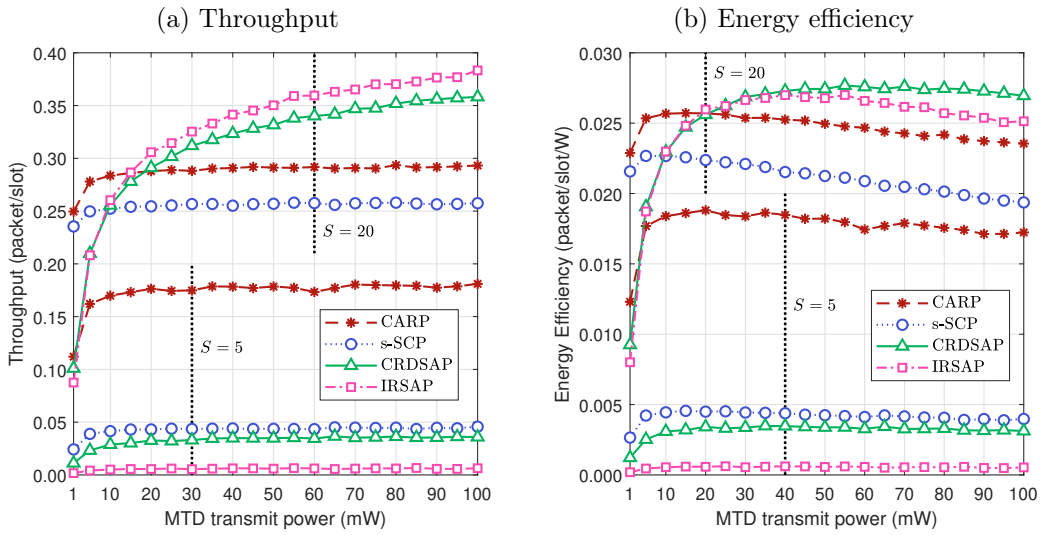
Fig. 23 shows the throughput and EE *vs.* the number of RIS elements, N. It is worth mentioning that we consider equal numbers of RIS elements in each direction, *i.e.*, $N_x = N_z$. In this result, for $S = 5$, again CARP attains remarkable performance both in throughput and EE, significantly outperforming the other policies. Also, for $S = 20$, $N < 150$, respectively CARP and s-SCP are the best policies. Conversely, for $S = 20$ and $N > 150$, the IRSAP becomes the best policy both in terms of throughput and EE.

Figure 21 – (a) $G \times K$ and (b) $\eta_{\text{EE}} \times K$ for two different numbers of time slots, $S \in \{5, 20\}$. $\rho_{\text{MTD}} = 10$ mW, $N = 100$.



Source: The authors.

Figure 22 – (a) $G \times \rho_{\text{MTD}}$ and (b) $\eta_{\text{EE}} \times \rho_{\text{MTD}}$ for two different numbers of time slots, $S \in \{5, 20\}$. $K = 10$, $N = 100$.

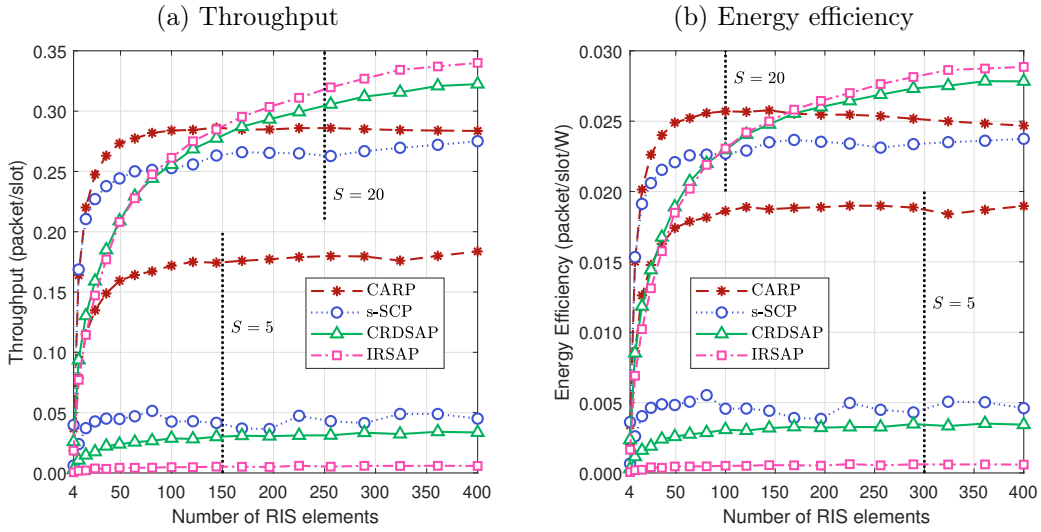


Source: The authors.

8.4.1 Optimal Throughput and EE w.r.t. Number of Time Slots

In the next three figures, we present the *optimal throughput* and *optimal EE* w.r.t. the number of time slots, S, considering increasing the a) number of contending MTDs (Fig. 24), b) MTD transmit power (Fig. 25), and c) number of RIS elements (Fig. 26). The optimal throughput and EE w.r.t. the number of time slots, S, *vs.* the number of contending MTDs, K, is revealed in Fig. 24. One can see that the attained throughput for all four policies decreases with K until reaching a floor. Conversely, the EE is monotonically

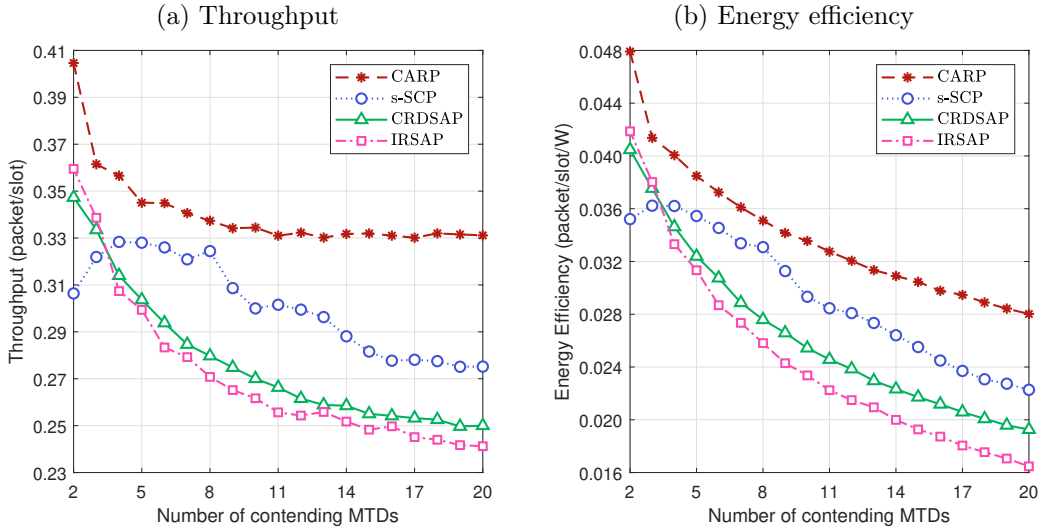
Figure 23 – (a) $G \times N$ and (b) $\eta_{\text{EE}} \times N$, for two different values of number of time slots, $S \in \{5, 20\}$. $K = 10$, $\rho_{\text{MTD}} = 10$ mW.



Source: The authors.

decreasing with K, since the power consumption is proportional to the number of contending MTDs. Looking at the throughput floor attained by CARP, one can notice its outstanding advantage to operate in crowded networks over the other policies.

Figure 24 – (a) Optimal G and (b) optimal η_{EE} w.r.t. S *vs.* the number of contending MTDs, K. $\rho_{\text{MTD}} = 10$ mW, $N = 100$.

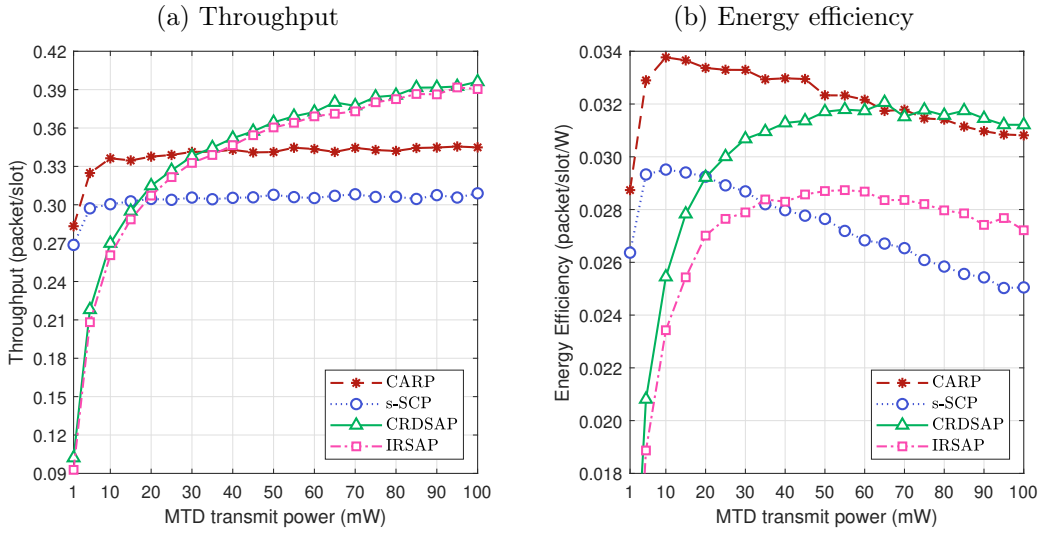


Source: The authors.

Fig. 25 presents the optimal throughput and EE w.r.t. the number of time slots, S, *vs.* the MTD transmit power, ρ_{MTD}. In this result, the throughput analysis is similar to that of Fig. 22a, except that the IRSAP performance is near the same as CRDSAP. Differently, under the EE perspective, the CARP policy attains remarkable results for low values of

transmit power ($\rho_{\text{MTD}} < 60$ mW), being a promising access policy for *very low-power* IoT networks. Still on EE, one can see that, for higher transmit power values, the IRSAP performance is significantly degraded when compared with CRDSAP. This is due to the fact that IRSAP allows the transmission of up to S packet replicas by each contending MTD, increasing significantly the number of transmitted replicas during the UL access phase.

Figure 25 – (a) Optimal G and (b) optimal η_{EE} w.r.t. S *vs.* the MTD transmit power, ρ_{MTD}. $K = 10$, $N = 100$.

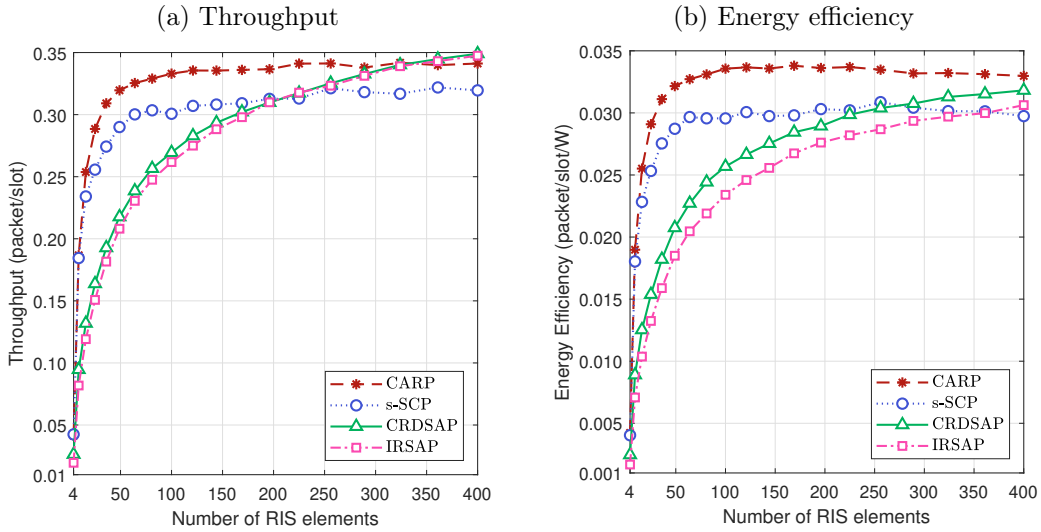


Source: The authors.

Fig. 26 reveals the optimal throughput and the optimal EE w.r.t. the number of time slots, S, *vs.* the number of RIS elements, N. Again, the CARP policy attains the best EE when compared with the other policies. It is worth mentioning that, after the maximum value is reached, the EE slowly decreases with N, since the power required by each RIS element is quite low. This demonstrates the advantage of RIS in providing reasonable network performance at the cost of an extremely-low power consumption when compared with conventional network solutions that use active elements.

Analyzing Figs. 24–26, one can conclude that the RIS-aided communication is an energy-efficient technology suitable for green IoT networks, providing high throughput gains with a marginal impact on the EE, even for surfaces with large numbers of reflecting elements. In addition, the *access policies* that exploit information of the channel qualities contribute significantly with the reduction of the power consumption at the MTDs-side. This is paramount in the proposed IoT network setup, since the MTDs are powered by limited-power sources such as batteries, while the AP and RIS may be connected to the power grid.

Figure 26 – (a) Optimal G and (b) optimal η_{EE} w.r.t. S *vs.* the number of RIS elements, N. $K = 10$, $\rho_{\text{MTD}} = 10$ mW.



Source: The authors.

8.5 Conclusions

In this paper, we develop a power consumption model, the throughput formula, and an expression for the overall EE for RIS-aided IoT networks that adopt a two time-scale RA protocol. Moreover, we adapt schemes from [4], [5], and [6] for respectively developing the access policies *s*-SCP, CRDSAP, and IRSAP. The numerical results on the RA protocol performance reveal that the access policies that exploit information of the channel qualities are suitable for green IoT networks, simultaneously attaining competitive EE and throughput with low power consumption at the MTDs-side.

8.6 References

[1] C. Zhu, V. C. M. Leung, L. Shu, and E. C.-H. Ngai, "Green internet of things for smart world," *IEEE Access*, vol. 3, pp. 2151–2162, Nov. 2015.

[2] Z. Chen, Y. Feng, Z. Tian, Y. Jia, M. Wang, and T. Q. S. Quek, "Energy efficiency optimization for irregular repetition slotted ALOHA-based massive access," *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 982–986, May 2022.

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

[4] V. Croisfelt, F. Saggese, I. Leyva-Mayorga, R. Kotaba, G. Gradoni, and P. Popovski, "A random access protocol for RIS-aided wireless communications," Mar. 2022, arXiv:2203.03377.

[5] E. Casini, R. De Gaudenzi, and O. Del Rio Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1408–1419, Apr. 2007.

[6] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 477–487, Feb. 2011.

[7] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, May 2020.

9 Paper C: Reconfigurable Intelligent Surfaces to Enable Energy-Efficient IoT Networks

João Henrique Inacio de Souza, José Carlos Marinello Filho, Taufik Abrão, and Cristiano Panazio

This paper has been published in the *proceedings of the 2022 Symposium on Internet of Things (SIoT),* São Paulo, Brazil, 24-28 Oct. 2022, pp. 1-4, doi: `https://doi.org/10.1109/SIoT56383.2022.10070317`.

Abstract

In this article, we study the uplink (UL) channel of a cellular network of Internet of Things (IoT) devices assisted by a reconfigurable intelligent surface (RIS) with a limited number of reflecting angle configurations. Firstly, we derive an expression of the required transmit power for the machine-type devices (MTDs) to attain a target signal-to-noise ratio (SNR), considering a channel model that accounts for the RIS discretization into sub-wavelength reflecting elements. Such an expression demonstrates that the transmit power depends on the target SNR, the position of the MTD in the service area, and the RIS setup, which includes the number of reflecting elements and the available reflecting angle configurations. Secondly, we develop an expression for the expected battery lifetime (EBL) of the MTDs, which explicitly depends on the MTD transmit power. Numerical simulations on the energy efficiency (EE) evaluated via the EBL demonstrate the benefits of adopting RISs to enable energy-efficient IoT networks.

Index Terms: Reconfigurable intelligent surface (RIS), Internet of Things (IoT), energy efficiency (EE).

9.1 Introduction

RIS is a promising low-cost and low-power *physical layer* (PHY) technology to enable sustainable wireless networks. This technology can produce controllable anomalous reflection of the incoming signal to create hot spots in areas of interest in the communication cell without any power amplifier (PA), consuming much less energy than common active PHY technologies [1]. On the matter of EE, it is shown in [2] that RIS-assisted networks

can achieve up to a threefold increase in EE when compared with networks assisted by multiple-antenna amplify-and-forward relays. This benefit on EE is particularly favorable for IoT networks, which need to achieve low-power consumption and extended coverage to serve hundreds of battery-powered MTDs. In this sense, in [3, 4], efficient random access protocols are proposed for RIS-assisted networks, introducing frameworks to coordinate the optimization of the RIS reflecting coefficients with the device transmissions. Moreover, in [5], the EE of the random access channel in a RIS-assisted IoT network is evaluated, revealing that simultaneously optimizing the RIS setup and the transmission protocol is paramount to attain competitive EE and throughput combined with low-power consumption.

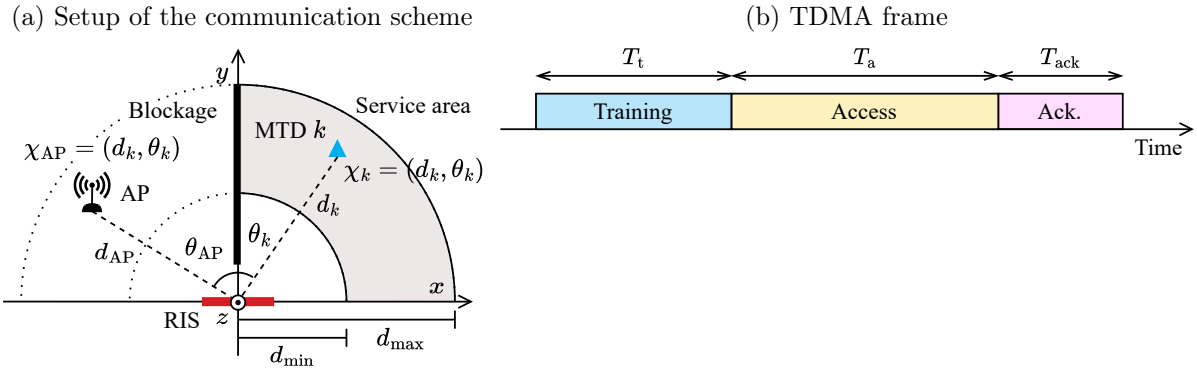
In this article, we study how RISs can improve the EBL of MTDs in IoT networks, focusing on the PHY aspects. Firstly, we derive an expression for the required transmit power to attain a target SNR considering a channel model that accounts for the RIS discretization into sub-wavelength reflecting elements [4]. Secondly, we develop an expression for the EBL of the MTDs, revealing that the numbers of reflecting elements and reflecting angle configurations at the RIS have significant impact on both system EE and network coverage. Differently from the work [5], in this article we focus on the EE problem exclusively at the devices-side, which needs more attention since the IoT devices are powered by limited-capacity batteries that are in general expensive and difficult to replace. Numerical evaluations demonstrate that the RIS can significantly improve the EBL of the MTDs by reducing the transmit power. Nevertheless, since the EBL is simultaneously limited by the number of RIS elements and configurations, the RIS setup must be carefully designed to achieve energy-efficient networks that can fully exploit the potential of the RIS.

9.2 System Model

In this section, we describe the system model. We consider a communication setup as depicted in Fig. 27a, where the single-antenna MTDs are connected with the single-antenna AP via the reflected path produced by the RIS, since the direct radio links between the AP and the MTDs are highly attenuated due to a blockage. In the communication scheme, the AP and MTDs are located at the xy-plane. Let $\chi_k = (d_k, \theta_k)$ denotes the position of the k-th MTD in the service area, where $d_{\min} \leq d_k \leq d_{\max}$ is the distance between the MTD and the origin, and $0 \leq \theta_k \leq \frac{\pi}{2}$ is the angle formed by the MTD with the RIS boresight. Similarly, let $\chi_{\text{AP}} = (d_{\text{AP}}, \theta_{\text{AP}})$ denotes the position of the AP.

The RIS is a thin surface located at the xz-plane with center at the origin. The surface is constituted by $N_x \in \mathbb{Z}_+$ reflecting elements in the x-axis direction and $N_z \in \mathbb{Z}_+$ in the z-axis direction, totaling $N = N_x N_z$ elements. From the operational standpoint, we consider that the RIS introduces a phase shift $0 \leq \theta_r \leq \frac{\pi}{2}$ with marginal amplitude attenuation to

Figure 27 – (a) AP, RIS, and service area of the communication cell where the MTDs are located; (b) TDMA frame.



Source: The authors.

the impinging signal. Importantly, the introduced phase shift belongs to a limited set of reflecting angle configurations, proposed to reduce the complexity of the RIS hardware and its control mechanism. Let $C \in \mathbb{Z}_+$ denotes the number of available configurations, and $\delta \triangleq \frac{\pi}{2(C+1)}$ denotes the angular resolution of the RIS. The set of reflecting angle configurations is defined as:

$$\Theta \triangleq \{\theta_r \mid \theta_r = c\delta, c = 1, \ldots, C\}. \tag{9.1}$$

In this article, we focus on the UL channel, assuming that the packet generation by each MTD is modeled as a Poisson process. The UL transmissions are carried out through time-division multiple access (TDMA) frames with duration less than the channel coherence time, $T_c > 0$. The TDMA frame, depicted in Fig. 27b, is comprised by a *training* block of length $T_t > 0$, an *access* block for data transmission with length $T_a > 0$, and an *acknowledgement* block of length $T_{\text{ack}} > 0$. In the training block, in coordination with the RIS, the AP broadcasts a signal to the MTDs to prepare for the data transmission. Then, the MTDs transmit their data signals in the access block. Lastly, the AP sends a message to recognize the successfully decoded data packets in the acknowledgement block. To focus purely on the PHY aspects of the system, we adopt this generic description of the TDMA frame. For comprehensive information on designing transmission protocols for RIS-assisted networks for MTDs, see [3, 4].

During the access block, the received data signal of the k-th MTD at the AP considering single-device transmission is equal to:

$$\mathbf{y} \triangleq \sqrt{\rho_k} h_k \mathbf{v}_k + \mathbf{w}, \tag{9.2}$$

where $\rho_k > 0$ is the MTD transmit power, $h_k \in \mathbb{C}$ denotes the channel coefficient, $\mathbf{v}_k \in \mathbb{C}^L$ such that $\mathbb{E}\{\|\mathbf{v}_k\|_2^2\} = 1$ denotes the transmitted data signal with $L \in \mathbb{Z}_+$ symbols, and $\mathbf{w} \in \mathbb{C}^L$ such that $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_L)$ denotes the additive white Gaussian noise vector,

where $\sigma_w^2 > 0$ is the noise power. From (9.2), the instantaneous SNR for the k-th MTD is equal to:

$$\gamma_k \triangleq \frac{\rho_k |h_k|^2}{\sigma_w^2}. \tag{9.3}$$

Based on the channel model of [4, 6] for communication schemes as depicted in Fig. 27a, the SNR can be rewritten as:

$$\gamma_k = \frac{\rho_k \beta_k |A_k|^2}{\sigma_w^2}, \tag{9.4}$$

where $\beta_k > 0$ denotes the *total path loss* and $A_k \in \mathbb{C}$ denotes the *array factor* due to the RIS discretization into sub-wavelength reflecting elements [4], respectively defined as:

$$\beta_k = \frac{\beta_0}{d_{\mathrm{AP}}^2 d_k^2} \cos^2 \theta_k, \tag{9.5}$$

$$A_k = N_z \sum_{n=1}^{N_x} e^{j2\pi(\sin\theta_k - \sin\theta_r)n}, \tag{9.6}$$

where $\beta_0 > 0$ denotes the path loss at the reference distance and $j = \sqrt{-1}$. From (9.6), one can notice that the RIS array factor directly depends on the reflecting angle θ_r, which is limited by the set Θ defined in eq. (9.1). Finally, substituting (9.5) and (9.6) in (9.4), we obtain the following expression of the instantaneous SNR:

$$\gamma_k = \frac{\rho_k \beta_0 N_z^2 |\sum_{n=1}^{N_x} e^{j2\pi(\sin\theta_k - \sin\theta_r)n}|^2 \cos^2 \theta_k}{\sigma_w^2 d_{\mathrm{AP}}^2 d_k^2}. \tag{9.7}$$

Therefore, one can note that, besides the MTD and AP positions, the RIS setup, specifically the number of reflecting elements, $N = N_x N_z$, and the reflecting angle configuration, $\theta_r \in \Theta$, has a significant impact on the instantaneous SNR.

9.3 Device Energy Consumption Model

In this section, we develop a model for the MTDs energy consumption. We propose an expression for the EBL as EE metric specially focused on the devices-side. Let $\bar{\gamma} \in \mathbb{R}$ denotes the target SNR at the AP for an MTD to experience reliable communication. From (9.7), we can derive the transmit power of the k-th MTD to attain the average SNR $\bar{\gamma}$ as:

$$\rho_k(\bar{\gamma}) = \bar{\gamma} \left(\frac{\beta_0 N_z^2 |\sum_{n=1}^{N_x} e^{j2\pi(\sin\theta_k - \sin\theta_r)n}|^2 \cos^2 \theta_k}{\sigma_w^2 d_{\mathrm{AP}}^2 d_k^2} \right)^{-1}. \tag{9.8}$$

From (9.8), one can notice that the transmit power depends on the position of the MTD, $\chi_k = (d_k, \theta_k)$, as well as on the number of elements at the RIS, $N = N_x N_z$, and the angle of reflection of the RIS, θ_r. Deriving the radiofrequency (RF) transmit power is paramount to analyze the MTD EE as it constitutes most of the dissipated power during the TDMA frame.

To develop the model for the *device energy consumption*, we consider the MTDs under two device modes: *i) transmission* mode, and *ii) receive* mode. From Fig. 27b, one can perceive that the MTDs are in the receive mode during the training and acknowledgement blocks, while they are in the transmit mode during the access block. Therefore, the *energy consumption* $E(\chi_k) > 0$ of the k-th MTD at the position χ_k is defined by:

$$E(\chi_k) \triangleq E_s + T_t P_{\mathrm{Rx}} + T_a(P_c + \xi\rho_k) + T_{\mathrm{ack}} P_{\mathrm{Rx}}, \tag{9.9}$$

where $E_s > 0$ denotes the static energy consumption, $P_{\mathrm{Rx}} > 0$ denotes the power dissipated in the receive mode, $P_{\mathrm{Tx}} = P_c + \xi\rho_k$ denotes the power dissipated in the transmission mode, where $P_c > 0$ denotes the power dissipated by the device circuits, and $\xi > 1$ denotes the inverse of efficiency of the RF PA.

The EBL is a metric to evaluate the device EE given its battery capacity, energy consumption, and reporting period [7]. Let $T_r > 0$ denotes the expected length between consecutive *reporting periods* of the MTD. The EBL in seconds of the k-th MTD at the position χ_k is equal to:

$$L(\chi_k) \triangleq \frac{E_0 T_r}{E_s + T_a(P_c + \xi\rho_k) + (T_t + T_{\mathrm{ack}})P_{\mathrm{Rx}}}, \tag{9.10}$$

where $E_0 > 0$ denotes the battery capacity of the MTD in Joule. From the transmission protocol standpoint, the EBL depends mainly on the length of the TDMA frame blocks and the transmit power. Especially, as suggested by (9.8), the transmit power is strongly linked to the RIS setup, particularly the number of RIS elements and reflecting angle configurations. Therefore, carefully designing the RIS setup is essential to improve the system EE in IoT networks, particularly at the devices-side.

9.4 Numerical Results

In this section, we analyze the EE of the RIS-assisted IoT network using the EBL as the main metric. In the simulations, we calculate the required transmit power to attain the target SNR, defined in eq. (9.8), at each position χ_k. Moreover, we assume that the transmit power cannot exceed the maximum transmit power, $\rho_{\max} > 0$, *i.e.*, $\rho_k(\bar{\gamma}) \leq \rho_{\max}$. In this sense, when the MTD needs more power than $\rho_{\max}$ to attain the target SNR, such MTD is under outage. All the parameter values deployed to generate the numerical results are listed in Table 6.

Fig. 28 depicts $\rho_k(\bar{\gamma})$ required to attain the target SNR $\bar{\gamma} = 10$ dB across the service area of the communication cell. We assume that the RIS has two different reflecting angle configurations, *i.e.*, $C = 2$. Additionally, at each position, $\rho_k(\bar{\gamma})$ is calculated considering the best RIS configuration, that is, the configuration that results in the smallest transmit power. Analyzing Fig. 28, one can see that the minimum $\rho_k(\bar{\gamma})$ observed across the service area decreases with N due to the enhanced array factor of the RIS, defined in eq. (9.6).

Table 6 – Parameter values used to generate the numerical results.

Parameter	**Description**
$\chi_{\text{AP}} = (20\text{ m}, \pi/4)$	AP position
$d_k \in [20, 100]$ m	Distance of the MTD from the origin
$\theta_k \in [0, \pi/2]$	MTD-RIS angle
$N \in [4, 100]$	# RIS elements ($N_x = N_z$)
$C \in \{2, 4, 8, 16\}$	# RIS reflecting angle configurations
$T_c = 50$ ms	Channel coherence time
$T_a = 0.85T_c$	Training block length
$T_t = 0.10T_c$	Access block length
$T_{\text{ack}} = 0.05T_c$	Acknowledgement block length
$\beta_0 = -52$ dB	Path loss at the reference distance
$\sigma_w^2 = -94$ dBm	Noise power
$T_r = 300$ s	Length between consecutive reporting periods
$E_0 = 2.5$ kJ	Battery capacity
$E_s = 10\mu$J	Static energy consumption
$P_c = 1$ mW	Power dissipated by the MTD circuits
$P_{\text{Rx}} = 100$ mW	Power dissipated in the receive mode
$\xi = 1.33$	Inverse of efficiency of the RF PA
$\rho_{\max} = 24$ dBm	Maximum transmit power
$\bar{\gamma} = 10$ dB	Target SNR

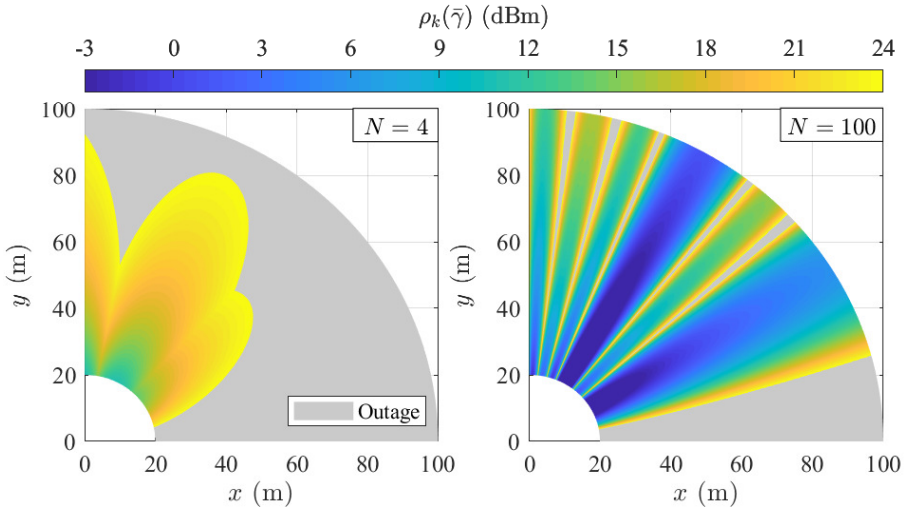
Source: The authors.

At the same time, the outage region area reduces as N increases, improving the network coverage. For $N = 100$, there are two main beams in the service area (dark blue color), where the required $\rho_k(\bar{\gamma})$ is less than the values in the other positions. Specifically, these beams are pointed towards the reflecting angles configured at the RIS, $\boldsymbol{\Theta}$, which in the analyzed scenario with $C = 2$ are equal to 30° and 60°. Hence, these beams represent the regions where the RIS can efficiently reflect much of the signal power.

Based on the results of Fig. 28, Fig. 29a depicts the EBL $L(\chi_k)$ in years across the service area for $C = 2$. From (9.10), it is straightforward to see that $L(\chi_k)$ must follow the inverse behavior of $\rho_k(\bar{\gamma})$. Accordingly, in Fig. 29a, one can notice that the increase on the RIS array factor due to the higher N improves the maximum $L(\chi_k)$, reducing the outage region area, and simultaneously improving the devices EE and the network coverage.

In order to analyze the effect of increasing the number of RIS reflecting angle configurations, Fig. 29b depicts $L(\chi_k)$ across the service area for $C = 8$. Comparing this result with the same N to Fig. 29a, one can notice that increasing C improves the EBL $L(\chi_k)$ across the service area and reduces the outage region size. Such improvements arise since the RIS with an increased number of reflecting angle configurations can beamform the signal towards a greater number of directions. Despite the improvement on coverage, one can notice that specifically for $N = 100$ there is a permanent outage region into the service

Figure 28 – $\rho_k(\bar{\gamma})$ across the service area. $C = 2$, $\bar{\gamma} = 10$ dB, and $\rho_{\max} = 24$ dBm.



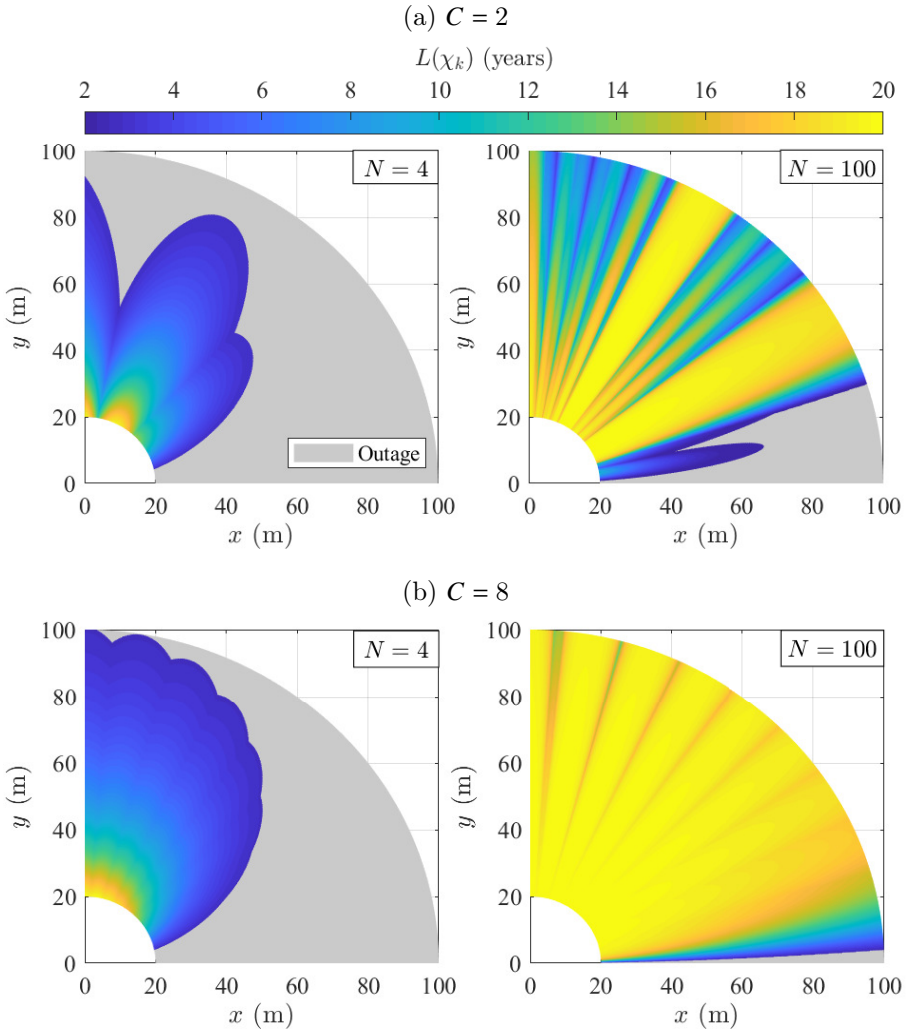
Source: The authors.

area near to $\theta_k = \pi/2$. Such region is due to the setup of the communication cell, depicted in Fig. 27a, where the MTDs with $\theta_k \approx \pi/2$ are located in approximately the same plane as the RIS, an unfavorable region for signal reflection as revealed by the cosine term in (9.5).

Fig. 30a depicts the average $\rho_k(\bar{\gamma})$ and $L(\chi_k)$ over the service area as a function of N. Following the behavior observed in the previous results, the average $\rho_k(\bar{\gamma})$ decreases with N, while the average $L(\chi_k)$ increases with it. Specifically for $C \in \{2, 4\}$ and $N > 36$, one can notice that the average $\rho_k(\bar{\gamma})$ has a clear oscillating behavior. This is due to the low number of main beams produced by the small number of RIS reflecting angle configurations C and the fact that increasing N may increase the signal power at the main beams at the expense of decreasing it at the side beams. Moreover, for a fixed C, one can notice that the average $\rho_k(\bar{\gamma})$ and, consequently, the average $L(\chi_k)$ are limited by N, demonstrating that the RIS setup, *i.e.*, N and C, must be carefully designed to avoid a setup that cannot be fully exploited by the network.

Lastly, Fig. 30b depicts the percentage of the outage region area as a function of N. This result reinforces that the outage region decreases asymptotically both with N and C, indicating that the RIS setup is directly associated with the network coverage. Indeed, such a result demonstrates that, even with large values of N and C, there is a permanent outage region for $\theta_k \approx \pi/2$ due to the communication cell setup. Therefore, if the aim is to enhance the network coverage in the simulated setup, a RIS with $(N, C) = (81, 8)$ is sufficient to get an outage region with area as small as 2%.

Figure 29 – $L(\chi_k)$ across the service area. $C \in \{2, 8\}$, $T_r = 300$ s, $\bar{\gamma} = 10$ dB, and $\rho_{\max} = 24$ dBm.

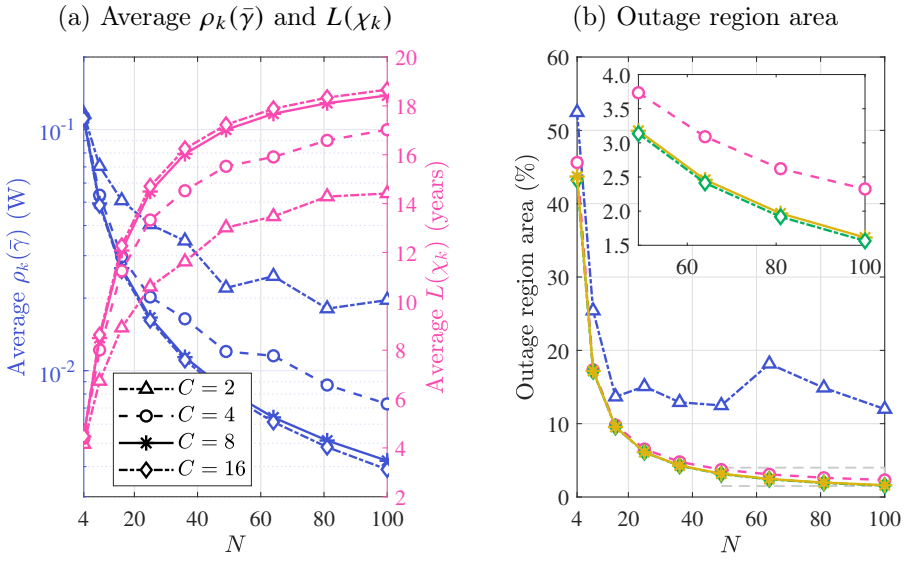


Source: The authors.

9.5 Conclusions

In this article, we study the design of energy-efficient IoT networks assisted by a RIS. Firstly, we derive an expression for the required transmit power to attain a target SNR considering the RIS discretization into sub-wavelength reflecting elements. Secondly, we develop an expression for the EBL of the MTDs. The numerical results demonstrate that the RIS can significantly improve the EBL of the MTDs by reducing the transmit power. Additionally, the RIS can enhance the network coverage as long as it has the right numbers of reflecting elements and reflecting angle configurations. Nevertheless, since the EBL is simultaneously limited by the number of RIS elements and configurations, the RIS setup must be carefully designed to achieve energy-efficient networks that can fully exploit the potential of the RIS.

Figure 30 – (a) Average $\rho_k(\bar{\gamma})$ and average $L(\chi_k)$ as a function of N; (b) Outage region area as a function of N. $C \in \{2, 4, 8, 16\}$, $T_r = 300$ s, $\bar{\gamma} = 10$ dB, and $\rho_{\max} = 24$ dBm.



Source: The authors.

9.6 References

[1] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

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

[3] X. Cao, B. Yang, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Han, H. V. Poor, and L. Hanzo, "Massive access of static and mobile users via reconfigurable intelligent surfaces: Protocol design and performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1253–1269, Apr. 2022.

[4] V. Croisfelt, F. Saggese, I. Leyva-Mayorga, R. Kotaba, G. Gradoni, and P. Popovski, "A random access protocol for RIS-aided wireless communications," in *2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*, 4-6 July 2022, pp. 1–5.

[5] J. H. I. de Souza, J. C. M. Filho, T. Abrão, and C. Panazio, "Energy efficiency and throughput of random access protocols for RIS-aided IoT networks," in *2022 IEEE 8th World Forum on Internet of Things (WF-IoT)*, 26 Oct.-11 Nov. 2022, pp. 1–6, to appear.

[6] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, May 2020.

[7] A. Azari, Č. Stefanović, P. Popovski, and C. Cavdar, "Energy-efficient and reliable IoT access without radio resource reservation," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 2, pp. 908–920, June 2021.

10 Paper D: Randomized Control of Wireless Temporal Coherence via Reconfigurable Intelligent Surface

João Henrique Inacio de Souza, Victor Croisfelt, Fabio Saggese, Taufik Abrão, and Petar Popovski

This paper has been presented in the *proceedings of the 2023 IEEE International Conference on Communications Workshops (ICC Workshops),* Rome, Italy, 28 May-01 June 2023, pp. 1535-1540, doi: `https://doi.org/10.1109/ICCWorkshops57953.2023.10283583`.

Abstract

A reconfigurable intelligent surface (RIS) can shape the wireless propagation channel by inducing controlled phase shift variations to the impinging signals. Multiple works have considered the use of RIS with time-varying configurations of reflection coefficients. In this work, we use the RIS to control the channel coherence time and introduce a generalized discrete-time-varying channel model for RIS-aided systems. We characterize the temporal variation of the channel correlation by assuming that the configuration of RIS' elements changes at every time step. The analysis converges to a randomized framework to control the channel coherence time by setting the number of RIS' elements and their phase shifts. The main result is a framework for a flexible block-fading model, where the number of samples within a coherence block can be dynamically adapted.

Index Terms: Reconfigurable intelligent surface (RIS), temporal correlation, fading.

10.1 Introduction

A RIS consists of a massive number of passive reflecting elements able to alter the phase shifts and possibly the amplitude of impinging wireless signals [1], thereby exerting control over the wireless propagation. Some RIS instances can be seen as passive holographic MIMO surfaces. Numerous use cases have been proposed to show how such control can benefit the communication between a transmitter (Tx) and a receiver (Rx), where the prevailing focus is on improving the communication performance [1]. However, relatively few works explore how to use the RIS to induce temporal diversity in the wireless channel and avoid prolonged unfavorable propagation to a given user.

To illustrate, consider the channel aging problem that occurs due to Rx mobility which makes the CSI acquired by the Tx unreliable over time. Works as [2–5] suggest the use of a RIS to deal with this problem by compensating for the Doppler effects of mobility. The focus of these works typically relies on optimizing the RIS' configurations of the elements aiming to minimize the channel aging effect. In [6], a continuous-time propagation model is given and is used to configure the RIS in such a way that the received power is maximized whereas the delay and Doppler spread are minimized. The authors in [7] study the spatial-temporal correlation implied by the RIS when it is embedded in an isotropic scattering environment. Nevertheless, these prior works do not analyze how the temporal channel statistics, such as the coherence time, can be modeled as a function of the properties of the RIS' elements. In this paper, we focus on studying how the RIS can shape temporal channel statistics by relying on a discrete-time-varying channel model. A closely related work is [8] where the authors proposed a RIS phase hopping scheme with the purpose of transforming a slow-fading into a fast-fading channel. This was achieved by randomly varying the RIS' configurations, significantly improving the outage performance without the need for CSI at the RIS. Nevertheless, this paper focused on how the random variation of phases impacts the outage performance, while here we analyze temporal channel correlation that stems from the properties of the RIS.

We propose a generalized discrete-time-varying channel model for RIS-aided communication systems, showing how the part of the propagation environment controlled by the RIS shapes the discrete temporal channel statistics. We characterize the temporal variation of the channel correlation as the RIS' reflections configuration changes at every time index. This analysis reveals how one can control the coherence time of the channel by changing the number of RIS' reflecting elements and their phase shift configurations. Our findings corroborate the results from [8], and prove the possibility of using the RIS to generate a flexible block-fading model.

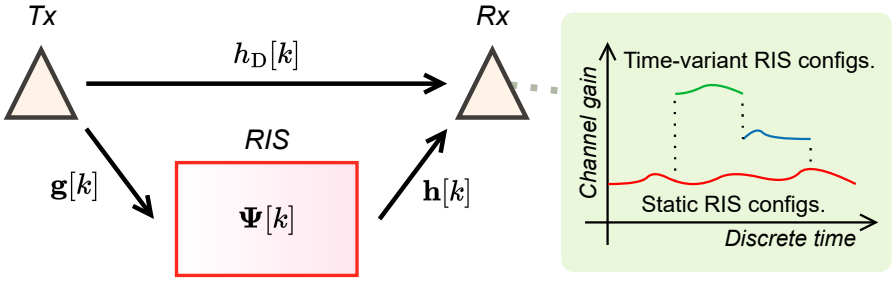
Notation. Boldface lowercase $\mathbf{a}$ and uppercase $\mathbf{A}$ letters represent vectors and matrices, respectively. Calligraphic letters $\mathcal{A}$ represent finite sets. Operators: transpose by $\{\cdot\}^{\intercal}$, complex conjugate by $\{\cdot\}^{*}$, and real part by $\Re\{\cdot\}$. Important functions are: $\lfloor\cdot\rfloor$ the floor function, $\delta[\cdot]$ the Kronecker's delta function, and $\mathrm{sinc}(\theta) = \sin(\theta)/\theta$. The expected value operator is $\mathbb{E}\{\cdot\}$ and, unless otherwise stated, it is taken w.r.t. the variable k. The complex Gaussian distribution is denoted as $\mathcal{CN}(\mu, \sigma^2)$ with mean μ and variance σ^2, whereas a uniform random distribution over the interval $[a, b]$ is $\mathcal{U}(a, b)$.

10.2 System Model

The communication setup consists of one single-antenna Tx, one single-antenna Rx, and one RIS with $N \in \mathbb{Z}_{+}$ passive reflecting elements, operating in narrowband communication

channel and in free space.[1] The wireless channel consists of two distinct radio paths from the Tx to the Rx, the direct path and the reflected path controlled by the RIS, see Fig. 31. The index of the complex samples in the discrete-time domain is denoted by $k \in \mathbb{Z}$.

Figure 31 – RIS-aided communication system, where the RIS' elements imposing time-variant reflection configurations can alter the channel response.



Source: The authors.

Considering the downlink, let $h_D[k] \in \mathbb{C}$ denote the channel coefficient from the Tx to the Rx, $g_n[k] \in \mathbb{C}$ denote the channel coefficient from the Tx to the n-th reflecting element of the RIS, and $h_n[k] \in \mathbb{C}$ denote the channel coefficient from the n-th reflecting element of the RIS to the Rx, $n \in \{1, \ldots, N\}$. Let the channel vectors from the Tx to the RIS and from the RIS to the Rx be $\mathbf{g}[k] = [g_1[k], \ldots, g_N[k]]^\intercal$ and $\mathbf{h}[k] = [h_1[k], \ldots, h_N[k]]^\intercal$, respectively. Consider then that the n-th reflecting element of the RIS can induce a phase shift of $\phi_n[k] \in [0, 2\pi)$ upon the impinging signal with marginal impact on the signal's amplitude. Thus, we denote the RIS' configuration impressed at time k as the matrix $\mathbf{\Psi}[k] = \text{diag}([\psi_1[k], \ldots, \psi_N[k]]^\intercal)$, whose n-th diagonal entry is the *reflection coefficient* $\psi_n[k] = e^{-j\phi_n[k]}$ of the n-th RIS element.[2] Using the narrowband model from [1], we assume that a RIS configuration is constant within the time of each sample k.[3] Moreover, we assume that the RIS' elements have a flat frequency response, preserving the coherence bandwidth of the equivalent channel, $B_c > 0$.[4] The equivalent channel $h_{eq}[k] \in \mathbb{C}$ from the Tx to the Rx is then [1]:

$$\begin{aligned} h_{eq}[k] &= h_D[k] + \mathbf{g}^\intercal[k]\mathbf{\Psi}[k]\mathbf{h}[k], \\ &= \underbrace{h_D[k]}_{\text{Uncontrollable}} + \underbrace{\textstyle\sum_{n=1}^{N} g_n[k]h_n[k]\psi_n[k]}_{\text{Controllable}}. \end{aligned} \tag{10.1}$$

[1] We assume a scenario with a single Rx and single-antenna equipment to simplify the analysis. Note that the obtained results can be directly generalized to multi-Rx and multi-antenna scenarios if statistical independence holds.

[2] The RIS is considered to have unitary attenuation to simplify the presentation. The generalization is considered to be straightforward.

[3] During the switching time from a configuration to another, the RIS can generate an unpredictable channel behavior. Throughout the paper, we assume that this effect is negligible considering that the switching time is much lower than the time reserved for each configuration.

[4] B_c is inversely proportional to the *channel delay spread*, determined by the multiple reflection delays of the signal between the Tx and Rx.

Eq. (10.1) is made up of two different time-variant terms: the *uncontrollable component* depending only on the properties of the propagation environment, and the *controllable components* controlled by changing the RIS' configurations.

10.3 Discrete-Time-Varying Channel Model

In this section, we present a discrete-time-varying model for the channels $h_{\mathrm{D}}[k]$, $\mathbf{g}[k]$, and $\mathbf{h}[k]$ in order to understand the impact of switching the reflection coefficients of the RIS over the time index k. Based on such a model, we show how the temporal correlation of the equivalent channel behaves by characterizing its ACF w.r.t. k.

10.3.1 The Model

Without loss of generality, let $h[k] \in \mathbb{C}$ denote a generic channel coefficient sample from one of the channels $h_{\mathrm{D}}[k]$, $g_n[k]$, and $h_n[k]$. To represent both LoS and NLoS components, we assume the time-variant Rician fading model as follows [9]:

$$h[k] = \bar{h} + \check{h}[k], \tag{10.2}$$

where $\bar{h} \in \mathbb{C}$ denotes the time-invariant LoS channel component and $\check{h}[k] \in \mathbb{C}$ denotes the time-variant NLoS channel component.[5] Specifically, the NLoS component is modeled as a stationary AR(1) random process with recurrence relation [10–12][6]

$$\check{h}[k] = \alpha_h \check{h}[k-1] + \sqrt{1-\alpha_h^2}\sigma_h w[k], \tag{10.3}$$

where $0 \leq \alpha_h < 1$ denotes the AR(1) parameter, $\sigma_h^2 > 0$ denotes the power of the NLoS component, and $w[k]$ is a white stationary complex Gaussian process such that $w[k] \sim \mathcal{CN}(0,1)$. From (10.2) and the properties of the AR(1), it is straightforward to demonstrate that:

$$\mathbb{E}\{h[k]\} = \bar{h} \text{ and } \mathbb{E}\{|h[k]|^2\} = |\bar{h}|^2 + \sigma_h^2. \tag{10.4}$$

Hence, the Rice factor $\kappa_h > 0$ of $h[k]$ is defined as the ratio between the powers of the LoS and NLoS components,

$$\kappa_h = \frac{|\bar{h}|^2}{\sigma_h^2}. \tag{10.5}$$

Given the channel model and the relationship between the powers of its time-invariant and -variant parts, we now define what are the intrinsic parameters to the environment, meaning that they are determined by the physical properties of the wireless propagation medium and the setup geometry, and *cannot* be directly controlled by the system designer.

[5] The time-variant nature of the NLoS component comes from changes in the communication environment, such as the ones caused by scatterers moving, appearing, and disappearing from the environment [3].

[6] In [10], the authors demonstrated that the AR model could be considered for the computer simulation of correlated fading channels, corroborating that low orders are appropriate for narrowband Doppler fading processes. Moreover, [11, 12] unveil that an AR(1) model is enough to capture most of the channel tap dynamics.

Definition 10.1. (Set of Environmental Parameters) We denote as $\mathcal{E}_h = \{\alpha_h, \kappa_h, \sigma_h^2\}$ the set of environmental parameters.

10.3.2 Temporal Correlation of the Equivalent Channel

We now carry out an analysis of the correlation among the channel samples over the time index k. Our first result is summarized in the following lemma.

Lemma 10.1. *Consider that the channel coefficients $h_{\mathrm{D}}[k]$, $g_n[k]$, and $h_n[k]$, $\forall n$, follow the time-variant Rician model in eq. (10.2). Then, the ACF $R_{h_{\mathrm{eq}}h_{\mathrm{eq}}} : \mathbb{Z} \to \mathbb{R}$ of the equivalent channel is given by*

$$R_{h_{\mathrm{eq}}h_{\mathrm{eq}}}[\tau] = \mathbb{E}\{h_{\mathrm{eq}}[k]h_{\mathrm{eq}}^*[k-\tau]\}, \tag{10.6}$$

$$= |\bar{h}_{\mathrm{D}}|^2 + \alpha_{h_{\mathrm{D}}}^{|\tau|}\sigma_{h_{\mathrm{D}}}^2 + \textstyle\sum_{n=1}^{N}(|\bar{g}_n|^2 + \alpha_{g_n}^{|\tau|}\sigma_{g_n}^2)(|\bar{h}_n|^2 + \alpha_{h_n}^{|\tau|}\sigma_{h_n}^2)R_{\psi_n\psi_n}[\tau] +$$

$$+ \textstyle\sum_{n=1}^{N}\sum_{\substack{n'=1 \\ n'\neq n}}^{N} \bar{g}_n\bar{g}_{n'}^*\bar{h}_n\bar{h}_{n'}^* R_{\psi_n\psi_{n'}}[\tau] + 2\Re\{\bar{h}_{\mathrm{D}}^* \textstyle\sum_{n=1}^{N} \bar{g}_n\bar{h}_n\mathbb{E}\{\psi_n[k]\}\}, \tag{10.7}$$

where $R_{\psi_n\psi_{n'}} : \mathbb{Z} \to \mathbb{R}$ is the cross-correlation function (CCF) of the RIS' reflection coefficients, calculated for the discrete-time delay τ as:

$$R_{\psi_n\psi_{n'}}[\tau] = \mathbb{E}\{\psi_n[k]\psi_{n'}^*[k-\tau]\}. \tag{10.8}$$

Proof. See Appendix A. □

From the above result, one can note that the channel ACF inherits the *uncontrollable* part and the *controllable* parts from the equivalent channel in eq. (10.1).

10.4 A Randomized Framework for Controlling the Temporal Correlation

We start by providing a general framework that describes how to control the temporal correlation by using Lemma 10.1 to set the number of RIS' reflecting elements N and/or designing their configuration $\{\psi_n[k]\}_{n=1}^N$. Note that through N we select a RIS of sufficient size to meet the system's temporal correlation requirements. Then, we study the case with uniformly distributed phase shifts, as in [8].

10.4.1 Temporal Correlation under Random Phase Shifts

We first make the following simplifying assumptions. **(1)** There is no direct path from the Tx to the Rx, *i.e.*, $h_{\mathrm{D}}[k] = 0$. This holds when obstacles block the direct path between the Tx and the Rx, as in dense urban scenarios and industries. **(2)** The LoS components are predominant in the channels from the Tx to the RIS, *i.e.*, $\kappa_{g_n} \to \infty$ dB , $\forall n$. Therefore,

these channels are static, *i.e.*, $g_n[k] = \bar{g}_n, \forall n$. This can be justified by the fact that the Tx and the RIS do not move and the deployment of the RIS is chosen so as to enhance the LoS components between the Tx and the RIS. Using these assumptions, the equivalent channel in (10.1) can be rewritten as:

$$h'_{\text{eq}}[k] = \bar{\mathbf{g}}^{\intercal}\mathbf{\Psi}[k]\mathbf{h}[k], \tag{10.9}$$

and its ACF is given by:

$$R_{h'_{\text{eq}}h'_{\text{eq}}}[\tau] = \sum_{n=1}^{N} |\bar{g}_n|^2(|\bar{h}_n|^2 + \alpha_{h_n}^{|\tau|}\sigma_{h_n}^2)R_{\psi_n\psi_n}[\tau] + \sum_{n=1}^{N}\sum_{\substack{n'=1 \\ n'\neq n}}^{N} \bar{g}_n\bar{g}^*_{n'}\bar{h}_n\bar{h}^*_{n'}R_{\psi_n\psi_{n'}}[\tau], \tag{10.10}$$

where $\bar{\mathbf{g}} = [\bar{g}_1, \ldots, \bar{g}_N]^{\intercal}$ and $\bar{\mathbf{h}} = [\bar{h}_1, \ldots, \bar{h}_N]^{\intercal}$. We further assume that the N channels from the RIS' elements to the Rx share the same set of environmental parameters $\mathcal{E} = \{\alpha, \kappa, \sigma^2\}$, where $\alpha_{h_n} = \alpha$, $\kappa_{h_n} = \kappa$, and $\sigma^2_{h_n} = \sigma^2, \forall n$. This assumption is valid when considering that the process which introduces the time variations is the same for all N channels and that the receptions occur in the far-field regime [10]. Now, let us assume that the phase shifts at a given time k are drawn from a uniform random distribution:

$$\phi'_n[k] \sim \mathcal{U}(\pi - \theta, \pi + \theta), \forall n, \tag{10.11}$$

where $\theta \in [0, \pi]$ is the phase shifts distribution parameter. Then, the CCF of the RIS' reflection coefficients is:

$$R_{\psi'_n\psi'_{n'}}[\tau] = \begin{cases} \delta[\tau] + (1 - \delta[\tau])\text{sinc}^2(\theta), & \text{if } n = n' \\ \text{sinc}^2(\theta), & \text{otherwise} \end{cases}, \tag{10.12}$$

where we used eq. (10.8). Substituting (10.12) into eq. (10.10) and given that $|\bar{\mathbf{g}}^{\intercal}\bar{\mathbf{h}}|^2 = \sum_{n=1}^{N}\sum_{n'=1}^{N} \bar{g}_n\bar{g}^*_{n'}\bar{h}_n\bar{h}^*_{n'}$ results in the ACF of the equivalent channel

$$\begin{aligned} R_{h'_{\text{eq}}h'_{\text{eq}}}[\tau] &= (1 - \text{sinc}^2(\theta))(\textstyle\sum_{n=1}^{N} |g_n|^2|\bar{h}_n|^2 + \sigma^2\|\bar{\mathbf{g}}\|_2^2)\delta[\tau] + \text{sinc}^2(\theta)(|\bar{\mathbf{g}}^{\intercal}\bar{\mathbf{h}}|^2 + \alpha^{|\tau|}\sigma^2\|\bar{\mathbf{g}}\|_2^2), \\ &\overset{(a)}{=} \sigma^2\|\bar{\mathbf{g}}\|_2^2((1 - \text{sinc}^2(\theta))(\kappa + 1)\delta[\tau] + \text{sinc}^2(\theta)(N\kappa\eta + \alpha^{|\tau|})), \end{aligned} \tag{10.13}$$

$$\text{with } \eta = \left|\frac{\bar{\mathbf{g}}^{\intercal}\bar{\mathbf{h}}}{\|\bar{\mathbf{g}}\|_2\|\bar{\mathbf{h}}\|_2}\right|^2. \tag{10.14}$$

Specifically, the equality (a) in eq. (10.13) results from the Rice factor in eq. (10.5), *i.e.*, from substituting the term $\|\bar{\mathbf{h}}\|_2^2 = \sum_{n=1}^{N} |\bar{h}_n|^2$ by $N\kappa\sigma^2$. Also, from the triangle inequality, η lies between $[0, 1]$ and is a measure of *orthogonality* between the LoS components of the channels $\mathbf{g}[k]$ and $\mathbf{h}[k]$, depending on N and the positions of the RIS, Tx, and Rx [13].

By using the results for the CCF and ACF, we derive the correlation coefficient between two channel samples delayed by $|\tau|$ samples,

$$\rho[\tau] = \frac{R_{h'_{\text{eq}}h'_{\text{eq}}}[\tau]}{R_{h'_{\text{eq}}h'_{\text{eq}}}[0]} = \frac{(1 - \text{sinc}^2(\theta))(\kappa + 1)\delta[\tau] + \text{sinc}^2(\theta)(N\kappa\eta + \alpha^{|\tau|})}{(1 - \text{sinc}^2(\theta))(\kappa + 1) + \text{sinc}^2(\theta)(N\kappa\eta + 1)}. \tag{10.15}$$

One can notice that the temporal channel correlation depends on *i*) the delay $|\tau|$ between the channel samples in discrete time, *ii*) the environmental parameters set $\mathcal{E}$, *iii*) the number of RIS' reflecting elements N, and *iv*) the parameter that determines the interval of the phase shifts' distribution θ. We make the following remarks about the obtained result.

Remark 10.1. Regarding κ and η, it is worth noting that, in the absence of a LoS path component from the RIS to the receiver, *i.e.*, $\kappa \to -\infty$ dB, or when the LoS components are perfectly orthogonal, *i.e.*, $\eta = 0$, the channel correlation is determined only by α and θ. On the other hand, if $\kappa > 0$ and $\eta > 0$, the correlation coefficient can also be altered by setting N.

Remark 10.2. Regarding the distribution parameter θ, we analyze how the temporal channel correlation behaves in the extreme values of its interval, $[0, \pi]$. When $\theta = 0$, the resulting *correlation coefficient* from (10.15) simplifies to:

$$\rho[\tau]|_{\theta=0} = \frac{N\kappa\eta + \alpha^{|\tau|}}{N\kappa\eta + 1}. \tag{10.16}$$

In this case, note that one can control the temporal correlation *only* by selecting the number of RIS' elements N. On the other hand, when $\theta = \pi$, the correlation coefficient becomes:

$$\rho[\tau]|_{\theta=\pi} = \delta[\tau]. \tag{10.17}$$

Now, observe that the channel samples are totally uncorrelated, corroborating with the findings of [8]. Recall that the authors of [8] used $\phi'_n[k] \sim \mathcal{U}(0, 2\pi)$ to transform a slow-fading channel into a fast-fading one, improving reliability-related metrics. Hence, by tuning θ and given N, we can control the temporal correlation to values in the interval $[\rho[\tau]|_{\theta=\pi}, \rho[\tau]|_{\theta=0}]$.

10.4.2 Controlling the Temporal Correlation

Based on Remark 10.2, we present a method to design the phase shifts distribution parameter θ to obtain the desired channel correlation between samples separated from each other by a desired delay. This is based on the following:

Definition 10.2. (Project Requirements) The tuple of *project requirements* is $p = (\tilde{\rho}, \tilde{\tau})$, where $0 \leq \tilde{\rho} \leq 1$ is the desired correlation coefficient of two-channel samples delayed by $\tilde{\tau} \in \mathbb{Z}_+$ samples.

Method. From eq. (10.15) and for a constant N, the value of θ for obtaining a channel correlation coefficient $\rho[\pm\tilde{\tau}] = \tilde{\rho}$ is:

$$\theta = \underline{\theta}(p) = \operatorname{sinc}^{-1}\left(\sqrt{\frac{(\kappa+1)\tilde{\rho}}{\tilde{\rho}\kappa + (1-\tilde{\rho})N\kappa\eta + \alpha^{|\tilde{\tau}|}}}\right), \tag{10.18}$$

where $\text{sinc}^{-1}(\cdot)$ is the inverse function of $\text{sinc}(\cdot)$ with codomain over the interval $[0, \pi]$.[7] By taking into account that $\theta \in [0, \pi]$, the argument in the right-hand side (RHS) of (10.18) must lie in the interval $[0, 1]$. Given this, we define the set of feasible project requirements as:

$$\mathcal{P}_{\text{feas.}}^{(\theta)} = \left\{ (\tilde{\rho}, \tilde{\tau}) \in \mathbb{R}_+ \times \mathbb{Z}_+ \,\middle|\, 0 \leq \tilde{\rho} \leq \frac{N\kappa\eta + \alpha^{|\tilde{\tau}|}}{N\kappa\eta + 1} \right\}. \tag{10.19}$$

From the above, one can note that the feasible channel correlation is upper-bounded by the environmental parameters $\mathcal{E}$ and the number of RIS reflecting elements N.

Remark 10.3. In Remark 10.1, we have observed that one can also control the correlation coefficient by changing N. We now give an additional result showing the achievable channel correlation when opting for designing N. We assume that the RIS, Tx, and Rx are positioned in a way that η does not vary with N.[8] Considering a constant θ, the value of N for obtaining a correlation coefficient $\rho[\pm\tilde{\tau}] = \tilde{\rho}$ is:

$$\underline{N}(p) = \left\lfloor \frac{(1 - \text{sinc}^2(\theta))(\kappa + 1) + \text{sinc}^2(\theta)(\tilde{\rho} - \alpha^{|\tilde{\tau}|})}{\text{sinc}^2(\theta)(1 - \tilde{\rho})\kappa\eta} \right\rfloor, \tag{10.20}$$

with $N = \underline{N}(p)$. Knowing that $N \geq 1$ and that the denominator of the argument at the RHS of (10.20) must be nonzero, the set of feasible project requirements can be derived as:

$$\mathcal{P}_{\text{feas.}}^{(N)} = \{ (\tilde{\rho}, \tilde{\tau}) \in \mathbb{R}_+ \times \mathbb{Z}_+ | \, \tilde{\rho}_{\min} \leq \tilde{\rho} < 1 \}, \text{ where} \tag{10.21}$$

$$\tilde{\rho}_{\min} = \frac{\text{sinc}^2(\theta)(\kappa\eta + \alpha^{|\tilde{\tau}|}) - (1 - \text{sinc}^2(\theta))(\kappa + 1)}{\text{sinc}^2(\theta)(\kappa\eta + 1)}. \tag{10.22}$$

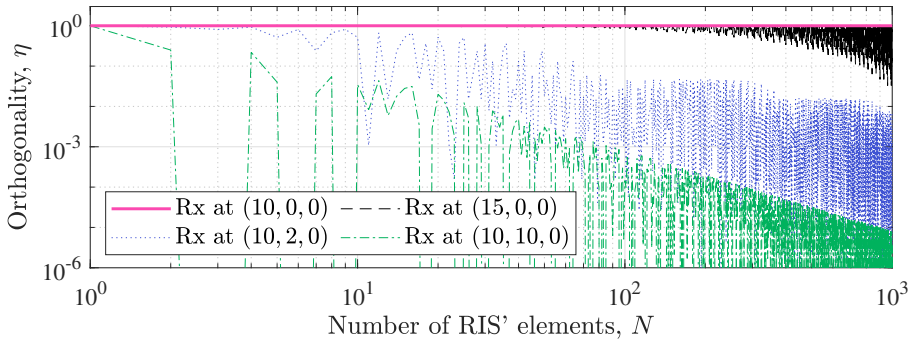
From this, one can note that the achievable channel correlation obtained by setting the number of RIS' elements N is lower-bounded by the environmental parameters $\mathcal{E}$ and θ.

To get an overview of the condition in which Remark 10.3 is valid, Fig. 32 depicts how η changes with N considering different Rx positions. The LoS channel vectors are calculated with the model in [13]. Considering the *right-handed Cartesian* coordinate system, the RIS is placed parallel to the xy-plane with center at coordinates $(0, 0, 5)$, while the Tx is at coordinates $(-10, 0, 0)$. When the Rx position is symmetric to the Tx one w.r.t. the RIS center, $\eta = 1$ is constant. So Remark 10.3 is valid for position $(10, 0, 0)$. In position $(15, 0, 0)$, it may be valid for $N < 100$ due to the low variation of η in this region. However, it does not apply for positions $(10, 2, 0)$ and $(10, 10, 0)$ due to the high-amplitude oscillations of η with N.

[7] In the domain $[0, \pi]$, the $\text{sinc}(\cdot)$ function is partially invertible since it becomes bijective. In the absence of a closed-form expression for $\text{sinc}^{-1}(\cdot)$, numerical methods can be used to calculate it with the required precision.

[8] We leave the analysis of the case where η varies with N for future works.

Figure 32 – Orthogonality between the LoS components as a function of N.



Source: The authors.

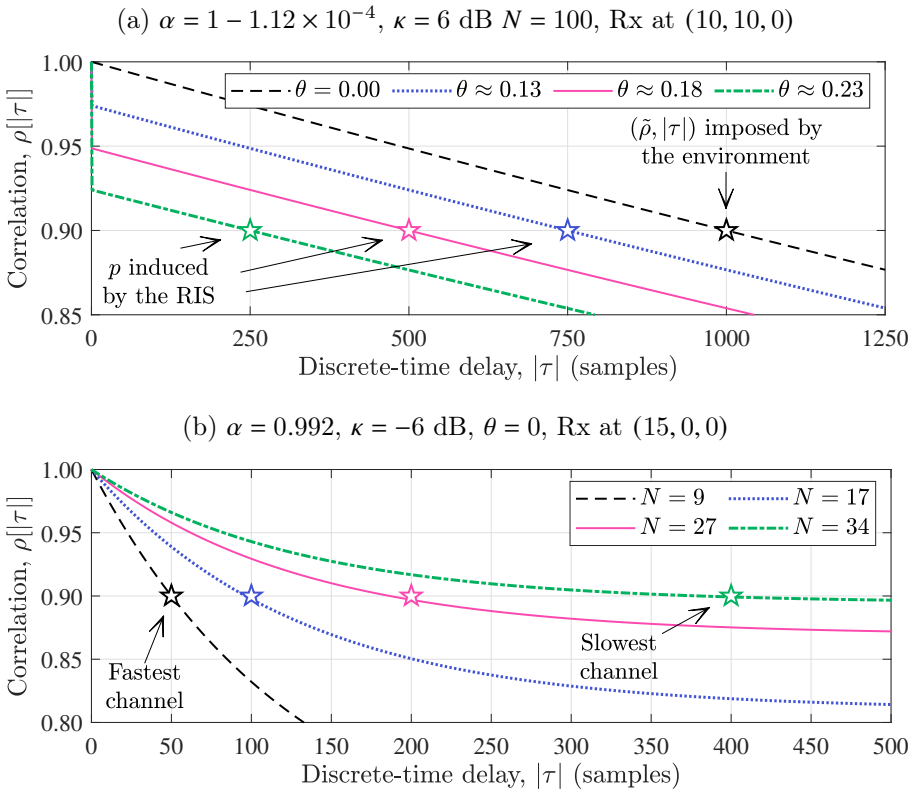
10.5 Simulation Results

In this section, we exemplify by numerical results how the method proposed in Subsection 10.4.2 can be applied to obtain a given project requirement $p = (\tilde{\rho}, \tilde{\tau})$. The results and their respective simulation parameters are given in Figs. 33 and 34. In the simulations, the coordinate system, the RIS and Tx positions, and the method to compute the LoS channel components are the same as in Fig. 32.

Fig. 33a depicts the correlation achieved by different θ with a fixed $N = 100$. From (10.11), recall that $\theta = 0$ implies static RIS phase shifts equal to π. Under this condition, α in this result is calculated by eq. (10.16) to obtain $\rho[1000]|_{\theta=0} = 0.9$. For the cases where $\theta > 0$, the phase shifts distribution parameter θ is calculated by eq. (10.18) to obtain a correlation of $\tilde{\rho} = 0.9$ at the time delays $\tilde{\tau} \in \{250, 500, 750\}$. This result shows that the proposed method can change the temporal channel statistics imposed by the environment. In the sequel, Fig. 33b depicts the correlation obtained by different N. In this result, α is calculated using the method above, but now to obtain $\rho[50]|_{\theta=0,N=9} = 0.9$. The values for $N > 9$ are calculated by eq. (10.20) to yield a correlation of $\tilde{\rho} = 0.9$ at the time delays $\tilde{\tau} \in \{100, 200, 400\}$. Such a result reveals that a fast channel, *i.e.*, a channel with a fast decay correlation, can be slowed down by increasing the number of elements of the RIS.

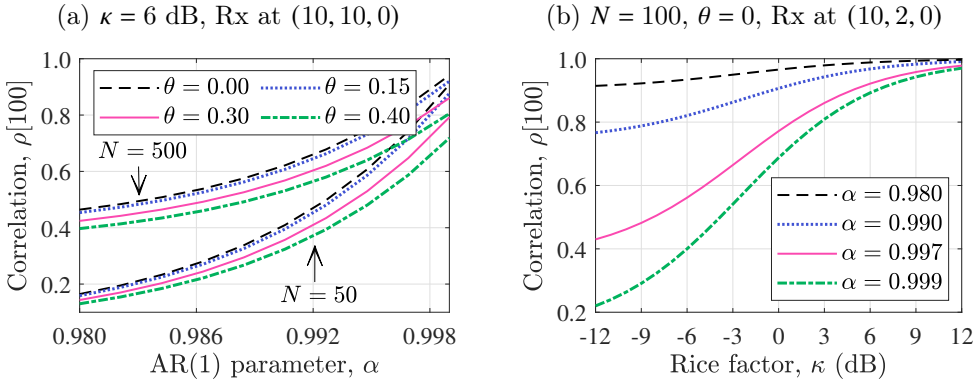
Fig. 34a shows the impact of θ and N in modifying the correlation at $|\tau| = 100$ as a function of α, representing different environmental conditions. It demonstrates that the correlation changes quickly with little change in α, justifying the use of the AR(1) model to represent both slow- and fast-fading channels. Then, Fig. 34b depicts the correlation at $|\tau| = 100$ as a function of κ and under different α. This result reveals that high κ yields correlation values close to 1 due to the significant increase in the power of the deterministic part of the channel $\bar{\mathbf{h}}$ relatively to the stochastic ones $\{\check{h}_n\}_{n=1}^{N}$. In other words, as expected, the higher the κ parameter, the lower the impact of the RIS in controlling the environment. It is worth mentioning that low κ is typical in scenarios with partially blocked LoS and/or environments with rich scattering.

Figure 33 – Channel temporal correlation. The markers indicate the points where the correlation should reach 0.9 according to the project requirements p.



Source: The authors.

Figure 34 – Channel temporal correlation as a function of (a) α and (b) κ.



Source: The authors.

10.6 Towards a Flexible Block-Fading Model

In this section, we extend the classical block-fading model [14, 15] to account for the channel correlation control based on the randomized framework proposed in Section 10.4. Define a *coherence block* as a resource block consisting of a number of subcarriers and time samples where the equivalent channel response z can be approximated as constant and flat-fading. Specifically, each coherence block has $\Delta_c = B_c T_c$ complex-valued samples,

where $T_c > 0$ is the channel coherence time. Moreover, this discrete-time channel's channel response (power gain) z follows a given distribution $z \sim f_Z$. For example, for Rayleigh fading channels, f_Z is exponential. The RIS-enabled control discussed in Section 10.4 can be used to create coherent blocks with different lengths, where the coherence time relates to the discrete-time interval $\tilde{\tau}$. Recall that the channel coherence time is defined as the range of time span values over which the channel ACF is approximately nonzero [15]. Therefore, using the RIS to shape the ACF of the equivalent channel is a path to control T_c and, consequently, change Δ_c. This generation of coherence blocks with a flexible number of samples can be done by setting θ and N to obtain a project requirement p as described respectively by eqs. (10.18) and (10.20). Particularly, this flexible block-fading model can improve how the resources are leveraged, enabling the on-demand creation of blocks according to the availability of services with different performance requirements over time.

10.7 Conclusion

In this paper, we have studied how a RIS can change the temporal statistics of the wireless propagation channel by analyzing the correlation among channel samples using the introduced discrete-time-varying channel model. Then, we proposed a randomized framework to control the relative channel coherence time by setting the number of RIS' elements and designing the distribution of their reflection coefficients, whose effectiveness is corroborated by simulation results. Our results demonstrate the possibility of redefining the resource allocation problem as we know it today by creating a flexible block-fading model based on the proposed framework.

10.8 Appendix A: Proof of the ACF of the Equivalent Channel

Recalling that $h_{\mathrm{D}}[k]$, $\mathbf{g}[k]$, $\mathbf{h}[k]$, and $\mathbf{\Psi}[k]$ are mutually independent, the ACF of eq. (10.1) is given by the sum:

$$R_{h_{\mathrm{eq}}h_{\mathrm{eq}}}[\tau] = \mathbb{E}\{h_{\mathrm{eq}}[k]h_{\mathrm{eq}}^*[k-\tau]\} = S_1 + S_2 + S_3, \text{ where} \tag{10.23}$$

$$S_1 = \mathbb{E}\{h_{\mathrm{D}}[k]h_{\mathrm{D}}^*[k-\tau]\}, \tag{10.24}$$

$$S_2 = \mathbb{E}\{(\mathbf{g}^{\mathsf{T}}[k]\mathbf{\Psi}[k]\mathbf{h}[k])(\mathbf{g}^{\mathsf{T}}[k-\tau]\mathbf{\Psi}[k-\tau]\mathbf{h}[k-\tau])^*\}, \text{ and} \tag{10.25}$$

$$S_3 = \mathbb{E}\{h_{\mathrm{D}}[k](\mathbf{g}^{\mathsf{T}}[k-\tau]\mathbf{\Psi}[k-\tau]\mathbf{h}[k-\tau])^*\} + \tag{10.26}$$

$$+ \mathbb{E}\{h_{\mathrm{D}}^*[k-\tau](\mathbf{g}^{\mathsf{T}}[k]\mathbf{\Psi}[k]\mathbf{h}[k])\}.$$

Now we evaluate each term independently. Initially, due to the model in (10.2) adopted for $h_{\mathrm{D}}[k]$, S_1 can be rewritten as:

$$\begin{aligned} S_1 &= \mathbb{E}\{(\bar{h}_{\mathrm{D}} + \check{h}_{\mathrm{D}}[k])(\bar{h}_{\mathrm{D}} + \check{h}_{\mathrm{D}}[k-\tau])^*\}, \\ &= |\bar{h}_{\mathrm{D}}|^2 + \mathbb{E}\{\check{h}_{\mathrm{D}}[k]\check{h}_{\mathrm{D}}^*[k-\tau]\} + \mathbb{E}\{\bar{h}_{\mathrm{D}}\check{h}_{\mathrm{D}}^*[k-\tau]\} + \mathbb{E}\{\bar{h}_{\mathrm{D}}^*\check{h}_{\mathrm{D}}[k]\}. \end{aligned} \tag{10.27}$$

Since $\bar{h}_{\mathrm{D}}$ is deterministic and $\check{h}_{\mathrm{D}}[k]$ is an AR(1) random process as per (10.3), the expectations are given respectively by:

$$\mathbb{E}\{\check{h}_{\mathrm{D}}[k]\check{h}_{\mathrm{D}}^*[k-\tau]\} = \alpha_{h_{\mathrm{D}}}^{|\tau|}\sigma_{h_{\mathrm{D}}}^2, \tag{10.28}$$

$$\mathbb{E}\{\bar{h}_{\mathrm{D}}\check{h}_{\mathrm{D}}^*[k-\tau]\} = \mathbb{E}\{\bar{h}_{\mathrm{D}}^*\check{h}_{\mathrm{D}}[k]\} = 0. \tag{10.29}$$

So, S_1 can be rewritten as:

$$S_1 = |\bar{h}_{\mathrm{D}}|^2 + \alpha_{h_{\mathrm{D}}}^{|\tau|}\sigma_{h_{\mathrm{D}}}^2. \tag{10.30}$$

Expanding the multiplications in S_2 results in

$$S_2 = \textstyle\sum_{n=1}^{N}\sum_{n'=1}^{N} P_1P_2P_3, \text{ where} \tag{10.31}$$

$$P_1 = \mathbb{E}\{g_n[k]g_{n'}[k-\tau]\}, \tag{10.32}$$

$$P_2 = \mathbb{E}\{h_n[k]h_{n'}[k-\tau]\}, \tag{10.33}$$

$$P_3 = \mathbb{E}\{\psi_n[k]\psi_{n'}[k-\tau]\} = R_{\psi_n\psi_{n'}}[\tau], \tag{10.34}$$

with $R_{\psi_n\psi_{n'}}$ defined in eq. (10.8). From the model in eq. (10.2) adopted for $\mathbf{g}[k]$ and $\mathbf{h}[k]$, and recalling that $\{g_n[k]\}_{n=1}^N$ and $\{h_n[k]\}_{n=1}^N$ are mutually independent, P_1 and P_2 result

$$P_1 = \begin{cases} |\bar{g}_n|^2 + \alpha_{g_n}^{|\tau|}\sigma_{g_n}^2, & \text{if } n = n' \\ \bar{g}_n\bar{g}_{n'}^*, & \text{otherwise} \end{cases}, \tag{10.35}$$

$$P_2 = \begin{cases} |\bar{h}_n|^2 + \alpha_{h_n}^{|\tau|}\sigma_{h_n}^2, & \text{if } n = n' \\ \bar{h}_n\bar{h}_{n'}^*, & \text{otherwise} \end{cases}. \tag{10.36}$$

While the results for $n = n'$ are based on the derivation of S_1, the results for $n \neq n'$ come from the mean of the time-variant channel coefficient in eq. (10.4). Therefore, given the results for P_1, P_2, and P_3, we can rewrite S_2 as:

$$\begin{aligned} S_2 = {} & \textstyle\sum_{n=1}^{N}\sum_{n'=1,n'\neq n}^{N} \bar{g}_n\bar{g}_{n'}^*\bar{h}_n\bar{h}_{n'}^* R_{\psi_n\psi_{n'}}[\tau] + \\ & + \textstyle\sum_{n=1}^{N}(|\bar{g}_n|^2 + \alpha_{g_n}^{|\tau|}\sigma_{g_n}^2)(|\bar{h}_n|^2 + \alpha_{h_n}^{|\tau|}\sigma_{h_n}^2)R_{\psi_n\psi_n}[\tau]. \end{aligned} \tag{10.37}$$

In S_3, the expectation of each multiplication can be rewritten as the multiplication of the expected value of each term:

$$S_3 = \bar{h}_{\mathrm{D}}\bar{\mathbf{g}}^{\mathsf{H}}\mathbb{E}\{\mathbf{\Psi}^*[k-\tau]\}\bar{\mathbf{h}}^* + \bar{h}_{\mathrm{D}}^*\bar{\mathbf{g}}^{\mathsf{T}}\mathbb{E}\{\mathbf{\Psi}[k]\}\bar{\mathbf{h}}. \tag{10.38}$$

Expanding the results of the multiplications in the summations:

$$S_3 = \bar{h}_{\mathrm{D}} \sum_{n=1}^{N} \bar{g}_n^* \bar{h}_n^* \mathbb{E}\{\psi_n^*[k-\tau]\} + \bar{h}_{\mathrm{D}}^* \sum_{n=1}^{N} \bar{g}_n \bar{h}_n \mathbb{E}\{\psi_n[k]\}. \tag{10.39}$$

Considering that $\{\psi_n[k]\}_{n=1}^{N}$ are wide-sense stationary (WSS) random processes, S_3 can be rewritten as:

$$S_3 = 2\Re\{\bar{h}_{\mathrm{D}}^* \sum_{n=1}^{N} \bar{g}_n \bar{h}_n \mathbb{E}\{\psi_n[k]\}\}. \tag{10.40}$$

Finally, substituting eqs. (10.30), (10.37), and (10.40) into eq. (10.23), we obtain eq. (10.7), completing the proof. □

10.9 References

[1] E. Björnson *et al.*, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 135–158, Feb. 2022.

[2] Y. Chen *et al.*, "Robust transmission for reconfigurable intelligent surface aided millimeter wave vehicular communications with statistical CSI," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 2, pp. 928–944, Feb. 2022.

[3] Y. Zhang *et al.*, "Channel aging-aware precoding for RIS-aided multi-user communications," *IEEE Trans. on Veh. Technol.*, Sept. 2022, early access.

[4] W. Jiang and H. D. Schotten, "Performance impact of channel aging and phase noise on intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 347–351, Jan. 2023.

[5] A. Papazafeiropoulos *et al.*, "Impact of channel aging on reconfigurable intelligent surface aided massive MIMO systems with statistical CSI," *IEEE Trans. on Veh. Technol.*, vol. 72, no. 1, pp. 689–703, Jan 2023.

[6] B. Matthiesen *et al.*, "Intelligent reflecting surface operation under predictable receiver mobility: A continuous time propagation model," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 216–220, Feb. 2021.

[7] S. Sun and H. Yan, "Small-scale spatial-temporal correlation and degrees of freedom for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2698–2702, Dec. 2021.

[8] K.-L. Besser and E. A. Jorswieck, "Reconfigurable intelligent surface phase hopping for ultra-reliable communications," *IEEE Trans. on Wireless Commun.*, vol. 21, no. 11, pp. 9082–9095, Nov. 2022.

[9] A. Kurt *et al.*, "An adaptive-iterative nonlinear interference cancellation in time-varying full-duplex channels," *IEEE Trans. on Veh. Technol.*, Sept. 2022, early access.

[10] K. Baddour and N. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, July 2005.

[11] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *Journal of Commun. and Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[12] H. S. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. on Veh. Technol.*, vol. 45, no. 2, pp. 353–357, May 1996.

[13] A. Albanese *et al.*, "MARISA: A self-configuring metasurfaces absorption and reflection solution towards 6G," in *41st IEEE Conf. on Comput. Commun.*, 2-5 May 2022, pp. 250–259.

[14] E. Björnson *et al.*, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. and Trends® in Signal Process.*, vol. 11, no. 3-4, pp. 154–655, 2017.

[15] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge University Press, 2005.

11 Paper E: Uplink Multiplexing of eMBB/URLLC Services Assisted by Reconfigurable Intelligent Surfaces

João Henrique Inacio de Souza, Victor Croisfelt, Radosław Kotaba, Taufik Abrão, and Petar Popovski

This paper has been submitted to the *IEEE Communications Letters* (IF: 4.1, Qualis: A1-Eng. IV), Jan. 2024, pp. 1-5. Pre-print available at `https://arxiv.org/abs/2305.04629`.

Abstract

This letter proposes a scheme assisted by a reconfigurable intelligent surface (RIS) for efficient uplink traffic multiplexing between enhanced mobile broadband (eMBB) and ultra-reliable-low-latency communication (URLLC). The scheme determines two RIS configurations based only on the eMBB channel state information (CSI) available at the base station (BS). The first optimizes eMBB quality of service, while the second reduces eMBB interference in URLLC traffic by temporarily silencing the eMBB traffic. Numerical results demonstrate that this approach, relying solely on eMBB CSI and without BS coordination, can outperform the state-of-the-art preemptive puncturing by 4.9 times in terms of URLLC outage probability.

Index Terms: Reconfigurable intelligent surface (RIS), enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and multiplexing.

11.1 Introduction

The 5G of mobile networks have been launched with features to support heterogeneous services with different requirements of QoS and traffic characteristics [1]. In particular, eMBB services require extremely high SE, while URLLC services demand high reliability and low latency [2]. Due to these very different QoSs, there is a need for multiplexing strategies that perform effectively in challenging channel conditions and better accommodate heterogeneous traffic from both eMBB and URLLC services. In particular, 5G NR has introduced the *preemptive puncturing scheme*, which relies on BS coordination to

interrupt the eMBB traffic and prioritizes the URLLC transmissions [3, 4]. Despite its fair DL performance, in the UL, such a strategy increases the URLLC latency, mainly due to the waiting time for the BS to grant scheduling responses.

Our work is motivated by the observation that an RIS can help to simultaneously deliver the QoSs of heterogeneous services through the proposal of new multiplexing schemes, whose study has received less attention in the research literature. An RIS is a thin sheet of composite material that can cover, *e.g.*, parts of walls and buildings. It can reflect incident signals to desired directions by dynamically configuring the phase shifts of the many elements that compose it [5]. In the literature, works like [6–8] study the fundamental limits and optimize the QoS of RIS-assisted URLLC applications. However, the multiplexing of eMBB/URLLC services in RIS-aided systems gives rise to two additional challenges [4]. First, the BS is unaware of the time of arrival of a URLLC packet and cannot reliably estimate its CSI to tailor the RIS configuration [2]. Second, controlling the RIS adds extra overhead that can violate the URLLC latency requirements.

In this letter, we propose a new RIS-assisted multiplexing scheme to support heterogeneous eMBB and URLLC UL traffic. Rather than depending on BS coordination, our scheme is based on the assumption that the RIS is equipped with an antenna, and the RIS is then capable of minimally processing the signal received by this antenna to detect URLLC traffic. Then, inspired by the preemptive puncturing of 5G NR, the RIS multiplexes the services using two configurations, which rely solely on eMBB CSI. The first configuration assists the eMBB UE by optimizing its signal strength and thus its SE. The second, motivated by [9], is designed to assist the URLLC UE (if detected) by temporarily silencing the eMBB interference. The proposed scheme is compared to benchmarks in terms of outage probability for the URLLC service and SE for the eMBB. Numerical results show that the proposed scheme can outperform the outage probability of preemptive puncturing by 4.9 times on average.

11.2 System Model

We consider a narrowband UL channel of a wireless system with one single-antenna BS, one hybrid RIS, and two single-antenna UEs, where the BS controls the RIS via an out-of-band control channel [10]. One of the UE uses the eMBB service, while the other uses the URLLC one.[1] Henceforth, we index the UEs by $\iota \in \{\mathrm{e}, \mathrm{u}\}$, where e and u refer to eMBB and URLLC, respectively. We assume an industrial scenario where the LoS paths between the BS and the UEs are blocked by obstacles. Thus, the RIS is deployed

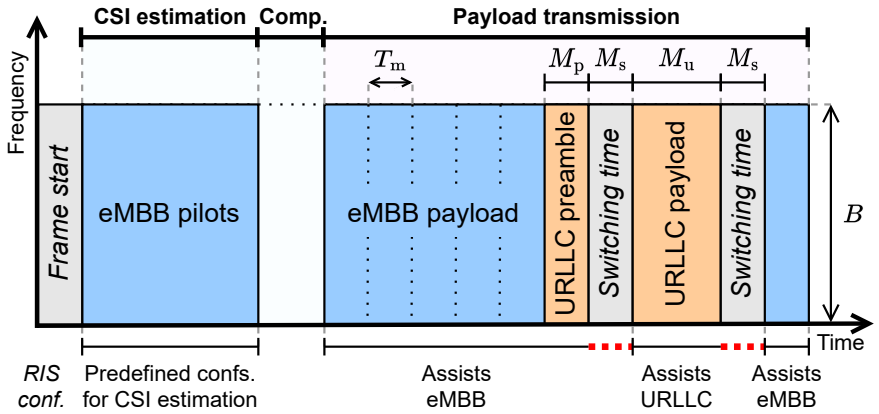
[1] To attain efficient multi-user communication, it is common to multiplex several UEs of each traffic type in a wideband channel. However, since we focus on how such multiplexing can be done, we consider a single UE of each traffic type in a narrowband channel, leaving the general case for a future extension of this work. Moreover, note that the resource allocation strategy is an independent problem not treated in this work.

to simultaneously have LoS to the BS and the UEs. The RIS is a square uniform planar array (UPA) of $N \in \mathbb{Z}_+$ half-wavelength spaced passive reflecting elements and one active element, placed at the center of the RIS. The RIS can perform light computational tasks by processing the signal received by its active element. Each passive element can induce a phase shift $\theta_n \in [0, 2\pi)$ to an impinging signal with marginal impact on its amplitude. On this basis, we define an RIS configuration as $\boldsymbol{\psi} = [\psi_1 \ \cdots \ \psi_N]^\intercal$, where $\psi_n = e^{-j\theta_n}$ is the reflection coefficient of the n-th RIS element.

11.2.1 Structure of a Uplink Frame

Fig. 35 depicts a UL frame, in which the time-frequency resources used by the UEs are organized into a grid. In the time domain, the length of the UL frame is $T > 0$, and it is divided into $M \in \mathbb{Z}_+$ identical mini-slots of duration $T_\mathrm{m} > 0$, which are indexed by $\mathcal{M} = \{1, \ldots, M\}$. In the frequency domain, the spectrum reserved for UL has the bandwidth $B > 0$. Due to different requirements, the UEs have different transmission time intervals (TTIs). To achieve high SE, the eMBB TTI spans over the entire frame. Conversely, the URLLC TTI spans over a limited number of $M_\mathrm{u} \in \mathbb{Z}_+$, $1 \leq M_\mathrm{u} \leq M$ contiguous mini-slots to guarantee the low-latency requirement. We denote as $\mathcal{M}_\mathrm{u} \subseteq \mathcal{M}$ with $|\mathcal{M}_\mathrm{u}| = M_\mathrm{u}$ the set of mini-slots of the URLLC TTI. Moreover, we consider that the eMBB UE is admitted and scheduled by the BS at the frame start, while the URLLC UE can start transmitting at any mini-slot.

Figure 35 – UL frame divided into CSI estimation, computing, and payload transmission.



Source: The authors.

11.2.2 Signal Model

Let the vector $\mathbf{h}_\mathrm{b} \in \mathbb{C}^N$ denotes the BS-RIS channels and the vectors $\mathbf{h}_\iota \in \mathbb{C}^{1\times N}$ denote the RIS-UEs channels. We further assume the block-fading model, *i.e.*, the channels remain constant for the entire frame duration. Therefore, the BS-UEs cascaded channels are

represented by the vector [9]

$$\mathbf{g}_\iota \triangleq \operatorname{diag}(\mathbf{h}_\iota)\,\mathbf{h}_\mathrm{b}. \tag{11.1}$$

Finally, we let the scalars $g_{\iota,n} = [\mathbf{g}_\iota]_n$ denote the BS-UEs cascaded channels that pass through the n-th RIS element. We denote as $\mathbf{x}_{\iota,m} \in \mathbb{C}^L$ the vectors with the $L \in \mathbb{Z}_+$ symbols transmitted by each UE such that $\mathbb{E}\{\|\mathbf{x}_{\iota,m}\|_2^2\} = 1$ in the m-th mini-slot. Thus, $\mathbf{x}_{\mathrm{u},m} = \mathbf{0}$ in the mini-slots with absence of URLLC traffic, that is, when $m \notin \mathcal{M}_\mathrm{u}$. Let $p_\iota > 0$ denote the UEs' transmit power for one mini-slot. Considering the RIS configuration $\boldsymbol{\psi}$, the received signal $\mathbf{y}_m(\boldsymbol{\psi}) \in \mathbb{C}^L$ at the BS is:

$$\mathbf{y}_m(\boldsymbol{\psi}) \triangleq \sqrt{p_\mathrm{u}}(\mathbf{g}_\mathrm{u}^\mathsf{H}\boldsymbol{\psi})\mathbf{x}_{\mathrm{u},m} + \sqrt{p_\mathrm{e}}(\mathbf{g}_\mathrm{e}^\mathsf{H}\boldsymbol{\psi})\mathbf{x}_{\mathrm{e},m} + \mathbf{w}_m, \tag{11.2}$$

where $\mathbf{w}_m \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is the AWGN.[2] Using (11.2), the SNR for the eMBB UE in the mini-slots without URLLC traffic is:

$$\Gamma_{\mathrm{e},m}(\boldsymbol{\psi}) \triangleq \frac{p_\mathrm{e}|\mathbf{g}_\mathrm{e}^\mathsf{H}\boldsymbol{\psi}|^2}{\sigma^2}, \forall m \in \mathcal{M}\backslash\mathcal{M}_\mathrm{u}. \tag{11.3}$$

Also, during the URLLC TTI, the SINR for the URLLC UE is given by

$$\Gamma_{\mathrm{u},m}(\boldsymbol{\psi}) \triangleq \frac{p_\mathrm{u}|\mathbf{g}_\mathrm{u}^\mathsf{H}\boldsymbol{\psi}|^2}{p_\mathrm{e}|\mathbf{g}_\mathrm{e}^\mathsf{H}\boldsymbol{\psi}|^2 + \sigma^2}, \forall m \in \mathcal{M}_\mathrm{u}. \tag{11.4}$$

With (11.3) and (11.4), one can measure the quality of the radio links established for the UEs by the RIS.

11.3 RIS-Assisted Multiplexing Scheme

Here, we introduce our proposed multiplexing scheme. First, we explain the three phases comprising a UL frame and give an overview of how our scheme works. Then, we discuss how the RIS could detect the URLLC traffic, and we parameterize the performance of an arbitrary detector by its miss detection rate, ϵ_m. Finally, to multiplex the eMBB and URLLC services, we design two RIS configurations that rely only on eMBB CSI.

11.3.1 Phases of the Multiplexing Scheme

From Fig. 35, a UL frame is divided into three phases. **1) CSI Estimation:** The BS estimates the eMBB CSI by using pilot symbols sent by the eMBB UE, while the RIS switches among predefined configurations (for more details, see [5]). **2) Computing:** The BS computes two RIS configurations based on the eMBB CSI, the *eMBB-oriented configuration* and the *URLLC-oriented configuration* to potentially multiplex the eMBB

[2] Following the approach in [4, 11, 12], we focus on the RIS configuration by assuming a single-antenna BS. In the case of a multiple-antenna BS, eq. (11.2) can be seen as the received signal after linear processing, where the effect of the beamformer applied by the BS is embedded into the vector of cascaded channels, $\mathbf{g}_\iota$. Hence, the proposed model and scheme are general.

and URLLC services within the given frame. Finally, these two configurations are sent to the RIS to be stored there for potential use in the next phase. **3) Payload Transmission:** In this phase, the scheduled eMBB UE transmits its payload whereas the RIS starts by being configured with the eMBB-oriented configuration. At any time in this phase, if the URLLC UE has a packet to send, it will start its TTI by sending a preamble over $M_{\mathrm{p}} \in \mathbb{Z}_{+}$ mini-slots to announce its intention to transmit payload data. By exploiting its active element[3] and the URLLC preamble, the RIS can *detect* the start of the URLLC traffic. Suppose URLLC traffic is detected. In that case, the RIS switches to the URLLC-oriented configuration to multiplex the URLLC TTI; otherwise, it keeps the eMBB-oriented configuration. In the former case, when the URLLC TTI is over, the RIS switches back to the eMBB-oriented configuration. Note that while all this occurs, the eMBB TTI is not interrupted, so part of the eMBB mini-slots will be erased by the URLLC traffic, and by the changes in the environment caused by the RIS.[4]

In practice, the RIS can impose a physical latency due to switching configurations. By assuming that the time the RIS takes to change its phase shifts is around some microseconds [5], we consider the RIS takes $M_{\mathrm{s}} \in \mathbb{Z}_{+}$ mini-slots to switch between configurations. We also assume that the URLLC UE is aware of this *switching time* and waits for it before transmitting the payload. Hence, in case the URLLC traffic is detected, the RIS adds an overhead of M_{s} mini-slots for the URLLC UE and $2M_{\mathrm{s}}$ mini-slots for the eMBB one.

11.3.2 Detecting the URLLC Traffic

The URLLC traffic can be identified at the RIS using the preamble signaling to span over M_{p} mini-slots. To do so, one effective approach would involve the joint design of the preamble and detection algorithm to meet URLLC QoSs. For instance, in the context of URLLC service, a robust scheme presented in [13] introduces a brief preamble consisting of a single orthogonal frequency-division multiplexing (OFDM) symbol, detected at the receiver using differential detection. This algorithm can be seamlessly implemented in a hybrid RIS with limited signal processing capacity. However, we investigate the performance of an arbitrary detection scheme, assessing its general impact on multiplexing efficiency without relying on a specific scheme. We parameterize the performance of this detector by its *miss detection rate*, defined as $0 \leq \epsilon_{\mathrm{m}} \leq 1$, whose impact is evaluated in Section 11.5.

11.3.3 Computing the eMBB-Oriented RIS Configuration

The eMBB-oriented RIS configuration $\boldsymbol{\psi}^{\mathrm{e}} \in \mathbb{C}^{N}$ is set to maximize the eMBB SNR, which can be posed as the problem

[3] In practice, this active element could also be the antenna of the communication interface used to control the RIS [5], making it a viable solution.

[4] To decode the eMBB payload data at the BS with a limited number of erased mini-slots, erasure coding can be applied by the eMBB UE [2].

$$\max_{\boldsymbol{\psi}\in\mathbb{C}^N} \frac{|\mathbf{g}_{\mathrm{e}}^{\mathsf{H}}\boldsymbol{\psi}|^2}{\sigma^2}, \text{ s.t. } \mathrm{C}_1 : |\psi_n| = 1, \forall n. \tag{11.5}$$

Fortunately, this problem assumes a family of closed-form solutions known as coherent passive beamformers [10]. Based on this and with eMBB CSI,[5] $\boldsymbol{\psi}^{\mathrm{e}}$ can be computed as:

$$\boldsymbol{\psi}^{\mathrm{e}} = [e^{-j\theta_1^{\mathrm{e}}} \cdots e^{-j\theta_N^{\mathrm{e}}}]^{\mathsf{T}}, \text{ where } \theta_n^{\mathrm{e}} = -\arg(g_{\mathrm{e},n}) + \bar{\theta}, \forall n, \tag{11.6}$$

and $\bar{\theta} \in [0, 2\pi)$ due to the phase periodicity. For the sake of simplicity but without compromising generality, we take $\bar{\theta} = 0$. In such a case, from (11.1), the effective channel of the eMBB UE can be simplified to $\mathbf{g}_{\mathrm{e}}^{\mathsf{H}}\boldsymbol{\psi}^{\mathrm{e}} = \sum_{n=1}^{N} |g_{\mathrm{e},n}|$. Hence, the RIS yields the maximum array gain of N^2, providing high SE.

11.3.4 Computing the URLLC-Oriented RIS Configurations

At the BS, the computation of an RIS configuration to multiplex the URLLC TTI is challenging since only eMBB CSI is available.[6] Our idea is to find a configuration that mitigates the interference caused by the eMBB traffic to the URLLC one, temporarily silencing the eMBB UE. We present two methods for computing the URLLC-oriented configuration. The first one is a heuristic based on *phasors rotation* that tries to cancel out the channel gain of the eMBB UE by compensating the phase shifts of the RIS elements via subtraction. The second one is an *alternating projection* algorithm to approximate a configuration that nulls the eMBB interference.

Method 1: Phasors Rotation (PR)

Given the eMBB CSI, let us define the problem of finding an RIS configuration that minimizes the channel gain of the eMBB UE as

$$\min_{\boldsymbol{\psi}\in\mathbb{C}^N} |\mathbf{g}_{\mathrm{e}}^{\mathsf{H}}\boldsymbol{\psi}|, \text{ s.t. } \mathrm{C}_1, \tag{11.7}$$

with C_1 as in (11.5), which makes the problem to be not convex. However, notice that the channel gain is lower-bounded such that $|\mathbf{g}_{\mathrm{e}}^{\mathsf{H}}\boldsymbol{\psi}| \geq 0$. Therefore, we aim to find a configuration that makes this gain as close to zero as possible.

We start by presenting a heuristic algorithm with low computational complexity, which yields a sub-optimal solution to problem (11.7). The heuristic is based on the representation of the eMBB cascaded channels as phasors and the idea that we can individually rotate

[5] We assume perfect CSI knowledge at the BS to show the upper bound performance of the proposed scheme, leaving the study of the impact of imperfect CSI for future work.

[6] Due to the unpredictability of its traffic, the BS cannot acquire the URLLC CSI for the computing phase. Still, in the case of coherent data detection in the subsequent payload transmission phase, the BS must estimate the CSI from, *e.g.*, the demodulation reference signal (DMRS) multiplexed in the URLLC TTI symbols [3].

them so that they cancel each other out, nulling the eMBB channel. This is done by dividing the RIS elements into two sets, where the goal of one of the sets is to eliminate the contribution of the other. From (11.1), let the n-th eMBB cascaded channel be represented by the phasor

$$A_n e^{j\omega_n} \triangleq g_{\mathrm{e},n}^{*} \psi_n, \tag{11.8}$$

where $A_n = |g_{\mathrm{e},n}|$ is the phasor's amplitude, and $\omega_n = -\arg(g_{\mathrm{e},n}) - \theta_n$ is the phasor's angle. Let the nonempty sets $\mathcal{N}_0$ and $\mathcal{N}_\pi$ denote a partition of the RIS elements, *i.e.*, $\mathcal{N}_0 \cup \mathcal{N}_\pi = \{1, \ldots, N\}$ and $\mathcal{N}_0 \cap \mathcal{N}_\pi = \emptyset$. Accordingly, the RIS elements are configured with the following phase shift

$$\theta_n^{\mathrm{u}} = \begin{cases} -\arg(g_{\mathrm{e},n}), & n \in \mathcal{N}_0 \\ \pi - \arg(g_{\mathrm{e},n}), & n \in \mathcal{N}_\pi \end{cases}. \tag{11.9}$$

Using the configuration $\boldsymbol{\psi}^{\mathrm{u}} = [e^{-j\theta_1^{\mathrm{u}}} \cdots e^{-j\theta_N^{\mathrm{u}}}]^{\intercal}$, the cascaded channels belonging to sets $\mathcal{N}_0$ and $\mathcal{N}_\pi$ are out of phase, since $\omega_n = 0$ if $n \in \mathcal{N}_0$ and $\omega_n = -\pi$ if $n \in \mathcal{N}_\pi$. Therefore, from (11.8) and (11.9), the eMBB channel gain with $\boldsymbol{\psi}^{\mathrm{u}}$ is equal to

$$|\mathbf{g}_{\mathrm{e}}^{\mathsf{H}} \boldsymbol{\psi}^{\mathrm{u}}| = |\textstyle\sum_{n=1}^{N} A_n e^{j\omega_n}| = |\sum_{n\in\mathcal{N}_0} A_n - \sum_{n'\in\mathcal{N}_\pi} A_{n'}|. \tag{11.10}$$

Thus, to mitigate the interference caused by the eMBB traffic, we need to determine the sets $\mathcal{N}_0$ and $\mathcal{N}_\pi$ that approximately null (11.10). Since each cascaded channel can belong to either $\mathcal{N}_0$ or $\mathcal{N}_\pi$, minimizing (11.10) is a combinatorial optimization problem with 2^N candidate solutions. As this is not tractable even for moderate size RISs, in Algorithm 4 we present an intuitive method for determining $\mathcal{N}_0$ and $\mathcal{N}_\pi$, approximating a solution that minimizes (11.10) in feasible computation time.

Algorithm 4 Phasors rotation algorithm.

input: The channel vector $\mathbf{g}_{\mathrm{e}}$
output: The RIS configuration $\boldsymbol{\psi}^{\mathrm{u}}$

1: $A_n \leftarrow |g_{\mathrm{e},n}|$
2: $(\alpha_i)_{i=1}^{N} \leftarrow \mathrm{sort}(A_1, \ldots, A_N)$
3: $N^* \leftarrow \arg\min \{|\sum_{i=1}^{N'} \alpha_i - \sum_{i'=N'+1}^{N} \alpha_{i'}| \mid 1 \le N' \le N\}$
4: $\mathcal{N}_0 \leftarrow \{\kappa(1), \ldots, \kappa(N^*)\}$, $\mathcal{N}_\pi \leftarrow \{\kappa(N^*+1), \ldots, \kappa(N)\}$
5: Set $\theta_n^{\mathrm{u}}, \forall n$ according to eq. (11.9)
6: **return** $\boldsymbol{\psi}^{\mathrm{u}} \leftarrow [e^{-\theta_1^{\mathrm{u}}} \cdots e^{-\theta_N^{\mathrm{u}}}]^{\intercal}$

The algorithm works as follows. Initially, the amplitudes of the phasors representing the cascaded channels are computed. Then, the algorithm finds an integer number $1 \le N^* \le N$ such that the sum of the amplitudes of the N^* shortest phasors is as close as possible to the sum of the amplitudes of the $N - N^*$ remaining ones. Finally, the set $\mathcal{N}_0$ is created with the indices of the elements associated with the N^* shortest phasors, while $\mathcal{N}_\pi$ is created with the indices of the remaining ones. In the algorithm, $\kappa(\cdot)$ maps the indices of $(\alpha_i)_{i=1}^{N}$ to $\{A_n\}_{n=1}^{N}$.

Method 2: Interference Nulling (IN)

The second method of finding an RIS configuration that nulls the eMBB interference at the BS can be cast as the following feasibility problem

$$\text{find } \boldsymbol{\psi} \in \mathbb{C}^N, \text{ s.t. } \mathrm{C}_1 \text{ and } \mathrm{C}_2 : \mathbf{g}_\mathrm{e}^\mathsf{H}\boldsymbol{\psi} = 0, \tag{11.11}$$

where C_1 is defined as in (11.5), while C_2 ensures that the eMBB traffic does not interfere at the BS. Similarly to (11.7), this problem is not convex due to the unit modulus constraints C_1. However, notice that, if any, a solution to the problem is any vector that belongs to the intersection between the sets defined by constraints C_1 and C_2. In this case, the alternating projection algorithm can approximate a solution with a high probability of convergence. In this sense, Algorithm 5 presents an adaption of the alternating projection algorithm in [9] that iteratively approximates a solution to problem (11.11). During each iteration $t \in \mathbb{Z}_+$, the vector of reflection coefficients $\boldsymbol{\psi}_{t-1}$ is projected sequentially onto the set $\{\boldsymbol{\psi} \in \mathbb{C}^N \mid \mathbf{g}_\mathrm{e}^\mathsf{H}\boldsymbol{\psi} = 0\}$, then onto the set $\{\boldsymbol{\psi} \in \mathbb{C}^N \mid |\psi_n| = 1, \forall n\}$. The projection operators consider the smallest Euclidean distance from $\boldsymbol{\psi}_{t-1}$ to a point in the projected set, derived according to [9, eq. (20)]. We use early stopping and maximum iterations as stopping criteria.

Algorithm 5 Alternating projection algorithm for interference nulling (IN).

input: The channel vector $\mathbf{g}_\mathrm{e}$ and the initial configuration $\boldsymbol{\psi}_0$
output: The RIS configuration $\boldsymbol{\psi}^\mathrm{u}$

$\mathbf{v} \leftarrow \mathbf{g}_\mathrm{e}/\|\mathbf{g}_\mathrm{e}\|_2^2$, $t \leftarrow 1$
repeat
  $\tilde{\boldsymbol{\psi}} \leftarrow \boldsymbol{\psi}_{t-1} - (\mathbf{v}^\mathsf{H}\boldsymbol{\psi}_{t-1})\mathbf{v}$
  $\boldsymbol{\psi}_t \leftarrow [\tilde{\psi}_1/|\tilde{\psi}_1| \cdots \tilde{\psi}_N/|\tilde{\psi}_N|]^\mathsf{T}$
  $t \leftarrow t + 1$
until stopping criterion is satisfied
return $\boldsymbol{\psi}^\mathrm{u} \leftarrow \boldsymbol{\psi}_{t-1}$

Fig. 36 illustrates how the configurations work by depicting the RIS power pattern produced by them. We can see that the eMBB-oriented configuration produces a strong beam centered around the direction of the eMBB UE, while the configurations obtained via phasors rotation (PR) and IN produce a null in this same direction.

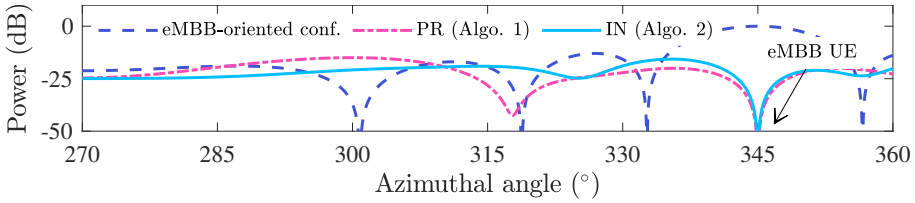
11.4 Analysis

This section introduces the metrics used to analyze the proposed multiplexing scheme.

Performance Analysis

By using the RIS configurations of Section 11.3, we present expressions for the outage probabilities achieved by the eMBB and URLLC UEs. First, considering $\boldsymbol{\psi}^\mathrm{e}$ in (11.6) and

Figure 36 – Normalized RIS power pattern produced by the proposed configuration designs. The eMBB UE is positioned in the azimuth angle of 345°.



Source: The authors.

assuming that a single URLLC TTI happens during each UL frame,[7] the average mutual information per mini-slot of the eMBB data stream at the BS is:

$$I_{\mathrm{e}}(p_{\mathrm{e}}) \triangleq (1-\xi)\log_2\left(1+\frac{p_{\mathrm{e}}(\sum_{n=1}^{N}|g_{\mathrm{e},n}|)^2}{\sigma^2}\right), \tag{11.12}$$

where the pre-log term $\xi = (M_{\mathrm{p}} + 2M_{\mathrm{s}} + M_{\mathrm{u}})/M$ accounts for the mini-slots erased by the URLLC TTI and RIS configuration switching. Similarly, the average mutual information per mini-slot of the URLLC data stream considering the RIS configuration $\boldsymbol{\psi}^{\mathrm{u}}$ computed by Algorithms 4 or 5 is derived as

$$I_{\mathrm{u}}(p_{\mathrm{u}}, p_{\mathrm{e}}) \triangleq \log_2\left(1+\frac{p_{\mathrm{u}}|\mathbf{g}_{\mathrm{u}}^{\mathsf{H}}\boldsymbol{\psi}^{\mathrm{u}}|^2}{p_{\mathrm{e}}|\mathbf{g}_{\mathrm{e}}^{\mathsf{H}}\boldsymbol{\psi}^{\mathrm{u}}|^2+\sigma^2}\right). \tag{11.13}$$

Therefore, given constant transmit power, the outage probabilities of the eMBB and URLLC UEs as functions of the SEs per mini-slot, $r_\iota > 0$, are respectively equal to

$$P_{\mathrm{e}}(r_{\mathrm{e}}) \triangleq \Pr(I_{\mathrm{e}}(p_{\mathrm{e}}) < r_{\mathrm{e}}) \text{ and } P_{\mathrm{u}}(r_{\mathrm{u}}) \triangleq \Pr(I_{\mathrm{u}}(p_{\mathrm{u}}, p_{\mathrm{e}}) < r_{\mathrm{u}}).$$

Latency Analysis

From Fig. 35, the latency introduced by the proposed scheme to the URLLC traffic is governed by: **1)** the transmission of the URLLC preamble, **2)** the delay due to processing of the preamble and switching between configurations at the RIS, and **3)** the transmission of the URLLC payload symbols. Neglecting the propagation delay, the latency introduced by the transmission of the URLLC preamble and payload are respectively $M_{\mathrm{p}}T_{\mathrm{m}}$ and $M_{\mathrm{u}}T_{\mathrm{m}}$. Moreover, we let $D_{\mathrm{proc}} > 0$ be a constant that accounts for the time needed to process the preamble at the RIS and detect the URLLC traffic. Also, recall that the delay for the RIS to switch to the URLLC-oriented configuration is $M_{\mathrm{s}}T_{\mathrm{m}}$. Then, the URLLC latency is:

$$D = M_{\mathrm{p}}T_{\mathrm{m}} + M_{\mathrm{u}}T_{\mathrm{m}} + D_{\mathrm{proc}} + M_{\mathrm{s}}T_{\mathrm{m}}. \tag{11.14}$$

[7] If the URLLC UE has multiple TTIs in a single frame, such that the total duration is in the same order as the eMBB TTI, allocating orthogonal resources to each service might be more efficient than multiplexing them.

If the processing delay is negligible compared to the mini-slot duration, a reasonable approximation for the URLLC latency is $D \approx (M_\mathrm{p} + M_\mathrm{s} + M_\mathrm{u})T_\mathrm{m}$. As a comparison, in the 5G NR preemptive puncturing, the latency increases in the order of the slot duration due to the BS coordination [3]. Given the transmission numerology $\mu = 2$ and two symbol mini-slots, the average URLLC latency obtained by preemptive puncturing is around 0.30 ms, against 0.17 ms for the proposed multiplexing scheme. Due to space limitations, we leave a detailed comparison for future work.

11.5 Numerical Results

Now, we present numerical simulations to discuss the performance of the proposed scheme for multiplexing eMBB and URLLC services. In the simulations, $\mathbf{h}_\mathrm{b}$ and $\mathbf{h}_\iota$ follow the LoS channel model in [14], with $\lambda = 0.1$ m, $\beta = 3.67$, $\gamma_0 = 1$, $d_0 = 1$ m, and $\sigma^2 = -90$ dBm.[8] Moreover, we consider $M_\mathrm{p} = M_\mathrm{s} = 1$ mini-slot and $M_\mathrm{u} = 2$ mini-slots. The RIS is placed at the origin of the coordinate system in the yz-plane, pointing towards the direction of the x-axis. The BS is at $\mathbf{q}_\mathrm{BS} = [\varrho_\mathrm{f}/\sqrt{2}\ \varrho_\mathrm{f}/\sqrt{2}\ 0]^\intercal$, where $\varrho_\mathrm{f} = \lambda(\sqrt{N}-1)^2/2$ is the far-field distance of the RIS. The region occupied by the UEs is a volume defined in spherical coordinates by the set $\{(\varrho, \vartheta, \varphi) \mid \varrho_\mathrm{f} \leq \varrho \leq 100\ \mathrm{m}, \vartheta_\mathrm{min} \leq \vartheta \leq \vartheta_\mathrm{max}, \frac{3\pi}{2} \leq \varphi \leq 2\pi\}$, where $(\varrho, \vartheta, \varphi)$ denotes, respectively, the radial distance, the polar, and azimuthal angles. Specifically, the angles $\vartheta_\mathrm{min}, \vartheta_\mathrm{max}$ are such that the z-coordinates of the UEs lie within $[-3, 3]$ m. The UEs' positions, drawn for each 10^7 realization, are uniformly distributed over this region.

Benchmarks: To compare the proposed algorithms for computing the URLLC-oriented configuration, we define the following benchmarks: **1)** *Random:* The phase shifts $\{\theta_n^\mathrm{u}\}_{n=1}^N$ are drawn from a uniform distribution over the interval $[0, 2\pi)$. **2)** *Missed URLLC preamble:* eMBB-oriented configuration, representing the case where the RIS always fails to detect the start of the URLLC traffic ($\epsilon_\mathrm{m} = 1$).[9] **3)** *Preemptive puncturing:* Following the procedure described in [3], the eMBB TTI is interrupted so as the URLLC UE has an interference-free TTI. Still, since there is no URLLC CSI knowledge at the BS, the RIS keeps the eMBB-oriented configuration. **4)** *Maximize URLLC SNR:* Ideally, assuming that URLLC CSI is available at the BS prior to its TTI, $\boldsymbol{\psi}^\mathrm{u}$ is set to perform coherent passive beamforming like in (11.6), but for the URLLC UE.

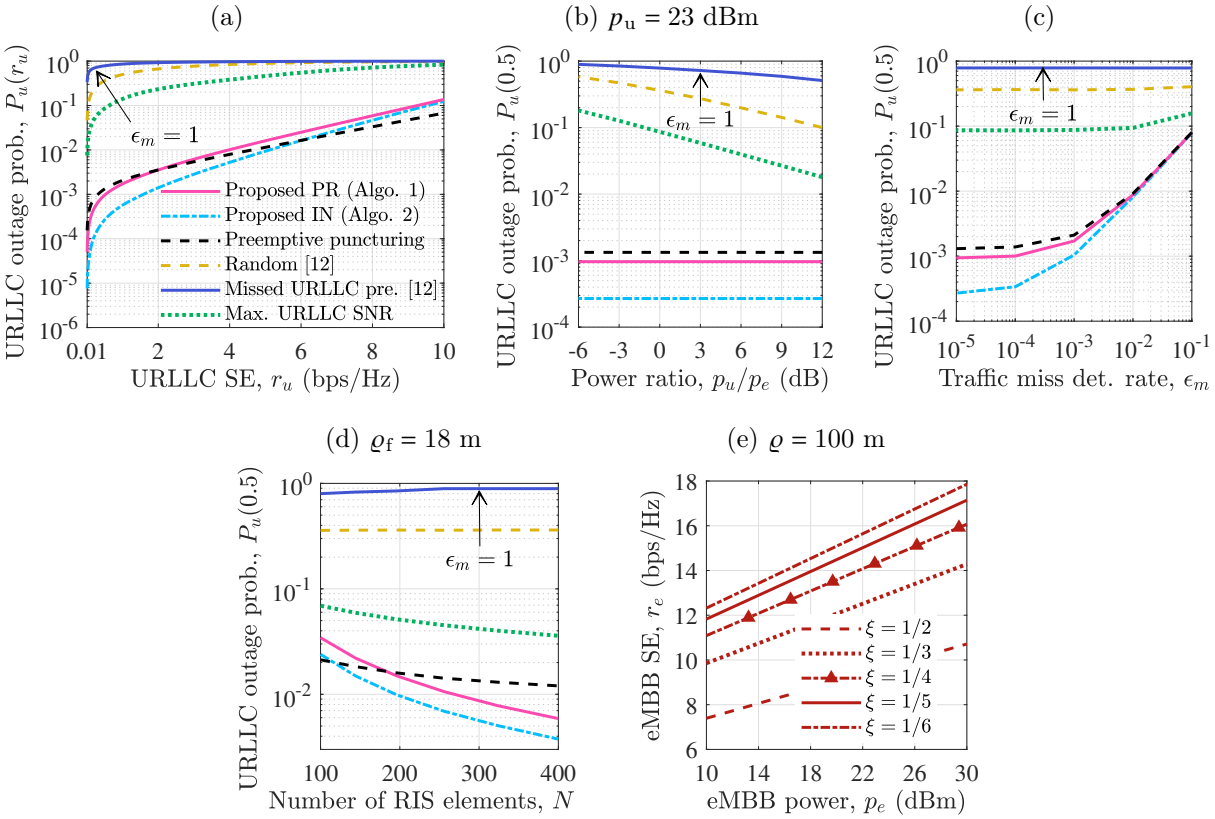
Figs. 37a and 37b present the URLLC outage probability as a function of the SE and the transmit power ratio, respectively. These results reveal the gains of the proposed PR and IN algorithms over the benchmarks, even relying only on eMBB CSI. Notice that the PR and IN performance is comparable to the preemptive puncturing, where there is no

[8] The operations of the proposed algorithms and benchmarks are independent of the channel model, applying to other, more general, models.

[9] The benchmarks **1)** and **2)** are respectively equivalent to the configurations to assist the URLLC and eMBB UEs in [12].

eMBB interference due to the BS scheduling. Remarkably, the IN algorithm outperforms the preemptive puncturing performance by 4.9 times in Fig. 37b. Observing Fig. 36, this might happen due to the several nulls produced by the eMBB-oriented configuration, which can impact the URLLC channel. Compared to the random and PR configurations, IN reduces the outage probability by up to 3 orders of magnitude and 3.5 times, respectively. However, the PR algorithm yields a better trade-off due to its simpler implementation.

Figure 37 – URLLC outage probability and eMBB SE at the cell edge. When not otherwise specified, the transmit powers are $p_{\mathrm{e}} = p_{\mathrm{u}} = 23$ dBm, the RIS has $N = 100$ elements, the URLLC traffic miss detection rate is $\epsilon_{\mathrm{m}} = 0$, and the RIS far-field distance is $\varrho_{\mathrm{f}} = 4$ m. Also, in (e), ξ is the fraction of eMBB mini-slots affected by the URLLC TTI and the RIS switching, as defined in (11.12).



Source: The authors.

Fig. 37c depicts the URLLC outage probability as a function of the URLLC traffic miss detection rate. Only for preemptive puncturing, ϵ_{m} is considered to be the failure rate for the URLLC scheduling request procedure. Notice that, for the proposed scheme, the traffic detection scheme is not considered a major limiting factor for the performance, mainly because there are simple architectures for it that yield extremely low detection error rates. For example, a scheme in [9] achieves a detection error rate of up to 10^{-7} at an SNR of 4 dB with a preamble comprised by a single OFDM symbol.

Fig. 37d depicts the URLLC outage probability as a function of the number of RIS elements. It is worth mentioning that, as the outage probability improves with a bigger

surface, the overhead for estimating the eMBB CSI increases proportionally to N. Hence, when selecting the number of RIS elements, a trade-off exists between the URLLC outage probability and the overhead in the CSI estimation phase.

Fig. 37e depicts the SE of the eMBB UE when placed at the cell edge as a function of the transmit power. Notice that, as the URLLC packet size is relatively small compared to the eMBB, the mini-slots erased by the URLLC traffic have minimal impact on the eMBB TTI, resulting in high eMBB SE.

11.6 Conclusions

In this letter, we have proposed an RIS-assisted UL multiplexing scheme to support the eMBB/URLLC coexistence. The scheme relies on two RIS configurations computed from eMBB CSI. The first configuration is a coherent passive beamformer to maximize the eMBB SNR. The second is computed by the proposed PR and IN algorithms and mitigates the eMBB interference in the URLLC traffic, temporarily silencing the eMBB traffic. Numerical results have demonstrated that the proposed scheme enables a balanced coexistence of the services in the absence of URLLC CSI. Future work might study how resource allocation can be solved while considering such a multiplexing scheme.

11.7 References

[1] M. Z. Chowdhury *et al.*, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. of the Commun. Society*, vol. 1, pp. 957–975, July 2020.

[2] P. Popovski *et al.*, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, Sept. 2018.

[3] "NR and NG-RAN overall description," *Document 3GPP TS 38.300 V17.5.0*, June 2023.

[4] M. Almekhlafi *et al.*, "Enabling URLLC applications through reconfigurable intelligent surfaces: Challenges and potential," *IEEE Internet of Things Mag.*, vol. 5, no. 1, pp. 130–135, Mar. 2022.

[5] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. on Sel. Areas in Commun.*, vol. 38, no. 11, pp. 2450–2525, July 2020.

[6] H. Xie *et al.*, "User grouping and reflective beamforming for IRS-aided URLLC," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2533–2537, Nov. 2021.

[7] Z. Li *et al.*, "Resource allocation for IRS-assisted uplink URLLC systems," *IEEE Commun. Lett.*, Apr. 2023, early access.

[8] H. Ren *et al.*, "Intelligent reflecting surface-aided URLLC in a factory automation scenario," *IEEE Trans. on Commun.*, vol. 70, no. 1, pp. 707–723, Jan. 2022.

[9] T. Jiang and W. Yu, "Interference nulling using reconfigurable intelligent surface," *IEEE J. on Sel. Areas in Commun.*, vol. 40, no. 5, pp. 1392–1406, May 2022.

[10] E. Björnson *et al.*, "Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 135–158, Mar. 2022.

[11] M. Almekhlafi *et al.*, "Joint resource allocation and phase shift optimization for RIS-aided eMBB/URLLC traffic multiplexing," *IEEE Trans. on Commun.*, vol. 70, no. 2, pp. 1304–1319, Feb. 2022.

[12] V. D. P. Souto *et al.*, "IRS-aided physical layer network slicing for URLLC and eMBB," *IEEE Access*, vol. 9, pp. 163 086–163 098, Dec. 2021.

[13] X. Jiang *et al.*, "Packet detection by a single OFDM symbol in URLLC for critical industrial control: A realistic study," *IEEE J. on Sel. Areas in Commun.*, vol. 37, no. 4, pp. 933–946, Apr. 2019.

[14] A. Albanese *et al.*, "MARISA: A self-configuring metasurfaces absorption and reflection solution towards 6G," in *IEEE INFOCOM 2022 - IEEE Conf. on Comput. Commun.*, 2-5 May 2022, pp. 250–259.