Wilson de Souza Junior

Reconfigurable Intelligent Surfaces Driven B5G Networks: An Analytical Performance Perspective

Londrina

July 30, 2023

Wilson de Souza Junior

Reconfigurable Intelligent Surfaces Driven B5G Networks: An Analytical Performance Perspective

A dissertation submitted to the Electrical Engineering Graduate Program at the State University of Londrina in fulfillment of the requirements for the degree of Master of Science in Electrical Engineering.

State University of Londrina – UEL

Department of Electrical Engineering

Associated Master Program in Electrical Engineering UEL

Supervisor: Prof. Dr. Taufik Abrão

Londrina

July 30, 2023

Junior, Wilson de Souza

Reconfigurable Intelligent Surfaces Driven B5G Networks: An Analytical Performance Perspective. Londrina, 2023. 103 p.

Supervisor: Prof. Dr. Taufik Abrão

Dissertation (Master of Science) – Department of Electrical Engineering – State University of Londrina

Wilson de Souza Junior

Reconfigurable Intelligent Surfaces Driven B5G Networks: An Analytical Performance Perspective

A dissertation submitted to the Electrical Engineering Graduate Program at the State University of Londrina in fulfillment of the requirements for the degree of Master of Science in Electrical Engineering.

Examination Board

Prof. Dr. Abolfazl Amiri
Nokia/Aalborg University - Denmark
Member

Prof. Dr. Jaime Laelson Jacob
State University of Londrina
Member

Prof. Dr. Taufik Abrão
State University of Londrina
Supervisor

Londrina
July 30, 2023

Resumo

Comunicação Massiva do Tipo Máquina (mMTC), Banda Larga Móvel Aprimorada (eMBB) e Comunicação de Baixa Latência e Ultra Confiabilidade (URLLC) desempenharão uma função muito importante nas redes de comunicação atuais Além da Quinta Geração (B5G) do que nas redes de comunicação de Quinta Geração (5G), cujo objetivo é operar em cenários futuros como casas inteligentes, fábricas inteligentes, transporte e dirigibilidade inteligentes, bem como dar suporte à comunicação ubíqua em todos outros ambientes práticos previstos para estarem inclusos, na então denominada, Indústria 4.0. Evidentemente, a demanda por taxas de dados nestes cenários serão fortemente aumentadas devido aos inúmeros serviços a serem ofertados. Ademais, o número de dispositivos conectados tem crescido exponencialmente, uma vez que a *Internet* das Coisas (IoT) será essencial para se atingir o futuro projetado. Para atender todos estes cenários com serviços de alta qualidade, novas tecnologias serão necessárias, uma vez que as tecnologias existentes não serão eficientes em atender simultaneamente todas as demandas dessas redes de comunicação. Além disso, uma vez que os recursos de rádios são escassos, é desafiador desenvolver novos métodos para fornecer qualidade de serviço aceitável aos usuários enquanto simultaneamente deve-se atender aos requisitos necessários para uma comunicação sustentável. Considerando os desafios mencionados anteriormente, nesta Qualificação de Mestrado nós discutimos o impacto de se aplicar uma nova tecnologia promissora em sistemas de comunicação denominada Superfície Inteligente Reconfigurável (RIS) através da sua forma convencional, bem como sua derivação avançada denominada RIS de Transmissão e Reflexão Simultânea (STAR-RIS), tendo em vista obter uma compreensão abrangente de como a RIS pode impactar métricas de desempenho importantes, nas atuais redes de comunicação, como confiabilidade, taxa de dados e eficiência energética. Associada à tecnologia habilitadora RIS, também integramos ao esquema de transmissão as estruturas de Múltiplas-Entradas Múltiplas-Saidas massiva (mMIMO) e o esquema de Múltiplo Acesso não-Ortogonal (NOMA) com intuito de caracterizar o potencial operacional destas tecnologias em conjunto com a RIS. Além disso, neste trabalho, nós focamos na comunicação sustentável através dos estudos em comunicação cooperativa com ajuda de Transmissão Simultânea de Energia e Informação (SWIPT), bem como desenvolvemos uma análise sobre seleção do número de antenas na estacão rádio-base e seleção do número de elementos da RIS visando incrementar a Eficiência Energética (EE) e Eficiência Espectral (SE) enquanto mantém altos índices de confiabilidade de comunicação. Por fim, neste trabalho, predominantemente, utilizamos uma extensa abordagem analítica para a descrição de cada problema e respectiva solução proposta buscando entender sistematicamente o impacto de cada parâmetro no desempenho do sistema como um todo, incluindo probabilidade de *outage*, EE, SE, confiabilidade, bem como reduzir a complexidade dos sistemas de transmissão, propondo algoritmos

quase-ótimos eficientes, tendo em vista adequar suas operações aos problemas práticos no mundo real.

Palavras-chave: Comunicação massiva do tipo máquina (mMTC); Comunicação de baixa latência e ultra-confiabilidade (URLLC); *Internet* das coisas (IoT); Superfície inteligente reconfigurável (RIS); Múltiplas-entradas múltiplas-saidas massiva (mMIMO); múltiplo acesso não-ortogonal (NOMA); transmissão simultânea de energia e informação (SWIPT).

Abstract

Massive Machine-Type Communications (mMTC), Enhanced Mobile Broadband (eMBB), and Ultra-Reliable Low Latency Communications (URLLC) use modes will play a more crucial key role in the Beyond Fifth Generation (B5G) communication systems than in the 5G communication systems, whose aim is to operate ubiquitously in the future scenarios such as smart homes, smart factories, transport and intelligent drivability, as well as other practical setups predictable to be included in the Industry 4.0. The demand for rates in these scenarios will powerfully increase due to the countless services offered. In addition, the number of connected devices has exponentially increased once the Internet of Things (IoT) became paramount for materializing the future. New technologies are required to provide improved services in these scenarios since the existing technologies will only be efficient in fulfilling some expected demands in these communications networks. Furthermore, since the radio resources are limited, developing new methods and technologies to provide Quality of Service (QoS) to the users while simultaneously regarding sustainable communications is challenging. Considering the aforementioned challenges, in this Master' Dissertation, we study the impact of deploying new promising technology in communication systems, named Reconfigurable Intelligent Surfaces (RIS), in their most diverse kinds, including the conventional RIS, the enhanced version, known as simultaneous transmitting and reflecting RIS (STAR-RIS), aiming to obtain a comprehensive understanding on how the RIS-based systems can impact key performance metrics, such as reliability, Spectral Efficiency (SE), and Energy Efficiency (EE). Associated with such enabler technology RIS, we also integrate the massive Multiple Input Multiple Outputs (mMIMO) and Non-Orthogonal Multiple Access (NOMA) technologies into the transmission scheme, aiming to understand and analyze their potential when operated jointly. Furthermore, in this work, we praising for sustainable communications by carrying out a study on cooperative communications with the aid of Simultaneous Wireless Information and Power Transfer (SWIPT) technology, as well as develop an analysis of the antenna number selection and RIS elements selection methodologies aiming to increase the EE, intending to improve the energy efficiency while holding the communication reliability. Finally, in this Master' Dissertation, we also predominantly utilized analytical methodologies to describe and assess each problem and its respective proposed solution to understand systematically the impact of each system parameter on the system performance, including the outage probability, EE, and SE, as well as to reduce the complexity of the proposed transmitting schemes, proposing quasi-optimal algorithm, aiming to adequate such schemes for real-world operations.

Keywords: Massive machine-type communications (mMTC), Ultra-reliable low latency communications (URLLC), Internet of things (IoT), Reconfigurable intelligent surfaces

(RIS), massive multiple inputs multiple outputs (mMIMO), non-orthogonal multiple access (NOMA), simultaneous wireless information and power transfer (SWIPT).

List of Figures

List of acronyms

1G	First Generation
4G	Fourth Generation
5G	Fifth Generation
AO	Alternating Optimization
AP	Access Point
B5G	Beyond the Fifth Generation
BS	Base Station
CDMA	Code-Division Multiple Access
CLT	Central Limit Theorem
CSI	Channel State Information
CD-NOMA	Code Domain non-Orthogonal Multiple Access
DL	Downlink
EE	Energy Efficiency
EH	Energy Harvested
eMBB	Enhanced Mobile Broadband
ER	Ergodic Rate
ES	Energy Splitting
FDMA	Frequency-Division Multiple Access
IoT	Internet of Things
IRS	Intelligent Reflecting Surface
KPI	Key Performance Indicator
LIS	Large Intelligent Surface
LoS	Line of Sight
MCs	Monte Carlo Simulation
mMIMO	Massive Multiple-Input Multiple-Output
mmWave	Milimeter Wave
MM	Moment Matching
mMTC	Massive Machine Type Communication
MS	Mode Switching
mULC	Massive Ultra-Reliable Low-Latency Communication
NLoS	Non Line of Sight
NOMA	Non-Orthogonal Multiple Access
OFDMA	Orthogonal Frequency-Division Multiple Access
OP	Outage Probability
PD-NOMA	Power Domain non-Orthogonal Multiple Access
QoS	Quality-of-Service
RF	Radio Frequency

RIS	Reconfigurable Intelligent Surface
SE	Spectral Efficiency
SI	Self Interference
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
STAR	Simultanous Transmitting and Reflecting
SWIPT	Simultaneous Wireless Information and Power Transfer
TDMA	Time-Division Multiple Access
TS	Time Switching
uMBB	Ubiquitous Mobile Broadband
UEL	State University of Londrina
UL	Uplink
ULBC	Ultra-Reliable Low-Latency Broadband Communication
URLLC	Ultra Reliable Low Latency Communications
VR	Visibility Region
ZF	Zero-forcing

List of Symbols

Chapter 2

N	Number of reflective elements in the RIS
D_1	Device 1
D_2	Device 2
B	Bandwidth
σ^2	AWGN power
R_1	Target rate for device 1
R_2	Target rate for device 2
ϕ_n	Phase-shift of n-th element in the RIS
m	Nakagami-m shape parameter
Ω	Nakagami-m spread parameter
α_k	Power coefficient for the k-th device
β	Large-scale fading coefficient
ρ	EH coefficient
ω	SI power
x_k	Transmitted information symbol for the k-th device
P_t	Transmitted power
P_H	Harvested power
$\bar{P}_H$	Mean harvested power
P_{th}	Threshold harvested power
P_{in}	Radio-frequency input power in the EH circuit
a	EH circuit constant
b	EH circuit constant
γ_{th}	SNR threshold
P_{out}	Outage probability
π	Pi constant

Chapter 3

M	Number of antennas in the base station
N	Number of reflective elements in the RIS
K	Number of single-antenna users
$\mathcal{M}$	Set of antenna elements
$\mathcal{N}$	Set of reflective elements
β_k	Large-scale fading for BS-k-th user link
ν_k	Large-scale fading for RIS-k-th user link
α	Large-scale fading for RIS-BS link
K_1	Rician factor
θ_n	Phase-shift of n-th element in the RIS
σ^2	AWGN power
p_k	Radio-frequency power for the k-th user
R_k	Data rate for the k-th user
$R_{\min}$	Minimum rate
P_{CP}	Power consumption at BS
P_{TX}	Maximum transmitted power
ϱ	Power amplification inefficiency
P_{BS}	Radio frequency chains circuit power consumption per transmit antenna
P_{RIS}	Power consumption of each reflecting element in the RIS
$\mathcal{L}$	Lagrangian
μ_k	Lagrangian multiplier
λ	Lagrangian multiplier
δ	Lagrangian multiplier
s_k	Transmitted information symbol for the k-th user
n_k	AWGN in the k-th user
$\mathcal{T}$	Number of Monte-Carlo simulations
π	Pi constant

Chapter 4

N	Number of total reflective elements
R_C	Conventional-RIS nomenclature
R_S	STAR-RIS nomenclature
N_C	Number of reflective elements in R_C
N_S	Number of reflective elements in R_S
$\mathcal{N}_C$	Set of reflective elements in R_C
U_O	Outdoor user
U_I	Indoor user
ξ_k	Amplitude coefficient in the n-th element in R_S
K_1	Rician factor
φ_n	Phase-shift of n-th element in the R_S
ϕ_n	Phase-shift of n-th element in the R_C
σ^2	AWGN power
P	Maximum transmitted power
α_O	Power coefficient for the outdoor user
α_I	Power coefficient for the indoor user
ρ	SNR
β	Large-scale fading
α	Path-loss exponent
d	Distance in meters
$\bar{\gamma}$	SNR threshold
m	Nakagami-m shape parameter
Ω	Nakagami-m spread parameter
x_i	Transmitted symbol for i-th user
z_i	AWGN in the i-th user
π	Pi constant

Chapter 5

M	Number of antennas in the base station
N	Number of reflective elements in the RIS
K	Number of single-antenna users
$\mathcal{M}$	Set of antenna elements
$\mathcal{N}$	Set of reflective elements
β_k	Large-scale fading for BS-k-th user link
ν_k	Large-scale fading for RIS-k-th user link
α	Large-scale fading for RIS-BS link
K_1	Rician factor
θ_n	Phase-shift of n-th element in the RIS
σ^2	AWGN power
p_k	Radio-frequency power for the k-th user
R_k	Data rate for the k-th user
$R_{\min}$	Minimum rate
P_{CP}	Power consumption at BS
P_{TX}	Maximum transmitted power
ϱ	Power amplify inefficiency
P_{BS}	Radio frequency chains circuit power consumption per transmit antenna
P_{RIS}	Power consumption of each reflecting element in the RIS
$\mathcal{L}$	Lagrangian
μ_k	Lagrangian multiplier
λ	Lagrangian multiplier
δ	Lagrangian multiplier
s_k	Transmitted information symbol for the k-th user
n_k	AWGN in the k-th user
$\mathcal{T}$	Number of Monte-Carlo simulations
π	Pi constant

Contents

1 Introduction

1.1 Motivation

In nowadays, there is a strong motivation to study, understand, and improve future communication systems, in view of addressing the growing challenges posed by the exponential increase in data demand and limited resources, particularly spectrum and energy, as well as the need for higher spectral efficiency, reduced latency, and increased network capacity. These systems will play a crucial role in supporting a wide range of advanced applications and services, such as 5G networks and beyond. It is evident that the past technologies will not suffice to meet these demands of future communication since future scenarios are projected to entail these several ambitious requirements; therefore, it becomes crucial to study and enhance new technologies for their future implementation in B5G communication networks.

Through the study of new technologies and innovative solutions, we can design systems that deliver superior performance, enhanced reliability, and an improved user experience. These emerging technologies promise to revolutionize how we connect, communicate, and interact with the world around us. Moreover, it is crucial to prioritize technologies with low implementation costs and high sustainability rates, alongside their performance impact exploring and addressing sustainability concerns, minimizing energy consumption, and reducing the environmental impact of communication networks. This sustainability-driven approach is crucial to ensure that future technologies are economically viable, socially responsible, and environmentally sustainable.

Therefore, dedicating efforts to study the future communication system is essential to drive innovation, anticipate the growing demands of society, and lay the foundations for an advanced and sustainable digital future. Hence, this Master's Dissertation aims to shed light on the benefits and drawbacks of selected technologies that have garnered significant attention within the academic community and industries recently, especially enabling technologies such as RIS, IoT, mMIMO, NOMA, and SWIPT. Additionally, we propose solutions to justify, or not, their respective implementations in future scenarios.

1.2 Beyound 5G services

The jointly rapid advent of the IoT, Mobile Broadband (MBB), and other fields have caused an exponential increase in the number of smartphones, smart devices, tablets, wearable electronics, virtual reality glasses, etc., associated with the network, imposing

a huge demand on mobile communications. Furthermore, the demands for high-quality wireless services imply notorious challenges to the existing cellular networks. It is expected that the average data consumption for every mobile user per month will increase from about 5GB in 2020 to around 250 GB in 2030 [Alwis et al. 2021], this is mainly due to the popularity of mobile video services such as YouTube, Netflix, Tik-Tok, Twitch, and other streaming services [Guo et al. 2021].

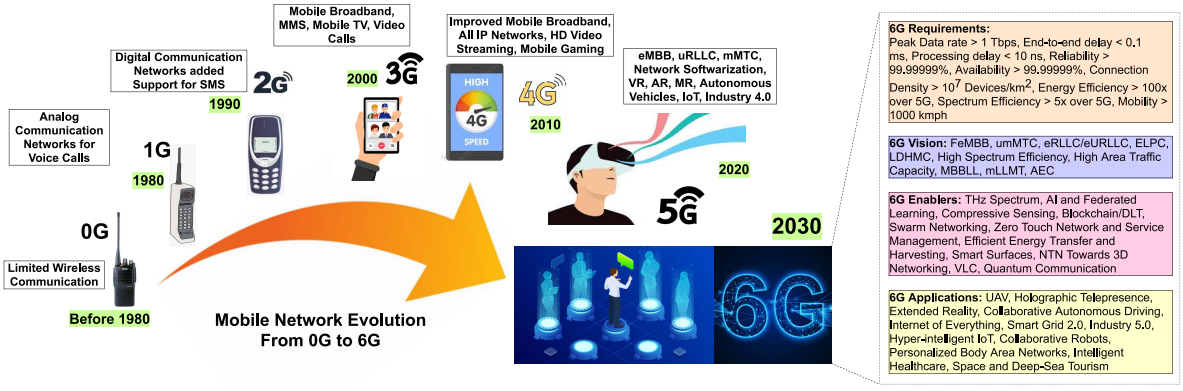


Figure 1 – The evolution of the communication system [Alwis et al. 2021].

Furthermore, nowadays, it is far from reality to provide ubiquitous coverage (encompassing the population in remote, sparse, rural, water surfaces and underwater areas) across the whole planet with sufficient capacity, acceptable QoS, and affordable cost. In this sense, the B5G systems, are expected to make use of the synergy of all kinds of applications such as terrestrial networks, satellite constellations, aerial platforms, etc., to provide ubiquitous connectivity [Guo et al. 2021, Huang et al. 2019], as illustrated in Fig. 1. Furthermore, the main services presented in the 5G, like [Liu et al. 2021, Carvalho et al. 2017]

1. **Enhanced mobile broadband (eMBB):** in this service, the broadband is utilized to provide high data rates and seamless access for human-centric applications, such as smart devices;

2. **Ultra-reliable low latency communications (URLLC):** in this service, the ultra-high reliability and low latency are essentials, aiming to support applications such as autonomous vehicles, smart grid, and industry 4.0;

3. **Massive machine-type communications (mMTC):** where the focus is to support dense connectivity with a very large number of connected low-cost and low-power devices,

should be present in the B5G communication systems, however, remarkably improved. In addition, there are some works in the literature that envisioned new applications based on these cited services to guide the deployment of B5G systems, such as sixth-generation (6G), for instance. In [Jiang et al. 2021], is conceived the following

1. **Ubiquitous mobile broadband (uMBB):** where the eMBB and mMTC are combined to provide global ubiquitous connectivity boosting, even more, the capacity of networks;

2. **Ultra-reliable low-latency broadband communication (ULBC):** where its focus is to support the applications of URLLC but also with extremely high throughput;

3. **Massive ultra-reliable low-latency communication (mULC):** that combines the characteristics of both mMTC and URLLC, facilitating the deployment of massive sensors in vertical industries.

1.3 Massive MIMO

Massive multiple-input multiple-output (mMIMO) is a technology that leverages dozens, hundred of antennas embedded in a physically large array at the base station (BS) [Marzetta et al. 2016], in order to provide better-operating conditions for the users simultaneously and emerges as the main candidate to achieve the aforementioned services. Through the elevated number of antennas, the mMIMO can create highly directive beams, providing better conditions of service for many users simultaneously, utilizing the same resource of frequency and time. Besides, these directive beams can concentrate the energy toward an intended user, reducing the transmission power, minimizing the interference between users e improving the spectral efficiency (SE). This technique is denominated space division multiple access (SDMA) and it is a main supporter of the 5G communication systems. The mMIMO scheme also provides benefits such as channel hardening and favorable propagation, which are interesting phenomena [Marzetta et al. 2016, Björnson, Hoydis e Sanguinetti 2017].

1. **Channel hardening:** The channel hardening denotes the phenom where the fading channel behaves as if it was a non-fading channel, *i.e.*, by increasing the number of antennas (M) embedded in the array at the BS, the channel of User 1, $\mathbf{h}_1 \in \mathbb{C}^M$ with the BS, behaves almost deterministically, due to the spatial diversity;

$$\frac{\mathbf{h}_1^H \mathbf{h}_1}{\mathbb{E}\left[\mathbf{h}_1^H \mathbf{h}_1\right]} \to 1, \qquad M \to \infty \tag{1.1}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator.

2. **Favorable propagation:** When the number of antennas at the BS increases rigorously, the channel directions of the channels of User 1 $\mathbf{h}_1 \in \mathbb{C}^M$ and User 2 $\mathbf{h}_2 \in \mathbb{C}^M$, become nearly orthogonal, and it is denoted as the favorable propagation phenom.

$$\frac{\mathbf{h}_1^H \mathbf{h}_2}{\sqrt{\mathbb{E}\left[\mathbf{h}_1^H \mathbf{h}_1\right] \mathbb{E}\left[\mathbf{h}_2^H \mathbf{h}_2\right]}} \to 0, \qquad M \to \infty \tag{1.2}$$

Although the mMIMO systems bring great gains, there are some challenges with respect to its implementation in practice, as: *a)* deployment of large aperture arrays; *b)* limitation in channel prediction; *c)*; low-cost arrays and the *d)* pilot contamination [Björnson, Hoydis e Sanguinetti 2017]. Independently, it is expected that this technology move toward the B5G communication systems.

Furthermore, by increasing the number of antennas at the BS, we have originated the so-called extra-large scale massive MIMO (XL-MIMO) [Carvalho et al. 2020, Marinello et al. 2020]. One of the differences between the mMIMO and XL-MIMO is that the spatially non-stationary wireless channel which is dominated by spherical wavefront propagation can appear. The commonly adopted boundary region which characterizes the spherical wavefront is named Rayleigh distance and is proportional to the square of the array aperture and the inverse of wavelength [Cui et al. 2022]

$$r_{rd} = \frac{2D^2}{\lambda}, \tag{1.3}$$

where D is the array aperture and λ denotes the wavelength. In 5G mMIMO communications, the Rayleigh distance is only several meters since the arrays are not very large [Wu e Dai 2023], however, in the XL-MIMO this distance can be potentially increased since typically in the XL-MIMO the aperture of the array is very large. In addition, the large arrays imply the so-called visibility region (VR) [Carvalho et al. 2020], which means that users when located in a specific area see only a geographical area of the array, where this area can be changeable according to the user position.

Therefore, these aspects must be addressed in order to consistently understand the system. Accordingly, a new field of study denominated as *near-field propagation* [Björnson e Sanguinetti 2020] is conceived.

1.4 Reconfigurable Intelligent Surfaces

In order to achieve all aforementioned requisites, new technologies are necessary aiming to support the B5G services. Among all emerging technologies, RIS-assisted communications have stood out. The RIS, also known in the literature as Large Intelligent Surface (LIS) and Intelligent Reflecting Surface (IRS), is a planar surface consisting

of a massive number of reflecting elements, with the objective of reconfiguring and redirecting the impinging signal, by changing its electromagnet properties (phase and/or magnitude)[Wu e Zhang 2020], as illustrated in Fig. 2, so that, the reflected signal can be focused toward a specific direction. Due to its capacity to control in a software-defined manner the wireless environment, RIS is promising due to its potential to mitigate the intrinsic effects present in the wireless channels. The specifics advantages of the RIS arrays can be summarized as follows [Gong et al. 2020]:

1. **Easiness in implementation:** The RISs are made of electromagnetic materials, namely metamaterials, which are embedded in the metasurfaces and have lightweight, small geometry leading to low cost. Furthermore, it can be in any shape, providing high versatility for its deployment and incorporation into built space. The RISs arrays can be employed in various structures, including stationary or moving objects and even on people. For example, RIS can be installed on building facades, ceilings, aerial platforms, billboards, unmanned aerial vehicles (UAVs), street lamps, vehicle windows, and, even on pedestrians' clothes;

2. **Spectral and Energy efficiency enhancement:** RISs have the potential to influence the propagation environment in a controlled manner, thus being able to compensate for the power loss over long distances. Therefore, there is the possibility of an LoS virtual connection between the base station and the users through the RIS, formed via the reflection of the impinging signals, reducing the power consumption in the communications, if well projected. Thus, an increase in the throughput becomes significant with the possibility of a direct link between the base station and the users, even if virtual. In this way, a wireless environment built by software can be meticulously designed in order to provide a substantial improvement in the QoS of the users;

3. **Sustainability:** A great paradigm for the next generations of wireless networks is the innovation and implementation of sustainable technologies, with high values of energy efficiency, called green communications. In this sense, the RISs, unlike conventional signal repetition methods (relays), such as amplify-and-forward (AF) and decode-and-forward (DF) [Björnson, Ozdogan e Larsson 2020], are capable of controlling the phase shift of each scattering element present in it, namely passive beamforming, focusing the incident signal in a specific direction and nulled at other directions, as opposed to the AF and DF relay, which solely apply a power amplifier. Due to this, it can be stated that RISs are more energetically efficient than both signal repetition methods. In addition, although additional radio frequency chains are not necessary, the RIS has a high potential to provide an elevated degree of spatial multiplexing with reduced power consumption, when compared with other promising physical layer technology such as mMIMO;

4. **Compatibility:** Due to the fact that RISs are only responsible for reflecting electromagnetic waves, their compatibility with full-duplex technology is a very interesting aspect and one that stands out over half-duplex technology, where orthogonal resources are needed to receive and propagate. In addition, RISs are compatible with the existing standards and hardware of the existing wireless networks.

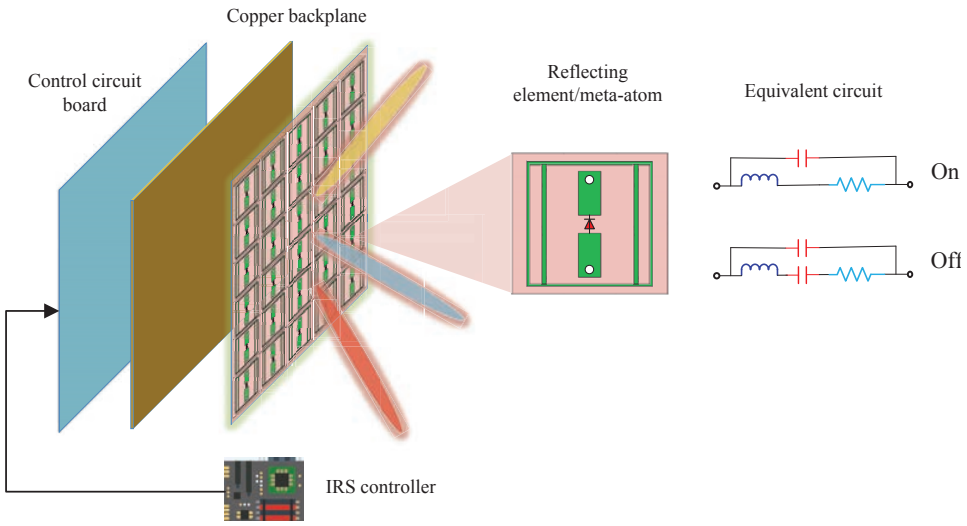


Figure 2 – Schematic illustration of a conventional RIS prototype [Wu e Zhang 2020].

Due to the aforementioned characteristics, RISs are recognized as a potential solution for solving a wide range of challenges in commercial and civilian applications. On this matter, like any emerging technology, new applications and a range of problems arise as a result, from the perspective of communication and signal processing, open problems such as **phase-shift optimization** of scattering elements, **channel acquisition** of additional links imposed by deploying the RIS are extremely necessary to be addressed and efficiently solved [Pan et al. 2021].

1. **Phase-shift optimization:** The phase-shift optimization is one of the main challenges concerning RIS. It should be noticed that this problem can be addressed in diverse manners, i.e., it is directly related to the optimization objective. Furthermore, it is noteworthy that the difficulty of this problem, in practice, is straightforwardly connected to the system parameters such as a) the number of users to be attached, b) the number of RIS elements, c) phase-shift levels of quantization of each element, etc. In the literature, there are many reports of conventional strategies being applied to this kind of problem such as convex and non-convex optimization, heuristic optimization, and even Machine Learning (ML) approaches through supervised/reinforcement learning;

2. **Channel Acquisition:** To obtain interesting performance gains, in general, accurate knowledge of the Channel State Information (CSI) between the RIS, Access Points

(APs), and users' devices is necessary. Besides, with high-order antenna arrays both in BS and planar RIS, the procedure to obtain a good estimate of CSI can be drastically hampered. Although conventional methods from the mMIMO scenario can be leveraged, the complexity in this new setup with the deployment of the RIS is highly elevated due to the additional links imposed. Furthermore, in scenarios where more than one single RIS is deployed, conventional procedures can be unaffordable. Alternatives to this problem can be considering the statistical channel information, which can reduce the complexity since estimating the CSI in all coherence time will not be necessary.

Although RIS brings diverse benefits, there are some variants of architectures that seek to address some problems in order to provide better performance. Among them, we can remark the **active RIS** and the **STAR-RIS**.

1. **Active RIS:** As an alternative for providing better conditions for the double-fading attenuation link, the active RIS emerges, where each reflecting element is supported by a set of active-load impedance, *i.e.* the active RIS is comparable to an active reflector that directly reflects the impinging signal by applying a power amplification. Since the active RIS has the ability to amplify the incident signal, fewer reflecting elements are required to achieve the same Signal-to-Noise Ratio (SNR) at the received when compared with the conventional passive RIS [Long et al. 2021]. Moreover, an interesting advantage of the active elements is their capability to estimate the channel between the transmitter-RIS in downlink and uplink at the RIS side [Taha, Alrabeiah e Alkhateeb 2021]. Therefore, it leaves room to develop the semi-passive/active RIS topology, where, both active and passive elements are present, leading to lower consumption with regard to the completely active RIS and less channel estimation overhead since with the channel partial knowledge, acquired by the active elements, the channel with respect to the passive elements also can be estimated in a prompt way through techniques as Machine Learning (ML), for instance;

2. **STAR-RIS:** Since the conventional RIS has the objective to reflect the incident signal, both the source and the destination have to be on the same side of the RIS. However, this physical constraint limits the flexibility of the conventional RIS. Aiming to solve this problem, STAR-RIS emerges in order to provide complete access, *i.e.*, with this topology the incident signal can be reflected and refracted enhancing the coverage from 180° to 360° [Liu et al. 2021, Xu et al. 2021], as depicted in Fig. 3. There are three practical protocols for operating STAR-RIS in wireless networks [Liu et al. 2021], namely energy splitting (ES), mode switching (MS), and time switching (TS), and each one has its particularities. The ES protocol provides high

flexibility since the reflecting elements can work simultaneously as reflecting and refracting; the MS protocol is easy to implement since the mode of operation of each element can be selected; in the TS protocol, the RIS operates in reflection and refraction mode in different times, providing independent transmission and reflection design.

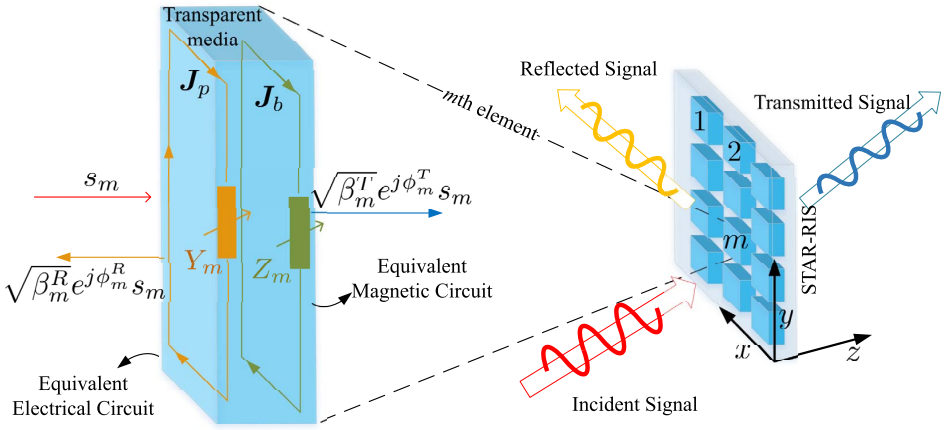


Figure 3 – Schematic illustration of an STAR-RIS prototype [Xu et al. 2021].

1.5 Non-Orthogonal Multiple Access

Over the decade previous generation's cellular networks from the First Generation (1G) to Fourth Generation (4G) have endorsed various distinct multiple access schemes such as Time-Division Multiple Access (TDMA), Frequency-Division Multiple Access (FDMA), Code-Division Multiple Access (CDMA) and Orthogonal Frequency-Division Multiple Access (OFDMA), *i.e.*, Orthogonal Multiple Access schemes (OMA). Since all the aforementioned multiple access schemes have as the principle the orthogonal resource utilization, they all suffer from orthogonality limitation, *i.e.*, one resource cannot be allocated to more than one user [Vaezi et al. 2019]. As a potential technology to achieve all the expected demand for the 5G and B5G systems, as massive connectivity, high throughput and provide services for machine-type communication, IoT applications, and human-centric communication, NOMA has grown up and has been severally studied in the academical community.

Several NOMA schemes have been proposed aspiring to explore the most sundry domains, the prevailing NOMA technique is split mainly into two categories, such as code domain NOMA (CD-NOMA) [Le et al. 2019] and power domain NOMA (PD-NOMA), however, different strands are also studied, as for instance, the waveform domain NOMA [Şahin e Arslan 2020].

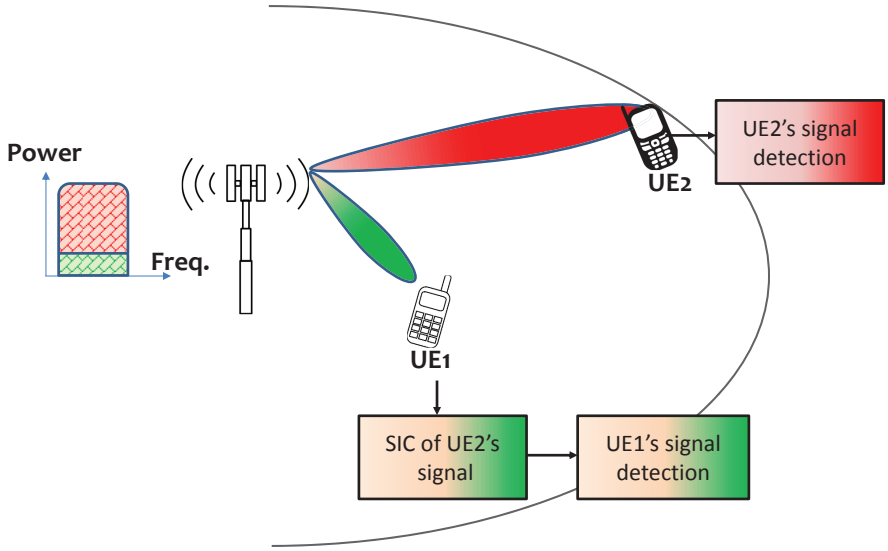


Figure 4 – Schematic illustration of a downlink NOMA via power domain multiplexing for two users [Vaezi et al. 2019].

The NOMA concept is fundamentally different from conventional OMA, *i.e.* the key principle of NOMA transmission is to encourage resource-sharing among users, where, multiple users are allowed to be served at the same time and frequency, for instance. The success of PD-NOMA, for example, is based on its capability to efficiently exploit the users' dynamic channel conditions. In the PD-NOMA, as depicted in Fig. 4, the user with better channel conditions, UE1 (denoted by green color), shares the same spectrum with the user with poor channel conditions, UE2 (denoted as red color). In order to decode the intending signal, UE1 adopts the successive interference cancellation (SIC) method to cancel the interference of UE2, whose signal is transmitting over the same spectrum, while UE2 decodes its signal considering the signal of UE1 as noise once it is allocated higher power for UE2' signal. Therefore, potentializing the system's performance and reliability, if suitably implemented [Ding et al. 2022].

In order to improve the fairness of the system, some extensions of NOMA also have attracted some attention in both the industrial and academic communities. The cooperative NOMA can be viewed as a potential solution for improving the communication reliability for the URLLC scenarios, as well as improving the system throughput for IoT scenarios. Cooperative re-transmission can be an alternative technique that seeks to combat the intrinsic effects of the channel, such as channel fading, path loss, and shadowing. The canonical model of cooperative re-transmission basically consists of three transmit/receive nodes, namely, a source node, a relay, and a destination node. The objective of cooperative re-transmission is to improve the communication between the source and receiver, through the help of the relay. However, the extra consumption of resource blocks is an unavoidable factor, for example, in a TDMA system, which requires

two temporal slots to complete the transmission, one slot is responsible for the transmission of the source to the destination and another is reserved for the re-transmission of the signal from the *relay* node to the destination node [Zeng et al. 2020].

In the literature, there are two main re-transmission protocols: AF and DF. The idea behind the AF protocol is simply to propagate a scaled version of the observed signal at the relay to the destination, while DF first seeks to decode the observed signal and then re-transmit. Recently, the combination of NOMA and cooperative relaying has gained a lot of attention due to its promising advantages. One of the main procedures regarding this combination is to use the user with better channel conditions as a relay in order to enhance conditions for the user with poor channel conditions. One can notice due to the SIC, the user-relay will necessarily detect the signal of the weak user, so this fact is crucial to the well-deployment of the system.

Aiming at improving the system, the application of simultaneous wireless information and power transfer (SWIPT) [Özyurt et al. 2022] in cooperative NOMA systems has recently emerged, which uses a technology implemented in the receiver capable of subdividing the power of the received signal aiming at decoding and the other portion of the power destined for re-transmission, battery recharge, among other applications. With this technology, the idea of cooperative NOMA with SWIPT has been gaining a lot of attention lately [Liu et al. 2016, Ren et al. 2022, Zhang et al. 2022, Shaheen e Soleymani 2022].

1.6 Simultaneous Wireless Information Power Transfer

Energy harvesting (EH) is the process of capturing ambient wasted or insignificant heat, sound, wind, and radio wave (radio frequency, RF) emissions and converting them into electrical energy for use in powering devices[Özyurt et al. 2022]. Among different energy scavenging approaches, SWIPT has gained popularity because RF signals have the capability to carry both information and energy. As making ubiquitous wireless communications possible in a self-sustainable fashion, SWIPT can be an indispensable solution for B5G providing the necessary energy for wireless charging of energy-constrained devices and for transmitting and receiving information.

The SWIPT process can be done in two well-defined ways, named Time Switching (TS) and Power Splitting (PS) [Özyurt et al. 2022]. In the TS structure, the receiving antenna is periodically switched between the information decoding (ID) and energy harvest (EH) circuits in an alternating fashion. On the other hand, the PS receiver splits the received signal into two parts with different power levels under a certain PS ratio. Subsequently, two streams are separately sent to the ID and EH units to allow simultaneous ID and EH operations. By modifying the PS ratio at the receiver, the information transmission rate and the level of harvested energy can be balanced and

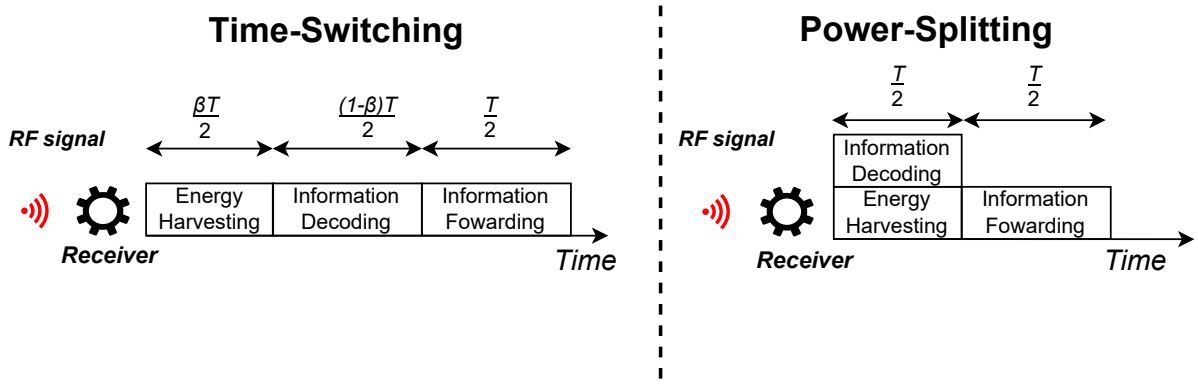


Figure 5 – Working principle of TS and PS SWIPT techniquies [Zeng et al. 2020].

optimized according to system needs

1.7 Key Concepts

In order to gain insight into and contribute to future communication systems design, it is essential to delve into key concepts that help us to understand and enhance communication systems. Among these concepts, Ergodic Rate and Outage Probability are of paramount importance, and they are extensively analyzed in this Master's Dissertation. These concepts play a crucial role in evaluating system performance, understanding capacity limits, guiding resource allocation, and facilitating comparative analysis. This knowledge is of significant importance for developing efficient communication solutions across diverse application domains. Additionally, the communication channels play a pivotal role in ensuring the consistency and validity of the studied techniques, given the numerous scenarios and conditions they encompass. Therefore, the following sections provide a description of these concepts, highlighting their significance and application in the field, as well as a short description of the adopted channel models.

1.7.1 Ergodic Rate

The Ergodic Rate (ER), also known as the Shannon (Ergodic) Capacity, defines the maximum data rate that can be sent over the channel with an asymptotically small error probability[Goldsmith 2005]. With this metric, we can provide valuable insights into communication systems' performance limits, optimization, and design. Mathematically, it can be defined as

$$\bar{R} = \mathbb{E}\left[\log_2(1+\gamma)\right] = \int_0^\infty \log_2(1+\gamma)p(\gamma)d\gamma, \tag{1.4}$$

where γ is the SNR at receiver and $p(\gamma)$ is the Probability Density Function (PDF) of γ.

1.7.2 Outage Probability

Studying the future behavior of system reliability is of paramount importance. In this regard, the concept of outage probability plays a crucial role as it quantifies the probability of a specific metric falling below a desired threshold; in other words, it represents the failure of a system to meet certain performance requirements. In this sense, with the outage probability, we can assess the system's robustness and identify potential vulnerabilities. Therefore, concerning the SNR (γ), we defined this metric as follows

$$P_{\text{out}} = \Pr\left(\gamma < \gamma_{th}\right) = \int_0^{\gamma_{th}} p(\gamma) d\gamma, \tag{1.5}$$

where γ_{th} is the SNR threshold.

1.7.3 Channel Model

In communication systems, basically, we can define the communication channels as: **line of sight** or **non-line of sight**. The line of sight (LoS) channel refers to a specific type of wireless communication channel characterized by direct, unobstructed propagation of electromagnetic waves between a transmitter and a receiver. In LoS channels, there is a clear and uninterrupted path between the transmitting and receiving antennas, with no significant obstacles or interference in the signal path. Non-line of sight (NLoS) channels represent a distinct category of wireless communication channels that contrast sharply with LoS channels. In NLoS scenarios, the propagation of electromagnetic waves between the transmitter and the receiver is characterized by multiple reflections, diffractions, and scattering events due to the presence of obstacles or environmental conditions that obstruct the direct path between the two endpoints. To effectively characterize both types for simulation purposes, specific approaches are employed. The Rician fading channel is the standard choice for characterizing both LoS band NLoS paths. A channel represented by the Rician fading model can be expressed as follows

$$\mathbf{H} = \sqrt{\frac{\kappa}{\kappa+1}}\bar{\mathbf{H}} + \sqrt{\frac{1}{\kappa+1}}\tilde{\mathbf{H}}, \tag{1.6}$$

where $\bar{\mathbf{H}}$ is the LoS component, $\tilde{\mathbf{H}}$ is the NLoS component, and the parameter κ in the Rician fading channel model is a critical factor that quantifies the power of the Line-of-Sight (LoS) path in comparison to the Non-Line-of-Sight (NLoS) paths. In other words, for low values of κ, the channel is primarily dominated by NLoS components, while for high values of κ, the channel is primarily influenced by the LoS path. When κ equals 1, it signifies a balanced or equilibrium state where both LoS and NLoS components have a significant impact on the channel characteristics. This parameter plays a crucial role in characterizing the behavior and performance of wireless communication channels in various real-world scenarios. In general, the LoS path can be given by the steering vector which can have different expressions dependent on the arrangement of the array

[Björnson, Hoydis e Sanguinetti 2017]. Besides, the NLoS path can be represented by a random normal distribution $\mathcal{C}(0,1)$. Furthermore, from a statistical perspective, there is a distribution that serves as a generalization of Rician fading, namely Nakagami-m distribution [Goldsmith 2005].

1.8 Analytical Techniques

In order to propose innovative solutions and conduct comprehensive analyses of the challenges posed by future communication systems, this Master's Dissertation emphasizes the importance of analytical techniques that prioritize analytical tractability and low implementation complexity. These techniques enable us to effectively address and gain a deep understanding of the problems at hand. In the following sections, we outline each of these analytical techniques, highlighting their relevance and applicability to our research.

1.8.1 Moment Matching

One of the most important techniques to estimate unknown parameters of a distribution function in statistics is the method of moments or Moment Matching (MM) [Kobayashi, Mark e Turin 2011], which is widely applied in the literature in the most diverse kinds due to its easiness of analytical manipulation providing interesting insights with an interesting accuracy. Such technique estimates the parameters of a specific random variable through its moments, where the most common are the mean and variance, denoted by the $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$. Let X be a random unknown variable, where its mean and variance are given respectively by μ_X and σ_X^2; we can represent X through a, for instance, Gamma distribution ($\Gamma(k,\theta)$) with k and θ being the shape and scale parameters of the Gamma distribution. Therefore, we should solve the following system of equations

$$k\theta = \mathbb{E}[X] = \mu_X, \qquad k\theta^2 = \mathbb{V}[X] = \sigma_X^2, \tag{1.7}$$

This technique has been and continues to be widely applied in the literature in several works for different kinds of applications, for instance, in [Xie et al. 2022], the authors show that the random variable representing the channel gain of STAR-RIS-aided system can be accurately approximated as a Gamma distribution through the MM. This enables a comprehensive understanding of the system behavior through its parameters and extends the possibility of enhancing the system performance according to the available resources. In [Tahir, Schwarz e Rupp 2020, Tahir, Schwarz e Rupp 2021], similarly, the authors utilized the MM technique to estimate the channel gain distribution by a Gamma random variable of a RIS-aided system. The Method of Moments (MM) technique is highly regarded for its power in representing unknown random variables, or even complex distributions, such as communication channels, SINRs, etc., utilizing the easiest analytical

description given by the well-known distributions like Nakagami-m, Gamma, Normal, Weibull, among others. In this Master's Dissertation, we leverage the MM technique to evaluate the performance of both cooperative systems in the work described in the forthcoming Chapter 2 and double-RIS systems described in the forthcoming Chapter 4. By employing this technique, we gain valuable insights into the behavior and efficiency of these systems, contributing to the advancement of future communication technologies.

1.8.2 Lagrangian Dual

In addition, in this work, we deal with intricate optimization problems, whose optimal solution can be very hard or even impossible to obtain. Therefore, in view of proposing near-optimal solutions, we have leveraged the Lagrangian dual method [Boyd, Boyd e Vandenberghe 2004], which has a great tradition in the literature to address optimization problems with constraints. Basically, its principle is to provide a sub-optimal solution to a non-convex problem by reformulating the original objective function and conceiving the Lagrangian ($\mathcal{L}$) represented by the original objective function with a weighted sum of the constraints. Let the following convex problem, expressed in the standard form [Boyd, Boyd e Vandenberghe 2004]

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(\mathbf{x}) && && (1.8)\\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, && i = 1, \ldots, m, && (1.8a)\\ & h_i(\mathbf{x}) = 0, && i = 1, \ldots, p, && (1.8b) \end{aligned}$$

with $\mathbf{x} \in \mathbb{R}^n$, $f_0(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$, $f_i(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ and $h_i(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$. The basic idea in Lagrangian duality is to take the constraints (1.8a) and (1.8b) into account by augmenting the objective function with a weighted sum of the constraint functions. Thus, the Lagrangian is defined as

$$\mathcal{L}(\mathbf{x}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \omega_i h_i(\mathbf{x}), \tag{1.9}$$

where λ_i and ω_i are the Lagrange multipliers associated with the i-th inequality constraint $f_i(\mathbf{x})$ and i-th equality constraint $h_i(\mathbf{x})$, respectively. Therefore the Lagrangian Dual function can be defined as

$$\begin{aligned} \underset{\mathbf{x},\omega}{\text{maximize}} \quad & \mathcal{L}(\mathbf{x}, \omega) && && (1.10)\\ \text{subject to:} \quad & \omega_i \geq 0, && i = 1, \ldots, p. && (1.10a)\\ & \nabla\mathcal{L} = \nabla f_0(\mathbf{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\mathbf{x}^\star) + \sum_{i=1}^{p} \omega_i^\star \nabla h_i(\mathbf{x}^\star) = \mathbf{0}, && && (1.10b) \end{aligned}$$

where the optimal solution $\mathbf{x}^\star$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions given by Eq. (1.10b).

In [Guo et al. 2020] the authors utilized the Lagrangian method to optimize the SE of a RIS-aided system, where the power constraint of precoding matrix is addressed, in [Tang et al. 2020] the power allocation, and power splitting coefficient of a cooperative NOMA system have been jointly optimized by considering the minimum harvested energy and power constraint leveraging the Lagrangian dual-method. In [Li et al. 2018], the Lagrangian method is deployed to optimize the EE of a massive MIMO system by activating a different number of antennas and optimizing the power allocation of the users with the power and minimum rate constraint. In this Master's Dissertation, specifically, we have addressed an EE optimization problem with constraints in the context of RIS-aided communication systems.

1.8.3 Fractional Optimization

Fractional optimization is a methodology that aims to solve fractional problems. These kinds of problems are mathematical problems where the objective function or constraints involve fractional (non-integer) exponents or ratios. Moreover, these problems have gained significant attention in various fields due to their ability to model complex phenomena and capture non-linear relationships. The key distinction from traditional optimization problems is the presence of fractional terms, which introduce non-convexities and non-linearities into the optimization framework. Basically, there are two types of fractional problems, which will be reviewed in the following with their respective common techniques of solving.

1.8.3.1 Single Ratio Fractional Problems

A fractional problem is a class of optimization problems involving fractional ratios. The simplest case is known as the single-ratio problem. A single-ratio problem is defined as

$$\underset{\mathbf{x}}{\text{maximize}} \qquad \frac{A(\mathbf{x})}{B(\mathbf{x})}, \tag{1.11}$$

where $\mathbf{x} \in \mathbb{R}^n$, $A(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}_+$ is a nonnegative function and $B(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}_{++}$ is a positive function. Since dealing with the problem given in Eq. (1.11) can be hard, an alternative approach is to reformulate the problem, whereby its numerator and denominator are decoupled. With this method, the joint optimization of $A(\mathbf{x})$ and $B(\mathbf{x})$ becomes easier. A classical approach, and utilized in this Master' Dissertation, is known as Dinkelbach's Transform. This classical technique reformulates the single-ratio problem in Eq. (1.11) as [Dinkelbach 1967]

$$\underset{\mathbf{x}}{\text{maximize}} \qquad A(\mathbf{x}) - \eta B(\mathbf{x}), \tag{1.12}$$

where η is an auxiliary variable, which is iteratively updated by

$$\eta^{(t+1)} = \frac{A(\mathbf{x}^{(t)})}{B(\mathbf{x}^{(t)})}, \tag{1.13}$$

where t is the iteration index. It has been proved, in [Dinkelbach 1967], that convergence is guaranteed by alternatively updating η according to Eq. (1.13) and solving Eq. (1.12) for $\mathbf{x}$ because η is nondecreasing after each iteration. The classical Dinkelbach method has been extensively utilized in order to solve several problems in literature. In [Shen e Yu 2018], the authors proposed to utilize this method to solve an EE maximization problem. The authors in [Li et al. 2018] utilized the same method to solve a resource allocation problem, where a joint antenna number selection and power allocation algorithm has been proposed.

1.8.3.2 Multiple Ratio Fractional Problems

Differently from single-ratio problems, the multiple-ratio fractional problem is a class of linear or non-linear optimization problems where multiple ratios are involved. This kind of problem can be defined as

$$\underset{\mathbf{x}}{\text{maximize}} \qquad \sum_{k=1}^{K} \frac{A_k(\mathbf{x})}{B_k(\mathbf{x})}, \tag{1.14}$$

It is worth noting that classic transforms for single-ratio problems cannot be easily generalized to multiple-ratio fractional problems. This is because although these classic transforms have the property that the original single-ratio fractional problem and the transformed problem have the same optimal solution, the optimal value of the objective function of the transformed problem is not necessarily the same as the original single-ratio fractional problem objective function value. Thus, when multiple ratios are involved, one cannot apply the transform to each ratio individually. To deal with the multiple-ratio fractional problem, there are some techniques such as [Shen e Yu 2018], [Jong 2012]. One common technique for addressing the multiple-ratio fractional problem is named quadratic transform, which was proposed in [Shen e Yu 2018]. The quadratic transform reformulates the multiple-ratio problem in Eq. (1.14) as

$$\underset{\mathbf{x}}{\text{maximize}} \qquad \sum_{k=1}^{K} 2\eta_k \sqrt{A_k(\mathbf{x})} - \eta_k^2 B_k(\mathbf{x}), \tag{1.15}$$

where, again, η_k is an auxiliary variable for $k = 1, \ldots, K$, updated iteratively as

$$\eta_k^{(t+1)} = \frac{\sqrt{A_k(\mathbf{x}^{(t)})}}{B_k(\mathbf{x}^{(t)})}. \tag{1.16}$$

In [Shen e Yu 2018], the authors prove the convergence of the proposed transform and also extend this technique for the complex case. In literature, the authors of [Shen e Yu 2018], solved a beamforming problem with the quadratic transform. Besides, in [Guo et al. 2020], a beamforming, phase-shift RIS angles, and power allocation jointly optimization problem has been solved by deploying the quadratic transform technique.

1.9 Research Objectives

The objectives of the Master's Dissertation can be sub-divided in general and specific, and are presented in the following:

1. **General:** Study how the RIS deployment can impact the system performance with respect to the throughput as well as the reliability and energy efficiency of the system, operating under URRLC and mMTC services scenarios for communication system B5G. In addition, since only the deployment is not sufficient to obtain the best performance of the system, in this work we seek to understand how to set up the RIS, *i.e*, how to select the RIS elements number efficiently, as well as configure its phase-shift of each element in a better way according to a particular objective. Furthermore, we aim to develop an extensive mathematical analysis of the adopted system model in order to derive closed-form expressions with tiny complexity and high accuracy aiming to explore its potential to avoid huge complexity optimizations.
2. **Specific:** The specific objectives are threefold:
   a) Provide an extensive literature review of the RIS theme in several contexts, especially on mMIMO and NOMA, identifying the target scenarios, and the open problems associated with the integration of these technologies.
   b) Proposed algorithms to solve practical problems related to the throughput and EE, in RIS-assisted mMIMO/NOMA systems, specifically: dealing with the number of RIS and antenna elements that will be selected to act desiring to reduce the energy consumption in addition to the optimization of power resources.
   c) Disseminate the achieved results and findings by means of scientific articles.

1.10 Contributions and Generated Publications

The results of this Master's Dissertation are presented by means of three scientific articles and one research proposal. The finished works are available in Appendices A, B, and C, while a report outlining the research proposal for the fourth work is placed as the Chapter 5. The following works have been submitted or are under review.

[A] JUNIOR, W. D. S.; ABRAO, T. **RIS-aided Cooperative FD-SWIPT-NOMA Outage Performance in Nakagami-m Channels**. *Submitted to IEEE Transactions on Cognitive Communications and Networking on Jul. 2023* (IF: 8.6, Eng. IV Qualis-CAPES: A1)

[B] JUNIOR, W. D. S.; ABRAO, T. **Energy Efficiency Maximization for Intelligent Surfaces Aided Massive MIMO with Zero Forcing.** Full paper. *Submitted to IEEE Transactions on Green Communications and Networking on Jun. 2023* (IF: 4.8, Eng. IV Qualis-CAPES: A3)

[C] JUNIOR, W. D. S.; ABRAO, T. **NOMA-aided Double RIS under Nakagami-m Fading: Channel and System Modelling**. Conference paper. Accepted and presented in the Vehicular Technology Conference (VTC2023-Spring). June 2023.

Moreover, the research currently under development has the title and abstract as follows:

[D] JUNIOR, W. D. S.; ABRAO, T. **On the Capacity of Massive MIMO Aided by Double-RIS under Rician Channels**

This study analyzes the rate performance of a downlink single-user double RIS-aided massive MIMO system. The system assumes Rician fading with purely LoS components and adopts MR precoding. We derive an expression for the upper bound of SE in this configuration. By leveraging this derived expression, we propose an optimization method for the phase shifts of RIS 1 and RIS 2, using only the long-term CSI. The proposed method, based on the gradient ascent algorithm, achieves significant rates without the need for instantaneous CSI optimization or the associated CSI estimation overhead. Our numerical results demonstrate the effectiveness of the proposed method in achieving high rates while minimizing the impact of CSI estimation requirements.

1.11 Organization of the Text

The remainder of this text is organized as follows. In Chapters 2, 3, 4, we outline the results reported by the scientific articles, while in Chapter 5 we describe the current research proposal in the context of this Master's Dissertation. In Chapter 6, we provide partial conclusions and comments. Appendices A, B and C reproduce, respectively, the three scientific full papers associated with this Master's Dissertation.

2 RIS-aided Cooperative FD-SWIPT-NOMA Outage Performance over Nakagami-m Channels

In this chapter, we proposed to analyze the impact of RIS structure in cooperative communications. The associate paper is presented in Appendix A. The contributions of the first part of this Master's Dissertation are fourfold:

- We integrate the SWIPT technique by utilizing a non-linear EH model for cooperative scenarios with NOMA aiming to investigate for the first time the performance of the adopted system model as well as its benefits and drawbacks.
- We showed that the harvested power in the good channel condition device can be precisely represented by a Gamma distribution.
- We derive novel general expressions for the outage probability as well as for the ergodic capacity. The derived expressions are parameterized in the system/channel parameters, therefore the effects of each parameter can be effectively scanned as a function of the number of RIS elements.
- The accuracy of the derived analytical expressions has been corroborated by the Monte Carlo simulations, providing effectiveness in the analysis.

The system model for this work is shown in Fig. 6. We consider two devices, namely the good channel conditions device (D_1) and the poor channel conditions device (D_2) which are paired to perform NOMA. The Source (BS) is considered equipped with just one antenna, configuring a single-input single-output (SISO) scenario. D_1 leverages of SWIPT technique to harvest power in order to act as an FD relay under the residual self-interference, without utilizing its own battery energy to cooperate. We adopt the non-linear EH model in order to capture the real behavior of the harvested power. Furthermore, we adopt a RIS whose reflecting elements are configured to cancel the phase of D_1 user since it will operate as a relay to assist and provide better conditions of operation to D_2. Since RIS is adopted, whole communication links undergo Nakagami-m fading, where the LoS and NLoS effects are present.

In this work, we intend to answer the following questions: *a) Can the RIS improve cooperative communication between two devices?*; *b) What are the impacts of RIS on cooperative communication, and is it truly worthwhile?*

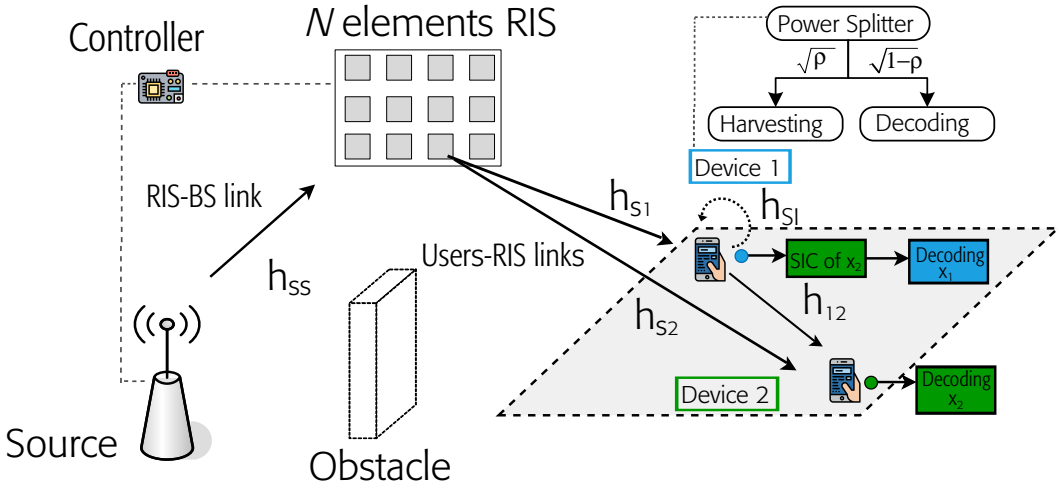
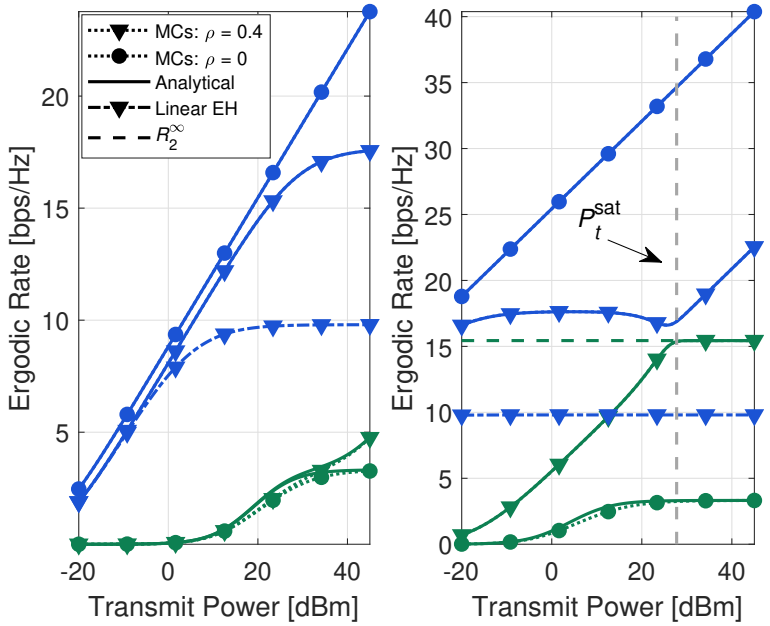


Figure 6 – System model studied in the Article attached in A.

2.1 Main Results

Fig. 7 depicts the ER of D_1 and D_2 *vs* transmit power (P_t) for $N = 400$ RIS-elements. Firstly, it is noticeable that the derived upper bounds expressions are very tight for the adopted parameters, closely matching the obtained results.



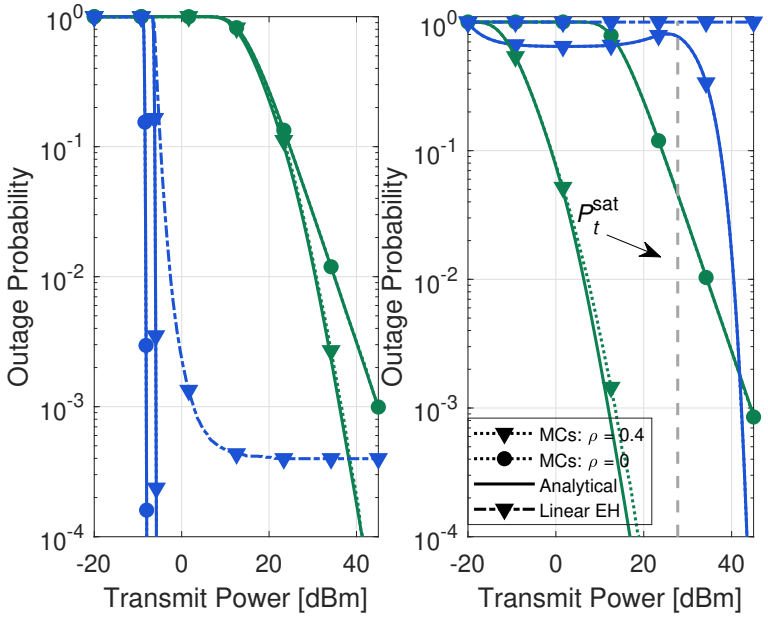
a) Scenario I b) Scenario II

Figure 7 – ER for D_1 and D_2, denoted as blue and green color, respectively. The number of elements of RIS $N = 20^2 = 400$, and energy harvested factor is adopted as $\rho = 0$ and $\rho = 0.4$ for non-cooperative and cooperative mode respectively.

In Fig. 7, the rate behavior is distinct for scenarios I and II. For scenario I,

we observe that in D_2, the rate increase becomes remarkable for transmit power values above 30 dBm. On the other hand, in D_1, the rate experiences a constant decrease, even for low values of P_t, and this decrease becomes more evident for high values of P_t. The reason for this difference lies in D_1 utilizing only $1 - \rho = 60\%$ of the received power for information detection, impacting its performance regardless of the transmit power. Moreover, for $P_t > 22$ dBm, the impact on the rate of D_1 in the cooperative scenario is further affected. The same impact occurs for scenario II, but it is more pronounced since the rate of D_1 is severely degraded for any value of P_t. This degradation is justified by the presence of residual self-interference. As the transmit power increases, the harvested power also increases, leading to higher residual self-interference, increased interference, and consequently decreased SINR of D_1.

In scenario II, we observe a further enhancement in the rate of D_2, demonstrating the potential of cooperative communication in scenarios where D_1 has high input power. However, it is worth noting that even with substantial rate increments, D_2 reaches a maximum limit when the transmit power increases considerably. We refer to this asymptotic behavior as R_2^{∞}. This observation is insightful, as it confirms that the cooperative system with RIS cannot provide an unlimited rate for D_2, even in the asymptotic regime. Additionally, this result shows that N does not influence the value of R_2^{∞}, indicating that the RIS cannot impact the asymptotic rate of D_2.



a) Scenario I b) Scenario II

Figure 8 – OP for D_1 and D_2, denoted as blue and green color, respectively. The number of elements of RIS $N = 20^2 = 400$, and energy harvested factor is adopted as $\rho = 0$ and $\rho = 0.4$ for non-cooperative and cooperative mode respectively.

In terms of OP, Fig. 8 again reveals distinct behaviors for scenarios I) and II. For scenario I, as P_t increases, the cooperative mode becomes beneficial, resulting in

a reduction in the outage probability of D_2 for $P_t > 30$. However, the impact of the cooperative mode on D_1 is minimal, indicating that it does not significantly affect its performance. In contrast, for scenario II, we observe an interesting reduction in outage probability for D_2 across a wide range of P_t. This reduction is particularly noteworthy as it comes without any additional power consumption at the source, making it a promising gain in terms of sustainability. However, the cooperative mode leads to an increase in outage probability for D_1, severely affecting its performance. Nevertheless, it is essential to note that despite the significant increment in OP for D_1, it does not reach an outage floor. This is a vital observation as, for the linear EH model, D_1 achieves an outage limit in both scenarios I and II. These findings highlight the benefits and trade-offs of cooperative FD-SWIPT-NOMA systems with RIS. The cooperative mode offers substantial improvements in outage probability for D_2, promoting sustainability without additional energy consumption. However, it may negatively impact the outage probability for D_1, requiring careful considerations in system design, mainly over ω, to strike a balance between performance gains and potential drawbacks.

2.2 Chapter Conclusions

In this work, we have investigated the downlink communication reliability and throughput of two devices in RIS-aided cooperative SWIPT-NOMA system operating under Nakagami-m fading; the D_1 device leverages of SWIPT technique equipped with a non-linear EH model to act as an FD-DF relay for assisting the D_2 device. So, we investigate the impact of the deployed RIS on the system performance, indicating that RIS can significantly improve the performance of both devices. Finally, the simulation results validate the correctness and effectiveness of the developed theoretical analysis. Based on the obtained results, we can answer the proposed questions.

a) Can the RIS improve cooperative communication between two devices?

The presented results clearly indicate that the cooperative SWIPT-NOMA system assisted by a RIS can substantially improve the overall system performance. While the cooperative system may lead to a decrease in the performance of D_1, it simultaneously enhances the rates of D_2 and improves overall reliability. Moreover, the presence of the RIS enhances the capabilities of D_1, reducing the negative impact of the cooperative communication in D_1. In addition, it can securely assist D_2 in challenging conditions, creating an intriguing synergy between these technologies.

b) What are the impacts of RIS on cooperative communication, and is it truly worthwhile?

The numerical findings demonstrate that as the size of the RIS increases, the received power at the user relay substantially improves, resulting in enhanced harvested power. Consequently, the interference caused by the residual self-interference increases, negatively impacting the performance of D_1. However, simultaneously, the cooperative link strengthens, leading to an overall improvement in the rate of D_2. Furthermore, the numerical results highlight the paramount importance of developing robust circuits to mitigate the residual self-interference, as it significantly influences the performance of D_1. Addressing this issue becomes crucial to improving the overall system performance.

3 On the Energy Efficiency Optimization for RIS-aided massive MIMO with Statistical CSI

In this Chapter with the associated paper included in Appendix B, we intend to investigate how the deployment of RIS structure can enhance the overall system EE. The contributions of the second part of this Master's Dissertation are outlined as follows:

- Based on a tight closed-form expression of the ER, we formulate a joint optimization problem concerning the users' transmit power ($\mathbf{p}$), the number of active BS antennas (M), and RIS phase shift angles ($\mathbf{v}$) variables under the constraints of overall power budget, ZF precoding feasibility, QoS, and unity module requirements.
- To deal with this problem, we apply the Lagrangian dual method, and some fractional programming techniques in order to obtain closed-form solutions for all optimization variables.
- The complexity of the proposed optimization algorithm has been accurately investigated.
- Extensive numerical results corroborate the effectiveness and efficiency of the proposed analytical EE optimization method for RIS-aided mMIMO systems with statistical CSI.

The communication scenario being addressed is illustrated in Fig. 9. Since the RIS is considered strategically to lie in a favorable position, the Rician fading model is considered to account for the combined effects of NLoS and LoS.

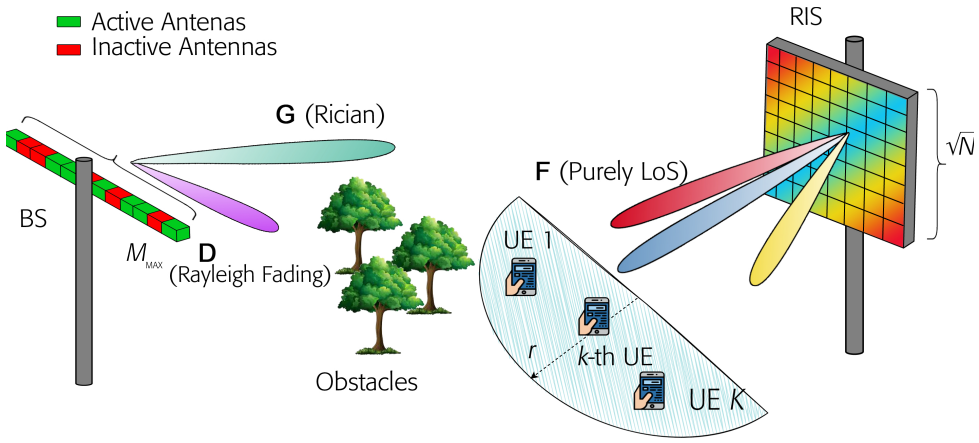


Figure 9 – Studied system model in the Article addressed in Appendix B.

Different from other works in the literature that focus on the optimization of EE using the instantaneous CSI strategy, in this work, we consider the statistical CSI optimization strategy, which is better explained in Fig. 10.

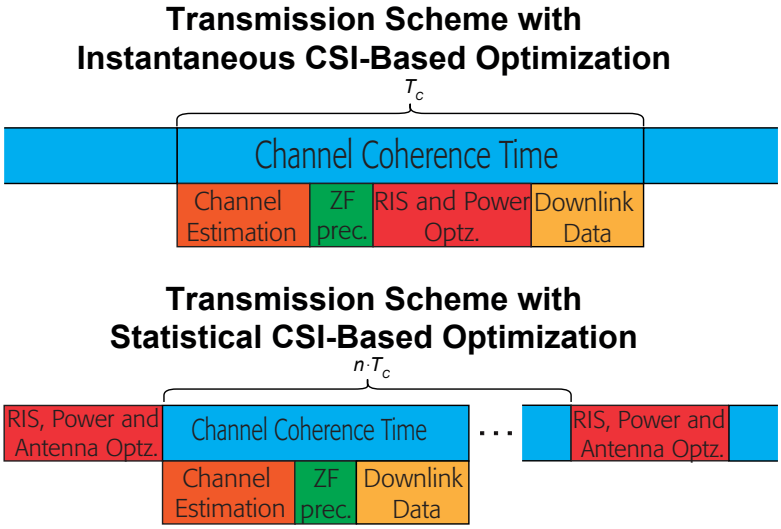


Figure 10 – Illustration of the advantages by adopting Statistical CSI-based parameters optimization, where the RIS phase-shift vector, transmit power allocation, and the number of BS antennas remain fixed for the users over multiple channel coherence times (nT_c); differently, for the Instantaneous CSI-based parameter optimization approach, where these variables must be updated at each T_c.

Besides, in this work, we adopt the Zero-forcing (ZF) linear precoding. Based on a closed-form expression for SE, we formulate an optimization problem aiming to maximize the EE with respect to the number of BS active antennas (M), power allocation ($\mathbf{p}$) for the users, and phase-shift matrix ($\mathbf{v}$), given by Eq. (13), Appendix B, reproduced here for sake of clarity.

$$\underset{M,\mathbf{p},\mathbf{v}}{\text{maximize}} \quad \eta_{\text{EE}} = \frac{\sum_{k=1}^{K} \log_2\left(1 + \frac{p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)}{\varrho\sum_{k=1}^{K} p_k + P_{\text{FIX}} + MP_{\text{BS}} + NP_{\text{RIS}}} \tag{3.1}$$

$$\text{subject to} \quad \sum_{k=1}^{K} p_k \leq P_{\text{TX}}, \tag{3.1a}$$

$$\widetilde{R}_k \geq R_{\min}, \quad \forall k = 1, \ldots, K, \tag{3.1b}$$

$$M > K, \tag{3.1c}$$

$$|e^{j\theta_n}| = 1, \quad \forall n = 1, \ldots, N, \tag{3.1d}$$

In short, we seek to answer the following questions: *a) Can the RIS, with a constant phase-shift matrix configured over multiple coherence times, achieve promising performance?; b) Is the statistical-CSI optimization strategy worth it in terms of EE?*

3.1 Main Results

Aiming to answer these crucial questions, we analyze Fig. 11, which illustrates the achievable EE performance as a function of the transmit power budget at BS, P_{TX}. We evaluated the performance for the four variable optimizations strategies: 1) only $\mathbf{p}$; 2) $\mathbf{p}$ and $\mathbf{v}$; 3) $\mathbf{p}$ and M; and 4) $\mathbf{p}$, $\mathbf{v}$ and M optimization, denoted as yellow, red, green, and blue color respectively. Moreover, we adopt as benchmarks the random all ($\mathbf{p}$, $\mathbf{v}$, and M) strategy, and the instantaneous CSI-based optimization strategy, which has been implemented based on Algorithm 1 proposed in [Miridakis, Tsiftsis e Yao 2023], denoted herein as black and purple color curves, respectively. We also plotted the average instantaneous EE obtained from the proposed statistical CSI-based optimization analytical approach evaluated by Monte Carlo Simulation (MCs), depicted in orange color.

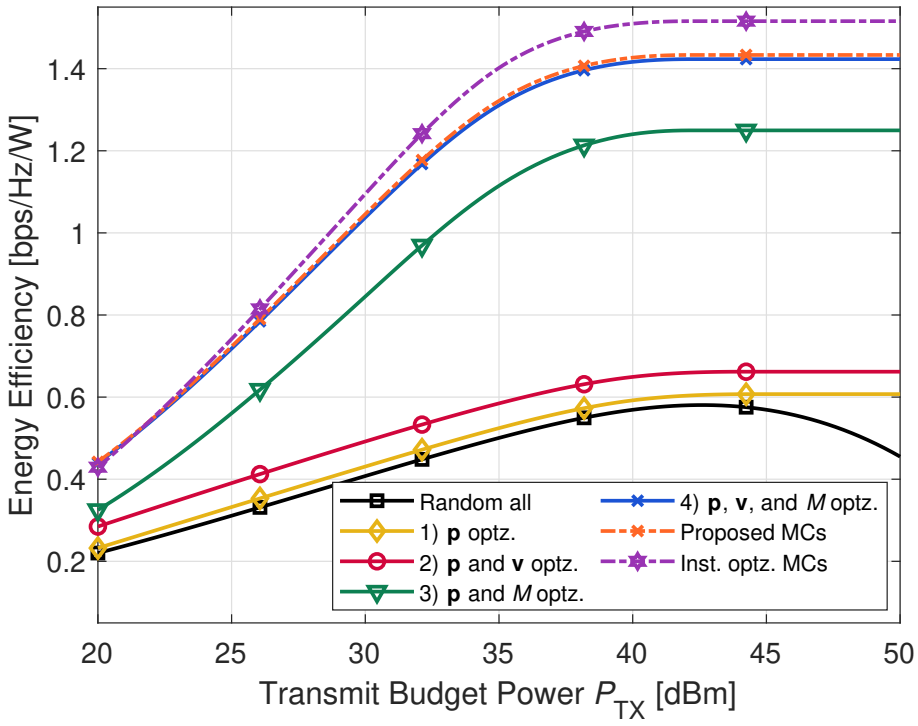


Figure 11 – Average EE *vs* the transmit power budget at the BS (P_{TX}). Performance evaluation of the proposed algorithm for four different approaches.

First and foremost, in Fig. 11, one can infer that the impact of optimizing the phase shift RIS angles is not negligible on the attainable EE, even considering the statistical optimization approach. It is noteworthy that the achieved EE improvement in scenarios in which the number of active BS antennas M is optimized, is considerably greater compared to those scenarios where M is not optimized. This can be readily verified by examining the obtained gain in the EE curve of strategy 1) compared to the curve of strategy 2), as well as in the curve of strategy 3) with respect to the curve of strategy 4).

Furthermore, also based on Fig. 11, one can see the potential to achieve higher EE through the statistical CSI optimization approach since this methodology enables us

to reduce the overhead associated with the phase-shift optimization during each channel coherence time while maintaining the fixed $\mathbf{v}$ over multiples channel coherence time with a minimal performance loss. As a result, the EE decreases from 1.510 to 1.433 bps/Hz/W, representing a loss of $\approx 5.3\%$, in high power regimes.

3.2 Chapter Conclusions

This work handles the EE maximization problem in a downlink RIS-aided mMIMO communication system operating under generalized Rician BS-RIS channels and ZF precoding. We proposed a statistical CSI-based optimization strategy, reducing the overhead of the variables optimization in every channel coherence time. Based on an ergodic rate (ER) lower-bound closed-form expression, the EE optimization problem with power and QoS constraints is formulated. We proposed an alternating-sequential optimization approach, that optimizes the power allocation vector ($\mathbf{p}$), the number of BS antennas (M), and the RIS phase-shift vector ($\mathbf{v}$). Our numerical results validate the effectiveness of the proposed Algorithm. Based on the obtained results, we can answer the proposed questions.

a) Can the RIS, with a constant phase-shift matrix configured over multiple coherence times, achieve promising performance?

According to the presented results in the work addressed in Appendix B, it can be observed that optimizing the phase shift angles of the RIS $\mathbf{v}$ with Statistical-CSI optimization strategy, can provide remarkable gains. This approach provides higher rates for the users, decreases the number of active antennas, and reduces power consumption. Additionally, it offers the advantage of an interesting overhead gain, as it does not necessarily need to be updated at every coherence time.

b) Is the statistical-CSI optimization strategy worth it in terms of EE?

According to Fig. 11, we can see that the Statistical-CSI optimization strategy can be promising in terms of EE, as the EE obtained from the Statistical-CSI optimization reduces from 1.510 to 1.433 bps/Hz/W, representing a loss of $\approx 5.3\%$, in high power regimes (maximum difference observed), when compared to the Instantaneous-CSI optimization strategy. This finding highlights the potential of the Statistical-CSI optimization approach to deliver favorable EE outcomes in RIS-aided communication systems.

4 NOMA-aided Double RIS under Nakagami-m Fading: Channel and System Modelling

In this Chapter, as addressed by the paper in Appendix C, we investigate a SISO double-RIS setup including channel and system modeling, with the application of the NOMA transmit scheme. The contributions of the third part of this Master's Dissertation are summarized in the following:

- With the aid of the MM technique and Central Limit Theorem (CLT), we statistically characterize the cascaded channel between the two RIS where all links are adopted as Nakagami-m fading.
- We derived a closed-form equation for outage probability and ergodic rate for this system setup, where different system and channel parameters are taken into account.
- We assessed the impact of the number of reflective elements of each RISs on the system performance.

The double RIS NOMA-aided communication scenario is illustrated in Fig. 12. We consider two RIS deployed, where the near BS RIS is a conventional RIS with N_C elements, and in order to provide a better degree of coverage, we adopt the near user RIS as STAR-RIS [Xu et al. 2021, Xu et al. 2021] topology with N_S elements, serving the outdoor and indoor users simultaneously.

Figure 12 – System model studied in the Article reproduced in Appendix C.

In order to assess the best performance of the adopted system model, we configure the STAR-RIS to coherently combine the phases of indoor/outdoor user, *i.e.*, the optimal phase-shift is adopted, while the conventional RIS is configured to coherently combine the phases of the outdoor user.

In this work, we intend to answer the following question *a) Is the performance of a double-RIS configuration better than the conventional single-RIS setup?; b)In which conditions the double-RIS setup outperforms the single-RIS setup for the SISO case?.*

4.1 Main Results

In view of answering these questions, we derived closed-form expressions for ergodic capacity, as well as the outage probability of indoor and outdoor users. Fig. 13, which illustrates the achievable ergodic sum-rate performance as a function of the split factor, defined as $\eta = \frac{N_C}{N}$, with $N \triangleq N_C + N_S$, for three different path-loss coefficients, $\alpha_t = \{2.8; 3.1; 3.4\}$, where N_C and N_S is the number of elements in the conventional RIS and the STAR-RIS respectively. We first notice that for all scenarios considered the analytically derived equations given by Eq. (13) in Appendix C are very accurate for any value of η. Furthermore, we can see that, by improving the conditions of the direct link between the source and the STAR-RIS, the ER performance of the system can be improved; the double-RIS setup can be outperformed by the single-RIS setup when $\alpha_t \leq 2.8$.

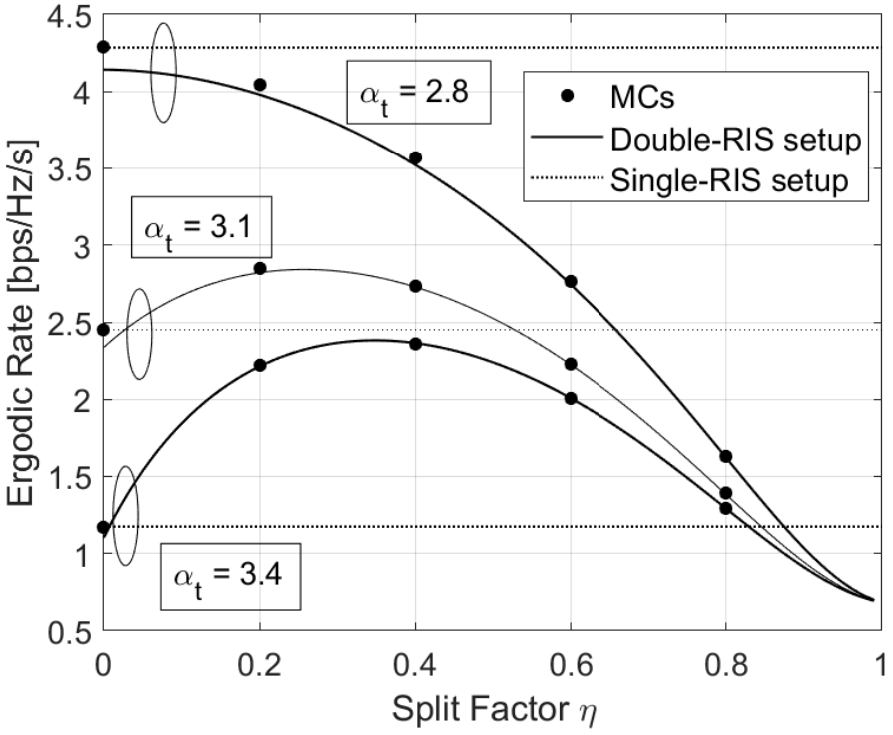


Figure 13 – Ergodic sum-rate *vs* split factor (η) over an NOMA-aided conventional and STAR-RIS system *vs* single-RIS setup with $N = 200$ and SNR $\rho = 35$ dB.

On the other hand, when $\alpha_t = \{3.1, 3.4\}$, the double-RIS setup can obtain higher data rates than the single-RIS setup for a wide range of η, revealing its great potential in scenarios where the direct link between the Source and STAR-RIS is poor. Finally, we can see the scenarios where the double-RIS setup performs well the number of meta-surface elements N_S and N_C can be strategically optimized.

Fig. 14 reveals a result of work embedded in Appendix C, where it illustrates the ergodic sum-rate *vs.* total number of reflective elements for three different values of η. As a benchmark, we include the random phase shift solution (RPS) for both a) STAR-RIS and Conventional RIS; b) only for Conventional RIS; and c) only for STAR-RIS. Besides, we also include the single-RIS setup with a near-user STAR-RIS deployed with N elements for a fair comparison, where the phase shift is set as the optimal phase shift.

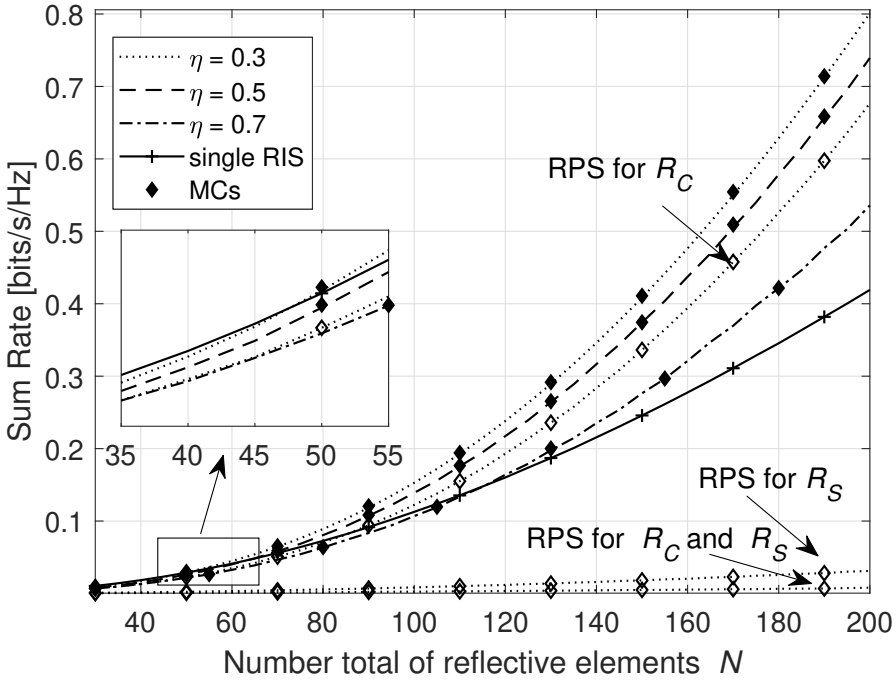


Figure 14 – Ergodic sum-rate *vs* total number of reflective elements (N) for three different values of η, with SNR $\rho = 35$ dB.

First, we observe the significance of optimizing the angles of STAR-RIS. This is due to STAR-RIS serving both indoor and outdoor users simultaneously, while Conventional RIS serves only the outdoor user. Additionally, the highest sum rate is achieved when $\eta = 0.3$, indicating the STAR-RIS must be larger than the Conventional RIS. This is attributed to the greater impact of STAR-RIS, as the indoor user rate, given by Eq. (13) in Appendix C, which grows with N_S^2. We also can notice that the single-RIS setup outperforms the double-RIS only in scenarios where N is low, typically $N < 50$; however, when $N > 50$ the double-RIS setup can provide greater sum rates for any value of η, till to the random phase shift configuration.

Another interesting finding in this work is that channel estimation is necessary for

defining the optimal phase shift in STAR-RIS. For instance, when $\eta = 0.3$, in the single-RIS setup, channel estimation is required for $N = 200$ reflecting elements. In contrast, in the double-RIS configuration, the channel estimation must be done for $N_S = N(1-\eta) = 140$ reflecting elements.

4.2 Chapter Conclusions

In this paper, we have characterized the system performance of a NOMA-aided conventional and STAR-RIS under Nakagami-m channels. We derived analytical equations for outage probability and ergotic rate using the CLT and MM techniques. The numerical (analytical and MC simulation) results reveal that in scenarios where the direct link between the Source and STAR-RIS is in poor conditions, the double-RIS outperforms the single-RIS setup increasing the users' rate and consequently decreasing the outage probability. Besides, within this scenario, when the values of R_C and R_S are suitably adjusted, the double-RIS has the potential to achieve very high rates than the single-RIS setup.

Moreover, another interesting finding, is that the double-RIS setup can achieve promising gains with lower channel estimation overhead than the single-RIS setup. Finally, the simulation results demonstrated that the analytical proposed equations are very accurate under the parameters of the system. Based on the obtained results, we can answer the proposed questions:

a) Is the performance of a double-RIS configuration better than the conventional single-RIS setup?

According to the findings in the work available in Appendix C, we can see that under specific system configuration and channel conditions, the double RIS setup exhibits significant potential to outperform the single RIS setup. This improvement is observed in terms of both outage probability and ergodic rate. Besides, we see that it is possible to achieve higher sum rates for the double-RIS setup with lower channel estimation overhead since the double-RIS can operate under random phase shift for the near BS RIS as illustrated in Fig. 14.

b)In which conditions the double-RIS setup outperforms the single-RIS setup for the SISO case?

According to Fig. 13, one can observe scenarios where the direct link between the source and the STAR-RIS is poor, the double-RIS has great potential for operating since it can provide alternate ways to improve the channel conditions for the users.

5 On the Capacity of Massive MIMO Aided by Double-RIS under Rician Channels

This research consists in analyzing the combined scenarios established by the previous Chapters 3 and 4, *i.e.*, the works in Appendix B and Appendix C. Specifically, in this Chapter, we proposed to seek to combine the mMIMO and double-RIS systems, differently from the previously discussed works which deploy a single-RIS structure. Specifically, we intend to approach a double-RIS setup as in [Zheng, You e Zhang 2021] but with deploying mMIMO technology and ZF precoding, with the goal of deriving the ergodic capacity expression in closed-form following the methodology proposed in [Zhi et al. 2022].

The motivation for deriving a closed-form expression for the ergodic rate of double-RIS mMIMO systems is that it can provide interesting insights about the double-RIS performance; in addition, such an approach allows avoid the necessity of optimizing the phase-shift matrix in each channel coherence time interval, since for that, the CSI knowledge is necessary, whereas the double-RIS setup imposes a huge complexity in the channel estimation process; hence, the proposed approach can be beneficial in the complexity viewpoint. Moreover, it can possibly achieve reasonable performance in terms of ER. Furthermore, the double-RIS setup can be justified due to the recent promising results reported in [Zheng, You e Zhang 2021] and the necessity of new algorithms addressing such double RIS-aided mMIMO system configuration.

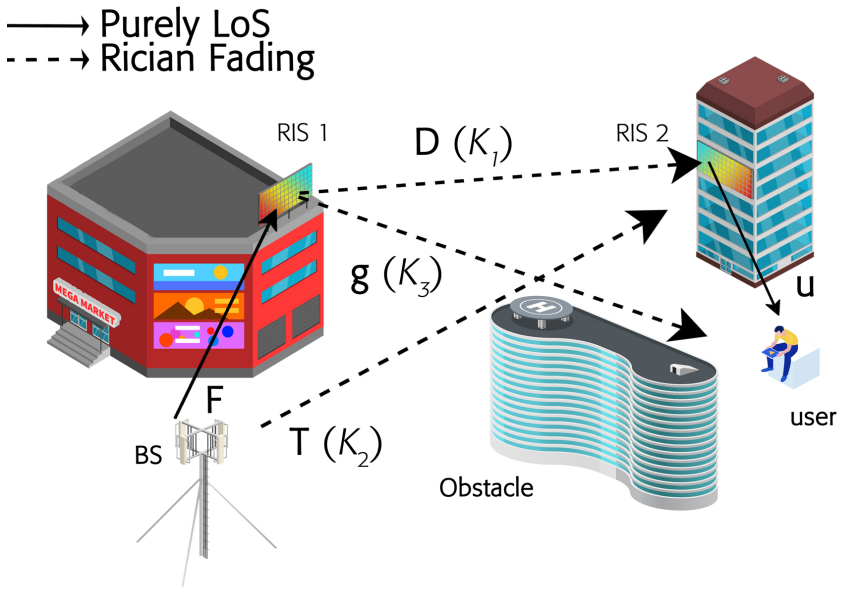


Figure 15 – A representation for the double-RIS-aided mMIMO communication in the current research.

In addition, other considerations can be adapted to become this study further attractive, for instance, we can also integrate a hybrid active-passive RIS architecture as in [Kang, You e Zhang 2022], in order to comprehensively explore the potential of the RIS communications in terms of SE and EE.

5.1 Addressed Problem

With the double-RIS deployed, naturally, the channel estimation task is hugely increased concerning the conventional single-RIS setup. Besides, in the recent literature, the most of works focus on optimizing the angles of the RIS in an instantaneous way, *i.e*, in each coherence time interval, it is necessary to estimate the CSI, so that the angles of the RIS are optimized. Evidently, all these tasks are costly to be processed in a short time period. In order to reduce this complexity related to the aforementioned task, deploying the statistical CSI optimization strategy result in an interesting solution. In order to explore the double-RIS setup over the single-RIS setup, in this Chapter, we look for deriving a closed-form expression for the ergodic rate of a double-RIS setup taking into account the statistical CSI approach instead of burden instantaneous CSI optimization. Furthermore, based on a closed-form expression, we intend to optimize the angles of both deployed RISs.[1]. In short, in this current work

1. We analyze the performance gains/drawbacks of the double-RIS setup over the single-RIS setup, especially integrated into the massive MIMO systems. Besides, we seek to understand how such performance is scalable regarding the SISO double-RIS setup studied in Chapter 4;

2. We derive a closed-form expression, easy to compute, for the ergodic rate of a double-RIS setup;

3. Based on a closed-form expression for the double-RIS setup, we optimize the phase-shift matrix of both deployed RISs in view of either increasing the SE of the users or the EE of the system;

4. We study the double-RIS mMIMO system behavior operating under different system and channel scenarios.

[1] It is noteworthy that the statistical CSI optimization procedure is effective in scenarios where the users are assumed static or with very low mobility, so the RIS phase-shift matrix can be set as fixed over a long time period.

5.2 Proposed Solution

Aiming to provide valuable insights, we leverage random matrix theory to derive a closed-form expression for the ergodic capacity of a double-RIS system. Furthermore, looking to obtain a tractable solution for the optimization of the phase-shift matrix of RISs we seek to apply the Alternate Optimization (AO) method. The AO can be seen as a conventional optimization method where in double RIS-aided communications it can optimize each phase-shift matrix separately and alternately, while another one is set fixed. Moreover, to optimize each phase-shift matrix, we apply the conventional gradient ascent method. The proposed solution is detailed in Section 5.5, and the numerical results can demonstrate its potential. For the considered problems, techniques such as Machine-Learning, through its strands such as reinforcement learning or supervised learning, and even heuristic evolutionary optimization, including the genetic algorithm, can be promising as well.

5.3 System Model

We consider a downlink double-RIS-aided mMIMO communication system, where a single-antenna user communicates with a base station (BS) equipped with M antennas and aided by two distributed RIS. The RISs are located near BS (referred to as RIS 1) and near the user (referred to as RIS 2). Let N be the total number of passive elements for the two RIS, where RISs 1 and 2 consist of N_1 and N_2 elements respectively, with $N_1 + N_2 = N$.

We can define the channel between the BS→RIS 1, the channel between RIS 1→user, the channel between RIS 2→user, the channel between RIS 1→RIS 2 and the channel between the BS→RIS 2 as $\mathbf{F} \in \mathbb{C}^{M\times N_1}$, $\mathbf{g} \in \mathbb{C}^{N_1\times 1}$, $\mathbf{u} \in \mathbb{C}^{N_2\times 1}$, $\mathbf{D} \in \mathbb{C}^{N_1\times N_2}$ and $\mathbf{T} \in \mathbb{C}^{M\times N_2}$ respectively. We consider the practical scenario where the direct link between the user and the BS is severely obstructed by obstacles (dense obstacles, *e.g.*, walls which generally occur in indoor scenarios). The RISs are assumed to be positioned strategically (according to Fig. 15) in order to channel matrix $\mathbf{F}$ to be purely LoS channel as well as the channel $\mathbf{u}$, besides, the channels are Rician fading and give as

$$\mathbf{D} = \sqrt{\frac{\beta_{\mathbf{D}}}{K_1+1}}\mathbf{D}_{\text{NLoS}} + \sqrt{\frac{\beta_{\mathbf{D}}K_1}{K_1+1}}\mathbf{D}_{\text{LoS}}, \qquad \text{RIS1} \rightarrow \text{RIS2} \tag{5.1}$$

$$\mathbf{T} = \sqrt{\frac{\beta_{\mathbf{T}}}{K_2+1}}\mathbf{T}_{\text{NLoS}} + \sqrt{\frac{\beta_{\mathbf{T}}K_2}{K_2+1}}\mathbf{T}_{\text{LoS}}, \qquad \text{BS} \rightarrow \text{RIS2} \tag{5.2}$$

$$\mathbf{g} = \sqrt{\frac{\beta_{\mathbf{g}}}{K_3+1}}\mathbf{g}_{\text{NLoS}} + \sqrt{\frac{\beta_{\mathbf{g}}K_3}{K_3+1}}\mathbf{g}_{\text{LoS}}, \qquad \text{RIS1} \rightarrow \text{User} \tag{5.3}$$

where K_1, K_2 and K_3 denotes the Rician factor. The LoS components of all links are expressed in terms of array responses (steering vector), which are dependent on the array geometry. Without loss of generality, in this work, we consider that the RIS 1, RIS 2, and BS are equipped with a Uniform Planar Array (UPA), thus, the LoS component are given respectively by

$$\begin{aligned}
&\mathbf{F} = \mathbf{a}_M(\varphi_{br_1}^{\text{AoD}}, \phi_{br_1}^{\text{AoD}})\mathbf{a}_{N_1}(\varphi_{br_1}^{\text{AoA}}, \phi_{br_1}^{\text{AoA}})^H && \text{BS} \rightarrow \text{RIS1}\\
&\mathbf{g}_{\text{LoS}} = \mathbf{a}_{N_1}(\varphi_{r_1u}^{\text{AoD}}, \phi_{r_1u}^{\text{AoD}}) && \text{RIS1} \rightarrow \text{User}\\
&\mathbf{u} = \mathbf{a}_{N_2}(\varphi_{r_2u}^{\text{AoD}}, \phi_{r_2u}^{\text{AoD}}) && \text{RIS2} \rightarrow \text{User}\\
&\mathbf{D}_{\text{LoS}} = \mathbf{a}_{N_1}(\varphi_{r_1r_2}^{\text{AoD}}, \phi_{r_1r_2}^{\text{AoD}})\mathbf{a}_{N_2}^H(\varphi_{r_1r_2}^{\text{AoA}}, \phi_{r_1r_2}^{\text{AoA}}) && \text{RIS1} \rightarrow \text{RIS2}\\
&\mathbf{T}_{\text{LoS}} = \mathbf{a}_M(\varphi_{br_2}^{\text{AoD}}, \phi_{br_2}^{\text{AoD}})\mathbf{a}_{N_2}^H(\varphi_{br_2}^{\text{AoA}}, \phi_{br_2}^{\text{AoA}}) && \text{BS} \rightarrow \text{RIS2}
\end{aligned}$$

where φ (ϕ) are the azimuth (elevation) angles and it can be Angle of Departure (AoD) or Angle of Arrival (AoA), with br_1, r_1u, r_2u, r_1r_2 and br_2 denoting the BS-RIS 1, RIS 1-user, RIS 2-user, RIS1-RIS2, and BS-RIS 2 path respectively. Since only UPA is adopted herein, the array response of BS, RIS 1, and RIS 2 share similar expressions and can be given as

$$\mathbf{a}_X(\varphi, \phi) = \left[e^{j\mathbf{k}(\varphi,\phi)^T\mathbf{r}_1}, \ldots, e^{j\mathbf{k}(\varphi,\phi)^T\mathbf{r}_X}\right]^T, \tag{5.4}$$

$$\mathbf{k}(\varphi, \phi) = \frac{2\pi d}{\lambda}\begin{bmatrix}\sin(\varphi)\cos(\phi)\\ \sin(\phi)\end{bmatrix}, \tag{5.5}$$

$$\mathbf{r}_i = \begin{bmatrix}\text{mod}(i-1, X_h)\\ \lfloor (i-1)/X_h \rfloor\end{bmatrix}, \qquad i = 1, \ldots, X \tag{5.6}$$

where $X \in \{M, N_1, N_2\}$, d denotes the element spacing (which is assumed to be uniform along the UPA, *i.e.*, the vertical and horizontal distance between elements are equal) and λ denotes the wavelength. The superimposed channel combining the two single-reflection links and the double-reflection link can be given as follows[2]

$$\mathbf{h} = \underbrace{\sqrt{\beta_{\mathbf{F}}\beta_{\mathbf{g}}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{g}}_{\text{single-reflection}} + \underbrace{\sqrt{\beta_{\mathbf{T}}\beta_{\mathbf{u}}}\mathbf{T}\boldsymbol{\Phi}_2\mathbf{u}}_{\text{single-reflection}} + \underbrace{\sqrt{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{D}\boldsymbol{\Phi}_2\mathbf{u}}_{\text{double-reflection}}, \tag{5.7}$$

where $\boldsymbol{\theta}_\mu = [e^{j\theta_1^\mu}, e^{j\theta_2^\mu}, \ldots, e^{j\theta_{N_\mu}^\mu}]^T$, denote the reflection coefficients of RIS μ, and $\boldsymbol{\Phi}_\mu = \text{diag}(\boldsymbol{\theta}_\mu)$ the phase-shift matrix of RIS μ, where $\mu \in \{1, 2\}$.[3] We assume the quasi-static flat-fading channel model for all the channels, which remain approximately constant within each channel coherence interval. During the download data transmission, the received signal at the user can be given as follows

$$y = \sqrt{\rho}\mathbf{w}^H\mathbf{h}x + z, \tag{5.8}$$

[2] It is noteworthy that although there is one more double-reflection link between BS→RIS 2→RIS 1→user, it is neglected in this paper due to high path-loss degradation.

[3] Herein we consider the reflection amplitudes of all surfaces as one, that is, $|e^{j\theta_n^\mu}| = 1 \; \forall \; \mu, n$ in order to evaluate the maximum performance of the system.

where ρ is the downlink transmit power allocated to the user; $\mathbf{w} \in \mathbb{C}^{M\times 1}$ is the precoding vector; x is the transmitted information symbol with $\mathbb{E}[|x|^2] = 1$ and z is the additive white Gaussian noise with zero mean and variance σ^2, $z \sim \mathcal{CN}(0, \sigma^2)$. The SNR, which is defined as the ratio between the signal power and the noise power, can be computed as:

$$\gamma = \frac{\rho}{|z|^2} \frac{|\mathbf{w}^H \mathbf{h}|^2}{||\mathbf{w}||^2}. \tag{5.9}$$

where we normalize the precoding power by $\frac{1}{||\mathbf{w}||^2}$. Hence, the achievable rate can be given by

$$R = \log_2 \left(1 + \frac{\rho}{|z|^2} \frac{|\mathbf{w}^H \mathbf{h}|^2}{||\mathbf{w}||^2}\right). \tag{5.10}$$

5.4 Instantaneous CSI-based Optimization

In this section, we formulate an optimization problem in order to maximize the rate by optimizing the active beamforming at BS and both passive beamforming at the RISs by considering the perfect CSI known at the BS. Accordingly, the problem can be formulated as

$$\max_{\mathbf{w},\mathbf{\Phi}_1,\mathbf{\Phi}_2} \quad \log_2 \left(1 + \frac{\rho}{|z|^2} \frac{|\mathbf{w}^H \mathbf{h}|^2}{||\mathbf{w}||^2}\right) \tag{5.11}$$

$$\text{s.}t. \ |e^{j\theta_n^\mu}| = 1, \qquad \forall n = 1, \ldots, N_\mu, \qquad \mu \in \{1, 2\}, \tag{5.11a}$$

where μ represents the first and second RIS.

This problem has been tackled by [Zheng, You e Zhang 2021] and through the AO technique, an improved near-optimal solution has been proposed. The solution proposed in [Zheng, You e Zhang 2021] optimizes all variables separately; thus, in the following, we give the solution of each variable in closed form.

5.4.1 Optimizing phase-shift matrix of RIS 1 with fixed precoding matrix and fixed phase-shift matrix of RIS 2

The optimal passive beamforming at RIS 1 when $\mathbf{w}$ and $\mathbf{\Phi}_2$ are fixed can be computed as [Zheng, You e Zhang 2021, Eq. 21]

$$\boldsymbol{\theta}_1^\star = e^{-j\left(\angle \mathbf{w}^H \mathbf{Q} + \angle \mathbf{w}^H \mathbf{T} \mathbf{\Phi}_2 \mathbf{u}\right)}, \tag{5.12}$$

with $\mathbf{Q} = \mathbf{F}\text{diag}(\mathbf{g}) + \sum_{i=1}^{N_2} e^{j\theta_2^i} \mathbf{u}_i \mathbf{F}\text{diag}(\mathbf{D}_{(:,i)})$.

5.4.2 Optimizing phase-shift matrix of RIS 2 with fixed precoding matrix and fixed phase-shift matrix of RIS 1

The optimal passive beamforming at RIS 2 when $\mathbf{w}$ and $\mathbf{\Phi}_1$ are fixed can be computed as [Zheng, You e Zhang 2021, Eq. 20]

$$\boldsymbol{\theta}_2^{\star} = e^{-j\left(\angle \mathbf{w}^H \mathbf{A} + \angle \mathbf{w}^H \mathbf{F} \mathbf{\Phi}_1 \mathbf{g}\right)}, \tag{5.13}$$

with $\mathbf{A} = \mathbf{T}\text{diag}(\mathbf{u}) + \mathbf{F}\mathbf{\Phi}_1\mathbf{D}\text{diag}(\mathbf{g})$.

5.4.3 Optimizing precoding matrix and fixed phase shift matrix of RIS 1 and phase shift matrix of RIS 2

The optimal active beamforming at BS when $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are fixed is the MR precoding, which maximizes the SNR given as [Zheng, You e Zhang 2021, Eq. 23]

$$\mathbf{w}^{\star} = \frac{\mathbf{h}}{||\mathbf{h}||} = \frac{\mathbf{F}\mathbf{\Phi}_1\mathbf{D}\mathbf{\Phi}_2\mathbf{u} + \mathbf{F}\mathbf{\Phi}_1\mathbf{g} + \mathbf{T}\mathbf{\Phi}_2\mathbf{u}}{||\mathbf{F}\mathbf{\Phi}_1\mathbf{D}\mathbf{\Phi}_2\mathbf{u} + \mathbf{F}\mathbf{\Phi}_1\mathbf{g} + \mathbf{T}\mathbf{\Phi}_2\mathbf{u}||}. \tag{5.14}$$

5.5 Statistical CSI-based Optimization

In this section, we proposed to maximize the ER which can reduce the overhead caused by the channel estimation. So, since in the previous section, we showed that MR is the optimal precoding, by plugging (5.14) in (5.10), one can compute the ergodic achievable downlink rate by applying the Jensen's inequality

$$\bar{R} = \mathbb{E}\left[\log_2\left(1 + \frac{\rho}{|z|^2}\mathbf{h}^H\mathbf{h}\right)\right] \overset{(a)}{\leq} \log_2\left(1 + \frac{\rho}{\sigma^2}\mathbb{E}\left[\mathbf{h}^H\mathbf{h}\right]\right), \tag{5.15}$$

where (a) utilizes Jensen's inequality based on the fact that the function $f(x) = \log_2(1+x)$ is concave with respect to x.

Theorem 1. *The ergodic capacity can be upper bounded by*

$$\bar{R} \leq \log_2\left(1 + \frac{\rho}{\sigma^2}\left[M\Upsilon + \mathbf{\Lambda}^H\mathbf{\Lambda}\right]\right), \tag{5.16}$$

with

$$\Upsilon = \frac{N_1\beta_{\mathbf{g}}\beta_{\mathbf{F}}}{K_3+1} + \frac{N_2\beta_{\mathbf{T}}\beta_{\mathbf{u}}}{K_2+1} + \frac{N_1N_2\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}}{K_1+1} \tag{5.17}$$

$$\mathbf{\Lambda} = \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}K_1}{K_1+1}}\mathbf{F}\mathbf{\Phi}_1\mathbf{D}_{\text{LoS}}\mathbf{\Phi}_2\mathbf{u} + \sqrt{\frac{\beta_{\mathbf{T}}\beta_{\mathbf{u}}K_2}{K_2+1}}\mathbf{T}_{\text{LoS}}\mathbf{\Phi}_2\mathbf{u} + \sqrt{\frac{\beta_{\mathbf{g}}\beta_{\mathbf{F}}K_3}{K_3+1}}\mathbf{F}\mathbf{\Phi}_1\mathbf{g}_{\text{LoS}} \tag{5.18}$$

Proof. To derive an expression in closed-form to $\bar{R}$, we should firstly compute $\mathbb{E}\left[\mathbf{h}^H\mathbf{h}\right]$, to this end we written $\mathbf{h}$ as follows

$$\mathbf{h} = \underbrace{\mathbf{F}\boldsymbol{\Phi}_1\left(\sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{g}}}{K_3+1}}\mathbf{g}_{\mathrm{NLoS}} + \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{g}}K_3}{K_3+1}}\mathbf{g}_{\mathrm{LoS}}\right)}_{\mathbf{x}_1}$$
$$\underbrace{\left(\sqrt{\frac{\beta_{\mathbf{T}}\beta_{\mathbf{u}}}{K_2+1}}\mathbf{T}_{\mathrm{NLoS}} + \sqrt{\frac{\beta_{\mathbf{T}}\beta_{\mathbf{u}}K_2}{K_2+1}}\mathbf{T}_{\mathrm{LoS}}\right)\boldsymbol{\Phi}_2\mathbf{u}}_{\mathbf{x}_2}$$
$$\underbrace{\mathbf{F}\boldsymbol{\Phi}_1\left(\sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}}{K_1+1}}\mathbf{D}_{\mathrm{NLoS}} + \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}K_1}{K_1+1}}\mathbf{D}_{\mathrm{LoS}}\right)\boldsymbol{\Phi}_2\mathbf{u}}_{\mathbf{x}_3},$$

Since $\mathbf{D}_{\mathrm{N}LoS} = \left[\mathbf{d}^1_{\mathrm{N}LoS}, \ldots, \mathbf{d}^{N_2}_{\mathrm{N}LoS}\right]$, $\mathbf{T}_{\mathrm{N}LoS} = \left[\mathbf{t}^1_{\mathrm{N}LoS}, \ldots, \mathbf{t}^{N_2}_{\mathrm{N}LoS}\right]$ and that

$$\mathbf{d}^n_{\mathrm{N}LoS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_1}), \qquad 1 \leq n \leq N_2 \tag{5.19}$$
$$\mathbf{t}^n_{\mathrm{N}LoS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M), \qquad 1 \leq n \leq N_2 \tag{5.20}$$
$$\mathbf{g}_{\mathrm{N}LoS} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_1}), \tag{5.21}$$

then, once the linear transformation for a standard Gaussian random vector is still a Gaussian random vector [Muirhead e J 2009], by leveraging of [Muirhead e J 2009, Theorem 1.2.6] we obtain

$$\mathbf{x}_1 \sim \mathcal{CN}\left(\sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{g}}K_3}{K_3+1}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{g}_{\mathrm{LoS}}, \frac{\beta_{\mathbf{F}}\beta_{\mathbf{g}}}{K_3+1}\mathbf{F}\mathbf{F}^H\right) \tag{5.22}$$
$$\mathbf{x}_2 \sim \mathcal{CN}\left(\sqrt{\frac{\beta_{\mathbf{T}}\beta_{\mathbf{u}}K_2}{K_2+1}}\mathbf{T}_{\mathrm{LoS}}\boldsymbol{\Phi}_2\mathbf{u}, \frac{N_2\beta_{\mathbf{T}}\beta_{\mathbf{u}}}{K_2+1}\mathbf{I}_M\right) \tag{5.23}$$
$$\mathbf{x}_3 \sim \mathcal{CN}\left(\sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}K_1}{K_1+1}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{D}_{\mathrm{LoS}}\boldsymbol{\Phi}_2\mathbf{u}, \frac{N_2\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}}{K_1+1}\mathbf{F}\mathbf{F}^H\right), \tag{5.24}$$

thus, by applying [Muirhead e J 2009, Theorem 1.2.14], we obtain the following

$$\mathbf{h} \sim \mathcal{CN}\left(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}\right), \tag{5.25}$$

where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ are respectively given by

$$\boldsymbol{\Lambda} = \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{g}}K_3}{K_3+1}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{g}_{\mathrm{LoS}} + \sqrt{\frac{\beta_{\mathbf{T}}\beta_{\mathbf{u}}K_2}{K_2+1}}\mathbf{T}_{\mathrm{LoS}}\boldsymbol{\Phi}_2\mathbf{u} + \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}K_1}{K_1+1}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{D}_{\mathrm{LoS}}\boldsymbol{\Phi}_2\mathbf{u}, \tag{5.26}$$

$$\boldsymbol{\Sigma} = \frac{\beta_{\mathbf{F}}\beta_{\mathbf{g}}}{K_3+1}\mathbf{F}\mathbf{F}^H + \frac{N_2\beta_{\mathbf{T}}\beta_{\mathbf{u}}}{K_2+1}\mathbf{I}_M + \frac{N_2\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}}{K_1+1}\mathbf{F}\mathbf{F}^H, \tag{5.27}$$

where, each element of $\mathbf{h}$ has variance of $\Upsilon = [\mathbf{\Sigma}]_{m,m} = \frac{N_1\beta_{\mathbf{F}}\beta_{\mathbf{g}}}{K_3+1} + \frac{N_2\beta_{\mathbf{T}}\beta_{\mathbf{u}}}{K_2+1} + \frac{N_1N_2\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}}{K_1+1}$. Therefore, the product $\mathbf{h}^H\mathbf{h}$ has a complex non-central Wishart distribution with M degrees of freedom, covariance Υ and non-centrality parameter $\dfrac{\mathbf{\Lambda}^H\mathbf{\Lambda}}{\Upsilon}$, represented by [Muirhead e J 2009, Definition 10.3.1]

$$\mathbf{h}^H\mathbf{h} \sim \mathcal{W}\left(M, \Upsilon, \frac{\mathbf{\Lambda}^H\mathbf{\Lambda}}{\Upsilon}\right), \tag{5.28}$$

Hence, we have the expectation of a non-central Wishart can be given in closed-form expression by [Siriteanu et al. 2012, Eq. (7)],[Siriteanu et al. 2014, Eq. (45)]

$$\mathbb{E}\left[\mathbf{h}^H\mathbf{h}\right] = M\Upsilon + \mathbf{\Lambda}^H\mathbf{\Lambda}. \tag{5.29}$$

This completes the proof. ■

We should notice that in this case by formulating the optimization problem intending to maximize the ergodic rate, just the passive beamforming at the RISs is considered since the active beamforming at BS has been considered in order to compute the ergodic rate, hence the optimization problem now is formulated as

$$\max_{\mathbf{\Phi}_1,\mathbf{\Phi}_2} \ \log_2\left(1+\rho\left[M\Upsilon + \mathbf{\Lambda}^H\mathbf{\Lambda}\right]\right) \tag{5.30}$$

$$\text{s.t. } |e^{j\theta_n^{\mu}}| = 1, \quad \forall n = 1,\ldots,N_\mu, \ \ \mu \in \{1,2\}. \tag{5.30a}$$

5.5.1 Optimizing phase-shift matrix of RIS 1 with fixed phase-shift matrix of RIS 2

The derivative of $\bar{R}$ with respect to the angles of RIS 1 can be given as

$$\frac{\partial\bar{R}(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1} = \frac{\frac{\rho}{\sigma^2}\left(2\mathbf{C}^H(\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)\right)}{\ln 2\left(1+\frac{\rho}{\sigma^2}(M\Upsilon + ||\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1||^2)\right)}, \tag{5.31}$$

with $\mathbf{b} = \sqrt{\frac{\beta_{\mathbf{u}}\beta_{\mathbf{T}}K_2}{K_2+1}}\mathbf{T}_{\text{LoS}}\mathbf{\Phi}_2\mathbf{u}$ and $\mathbf{C} = \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{g}}K_3}{K_3+1}}\mathbf{F}_{\text{LoS}}\text{diag}(\mathbf{g}_{\text{LoS}}) + \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}K_1}{K_1+1}}\mathbf{Q}$, where the matrix $\mathbf{Q} = \mathbf{F}_{\text{LoS}}\text{diag}\left(\mathbf{D}_{\text{LoS}}\mathbf{\Phi_2}\mathbf{u}\right)$

Proof. Notice that (5.16) can be written as

$$\bar{R} \leq \log_2\left(1+\frac{\rho}{\sigma^2}\left[M\Upsilon + (\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)^H(\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)\right]\right), \tag{5.32}$$

by taking the first order derivative of (5.32) with respect to $\boldsymbol{\theta}_1$ we obtain

$$\frac{\partial \bar{R}(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = \frac{\frac{\rho}{\sigma^2}\frac{\partial (\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)^H(\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}}{\ln 2\left(1+\frac{\rho}{\sigma^2}\left[M\Upsilon+(\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)^H(\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)\right]\right)}, \tag{5.33}$$

by utilizing [Petersen, Pedersen et al. 2008, Eq. (84)], the numerator of (5.33) can be evaluated as follows

$$\frac{\partial (\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)^H(\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = 2\mathbf{C}^H\left(\mathbf{b}+\mathbf{C}\boldsymbol{\theta}_1\right), \tag{5.34}$$

thus the proof is completed. ■

5.5.2 Optimizing phase-shift matrix of RIS 2 with fixed phase-shift matrix of RIS 1

The derivative of $\bar{R}$ with respect to the phase-shift of RIS 2 can be given as

$$\frac{\partial \bar{R}(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} = \frac{\frac{\rho}{\sigma^2}\left(2\mathbf{B}^H(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)\right)}{\ln 2\left(1+\frac{\rho}{\sigma^2}(M\boldsymbol{\Upsilon}+||\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2||^2)\right)}, \tag{5.35}$$

where $\mathbf{a}$ and $\mathbf{B}$ are given by

$$\mathbf{a} = \sqrt{\frac{\beta_{\mathbf{g}}\beta_{\mathbf{F}}K_3}{K_3+1}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{g}_{\text{LoS}}, \quad \mathbf{B} = \sqrt{\frac{\beta_{\mathbf{T}}\beta_{\mathbf{u}}K_2}{K_2+1}}\mathbf{T}_{\text{LoS}}\text{diag}(\mathbf{u}) + \sqrt{\frac{\beta_{\mathbf{F}}\beta_{\mathbf{D}}\beta_{\mathbf{u}}K_1}{K_1+1}}\mathbf{F}\boldsymbol{\Phi}_1\mathbf{D}_{\text{LoS}}\text{diag}(\mathbf{u}) \tag{5.36}$$

Proof. Notice that (5.16) can be written as

$$\bar{R} \leq \log_2\left(1+\frac{\rho}{\sigma^2}\left[M\Upsilon+(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)^H(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)\right]\right), \tag{5.37}$$

by taking the first order derivative of (5.37) with respect to $\boldsymbol{\theta}_2$ we obtain

$$\frac{\partial \bar{R}(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} = \frac{\frac{\rho}{\sigma^2}\frac{\partial (\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)^H(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2}}{\ln 2\left(1+\frac{\rho}{\sigma^2}\left[M\Upsilon+(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)^H(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)\right]\right)}, \tag{5.38}$$

by utilizing [Petersen, Pedersen et al. 2008, Eq. (84)], the numerator of (5.38) can be evaluated as follows

$$\frac{\partial (\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)^H(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} = 2\mathbf{B}^H\left(\mathbf{a}+\mathbf{B}\boldsymbol{\theta}_2\right), \tag{5.39}$$

thus the proof is completed. ■

Algorithm 1 Complete Solution for Optimization of $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$

Initialize $\boldsymbol{\Phi}_1 = \text{diag}(\boldsymbol{\theta}_1)$ and $\boldsymbol{\Phi}_2 = \text{diag}(\boldsymbol{\theta}_2)$ with random feasible values; $f^{(n)} = f_1^{(n)} = f_2^{(n)} = 0$; $\epsilon = 10^{-3}$
repeat
  Step 1: $\boldsymbol{\theta}_1$ *optimization with fixed* $\boldsymbol{\theta}_2$
  repeat
    $f_1^{(n)} = \bar{R}(\boldsymbol{\theta}_1^{(n)}, \boldsymbol{\theta}_2)$ in Eq (5.16);
    $\boldsymbol{\theta}_1^{(n+1)} = \exp\left(j\angle \arg_\nu \max \bar{R}\left(\boldsymbol{\theta}_1^{(n)} + \nu \frac{\partial \bar{R}(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}\right)\right)$;
    $f_1^{(n+1)} = \bar{R}(\boldsymbol{\theta}_1^{(n+1)}, \boldsymbol{\theta}_2)$ in Eq (5.16);
    $n = n + 1$;
  until $|f_1^{(n)} - f_1^{(n+1)}| \leq \epsilon$
  $F_1 = f_1^{(n+1)}$
  Step 2: $\boldsymbol{\theta}_2$ *optimization with fixed* $\boldsymbol{\theta}_1$
  repeat
    $f_2^{(n)} = \bar{R}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(n)})$ in Eq (5.16);
    $\boldsymbol{\theta}_2^{(n+1)} = \exp\left(j\angle \arg_\nu \max \bar{R}\left(\boldsymbol{\theta}_2^{(n)} + \nu \frac{\partial \bar{R}(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2}\right)\right)$;
    $f_2^{(n+1)} = \bar{R}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(n+1)})$ in Eq (5.16);
    $n = n + 1$;
  until $|f_2^{(n)} - f_2^{(n+1)}| \leq \epsilon$
  $F_2 = f_2^{(n+1)}$
until $|F_1 - F_2| \leq \epsilon$
Outputs: $\boldsymbol{\Phi}_1^{(n+1)}$ and $\boldsymbol{\Phi}_2^{(n+1)}$.

5.6 Simulation Results

In this section, we aim to investigate the performance of the single-user double-RIS-aided mMIMO system. The single antenna user is positioned in a (60; 0; 0) m, while the RIS 1 and RIS 2 are located at (10; 20; 30) and (50; 5; 30) respectively the BS is located at (0; 0; 20). All presented results have been averaged over 1000 realizations. Table 2 summarizes the adopted values for the main simulation parameters.

Table 2 – Adopted Simulation Parameters.

Parameter	**Value**
Double-RIS-Aided Massive MIMO System	
Tx. to Noise power ratio	$\frac{\rho}{\sigma^2} \in \{0; 80\}$ [dB]
Number of antennas at BS	$M = 64$
Number of reflecting meta-surfaces	$N = 50; N_1 = N_2 = N/2$
Path-loss models	$\beta_{\mathbf{F}} = \beta_{\mathbf{u}} = \beta_{\mathbf{D}} = \frac{1}{d^2}$ $\beta_{\mathbf{T}} = \beta_{\mathbf{g}} = \frac{1}{d^{3.5}}$
Rician coefficient	$K_1 = 15$, $K_2 = K_3 = 7.5$
Monte-Carlo Simulation	$\mathcal{T} = 1000$ realizations

Fig. 16 depicts the Capacity of a single-user system when the communication

is assisted by two RIS. Here, we apply two optimization techniques aforementioned, instantaneous CSI optimization and statistical CSI optimization. We firstly should notice that the performance of the single RIS system is outperformed by the double-RIS system, mainly when the $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are optimized in the instantaneous/statistical CSI mode. Furthermore, we can see that in mean, the statistical CSI can be very tight when compared with the instantaneous CSI, indicating the promising potential of this proposed optimization technique, since the optimization of $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are not necessary to be realized in each time coherence.

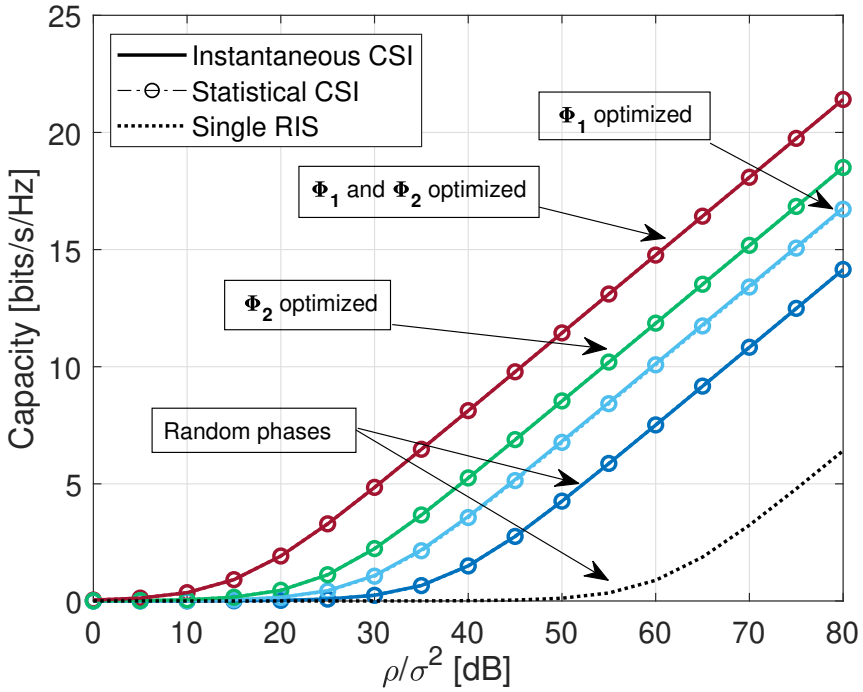


Figure 16 – The instantaneous CSI *vs.* the statistical CSI-based rate for four different cases: a) random phase-shift matrices; b) only $\mathbf{\Phi}_1$ optimization; c) only $\mathbf{\Phi}_2$ optimization and d) $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ optimization.

5.7 Future Directions

In the following, we intend further to explore the investigated scenario. Starting by changing the Rician factors (K_1, K_2, and K_3) we will investigate to understand which link can better contribute to the system when it is with the strongest LoS. Furthermore, the RISs' positions also should be investigated, where we will seek to understand how they can impact the performance of the system. Besides, we will explore the number of users in the cell aiming to comprehensively study a more realistic scenario, where a new technique of precoding should be explored since initially, we consider the MR precoding which is optimal in the single-user scenario but in the multi-user scenario can not operate with

desirable performance. In addition, the MR precoding at BS necessarily needs the CSI knowledge to be computed.

6 Conclusions

6.1 General Conclusions

In this Master's Dissertation, we study the physical characteristics, deployment, and operational principle of a new promising and potentially enabling technology called Intelligent Reconfigurable Surfaces. The findings of this study are presented by means of three scientific publications, each addressing different aspects and scenarios to provide comprehensive insights into the benefits and limitations of RIS-aided communication. The first scenario/work focuses on enhancing cooperative communication between two users by deploying RIS. We examine how RIS implementation can improve the performance and efficiency of such cooperative systems. In the following, in the second work, we study the EE problem for RIS-based systems, where an alternative strategy is utilized to deal with the difficulties imposed by the RIS. Furthermore, in the last work, we investigate the impact of deploying two RIS on system performance. We analyze the benefits and trade-offs associated with utilizing multiple RIS in different scenarios. Through these investigations, we aim to contribute to the understanding and advancement of RIS technology, paving the way for its practical implementation and deployment in various B5G communication scenarios. The main conclusions of these three works are:

- The RIS holds great potential in delivering substantial performance gains for communication systems. When properly deployed, it can effectively reduce power consumption at the BS, leading to enhanced energy efficiency. Moreover, it enables higher data rates for users located at the cell edge, and consequently better system fairness. The RIS also plays a significant role in enhancing the reliability of user information.

- The RIS also exhibits remarkable versatility as it can be interestingly integrated with other key technologies such as massive MIMO, NOMA, and others. This integration boosts the achievable capacity of communication systems, with an additional cost associated with its deployment and configuration.

- To achieve promising gains, the RIS is highly dependent on phase-shift configuration, which in turn is highly dependent on accurate channel estimation. However, obtaining accurate channel estimation procedures can be challenging in practice, particularly due to the large number of RIS elements. This introduces a significant overhead in the channel estimation process, which can pose difficulties, especially in scenarios with short channel coherence times.

6.2 Specific Chapter Conclusions

Here, we delve deeper into the key findings and specific conclusions of each research work conducted as part of this Master's Dissertation. These findings shed light on the important insights gained in each specific RIS-aided application scenario, providing valuable contributions to this research field.

6.2.1 Chapter 2 Conclusions

In the work discussed in Chapter 2, and shown in Appendix A, we analyze the RIS-aided cooperative FD-SWIPT-NOMA system. The conclusions and main findings are outlined below.

1. For low-middle values of power, the harvested power can be precisely approximated as a Gamma random variable through the moment matching technique. This statement allows us to derive analytical OP/ER expressions that are simple yet accurate for a wide range of values of several parameters of the system.

2. We show that the RIS can effectively improve the throughput of the poor channel condition device when the good channel condition device cooperates. The number of elements of RIS has a primordial impact on this configuration once the RIS potentialities the good channel condition device to harvest power, providing better conditions for the poor channel condition device.

3. We show that the number of elements of RIS does not impact the maximum data rate that the poor channel condition device can achieve, however, it can be achieved with lower power when the number of elements increases.

4. We show how it is vital to mitigate the residual self-interference value, in order to not harm the good channel condition device user performance. Furthermore, we show if the residual interference value is very high, it can influence the poor channel condition device performance, since the cooperative link can be useless in this situation.

5. The numerical/analytical results demonstrate that although the RIS can not mitigate the residual self-interference, it can potentialize the edge-cell performance till when both users are physically distant.

6.2.2 Chapter 3 Conclusions

In the work discussed in Chapter 3, available in Appendix B, we investigate the performance of EE in the RIS-aided mMIMO scenarios. The main conclusions and findings can be summarized as follows

- The statistical CSI optimization can lead to a very promising performance gain with lower complexity when compared with the instantaneous CSI optimization.
- The RIS has proven to be very promising as it can significantly improve the user's conditions even with the statistical CSI optimization, leading to higher EE.
- The proposed algorithm can provide higher EE when all optimization variables are optimized, since for this scenario the algorithm can manage efficiently the resources, providing lower power consumption, and few active antennas at BS.
- Our approach incurs a performance loss of $\approx 5.3\%$, compared to the instantaneous CSI-based optimization method. Besides, our proposed algorithm can achieve better EE gains with lower complexity than the Gradient Ascent method.

6.2.3 Chapter 4 Conclusions

In the work discussed in Chapter 4, and reproduced in Appendix C, we investigate the performance of a double-RIS setup. The main conclusions and findings can be summarized as follows

1. We show that the cascaded channel between the conventional and STAR RIS can be represented by a Gamma random variable by utilizing the moment matching technique.
2. We show that the double-RIS setup can outperform the single RIS setup achieving higher rates, mainly when the double-reflection BR-RIS-RIS-Users link has better conditions than the single-reflection link between the BS-RIS-Users. Besides, when the conventional RIS has a random phase-shift set, the double-RIS still can outperform the single-RIS setup.
3. We show how the placement of the RISs in the performance of the system is important. Furthermore, the size of each RIS has a fundamental impact on the system.

6.2.4 Chapter 5 Conclusions

In the work discussed in Chapter 5, we investigate the performance of a double-RIS setup with mMIMO. The partial conclusions can be listed as follows

1. The double-RIS system with mMIMO seems to be promising. Since channel estimation is a latent problem, the statistical CSI optimization should be better investigated.
2. The proposed algorithm for a single-user scenario presented promising results and should be extended for the most robust scenarios.

6.3 Future Research Directions

Here, the author emphasizes certain avenues for future research associated with this Master's Dissertation, that will be explored and further investigated in forthcoming studies:

- **Channel Estimation Overhead:** Although RIS can bring interesting potentials, in practice, the channel estimation overhead can be a crucial limiter. Since that RIS is, in general, composed of a huge number of passive/active elements, accurate channel estimation can be hard to obtain. Furthermore, the majority of works in the literature, consider perfect CSI estimation. Therefore the combination of these factors highlights the necessity of proposing new ways to configure and better deal with RIS-aided communication systems.

- **Self-Sustainable RIS:** The current deployment of RIS relies on an external power source, which is crucial for its operation. However, in order to achieve sustainable systems, self-sustainable RIS emerges as an up-and-coming solution. The concept of self-sustainable RIS is closely linked to Wireless Power Transfer (WPT), which has gained significant attention for its ability to extend the battery life of energy-constrained devices through electromagnetic waves. The fundamental idea behind self-sustainable RIS is to utilize the RIS not only in reflection mode but also in absorption mode, enabling it to harvest energy and sustain itself.

- **Near Field Communications:** Since millimeter waves (mmWave) technology is a potential candidate for operating in future scenarios due to its ability to utilize better the limited bandwidth resource, future array panels are expected to consist of a large number of antenna/metamaterial elements, resulting in larger panels in terms of width and length. This is an important consideration as it can significantly alter the communication scenario, transitioning from the conventional far-field to the near-field. In the far field, the impinging wave is assumed to be planar, while in the near field, it is spherical in nature. For this reason, the operation of RIS in these scenarios is crucial to be understood and new ways for RIS focusing should be proposed.

Bibliography

Alwis et al. 2021 ALWIS, C. D. et al. Survey on 6G Frontiers: Trends, Applications, Requirements, Technologies and Future Research. *IEEE Open Journal of the Communications Society*, v. 2, p. 836–886, 2021.

Björnson, Hoydis e Sanguinetti 2017 BJöRNSON, E.; HOYDIS, J.; SANGUINETTI, L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. *Foundations and Trends® in Signal Processing*, v. 11, n. 3-4, p. 154–655, 2017. ISSN 1932-8346.

Björnson, Ozdogan e Larsson 2020 BJöRNSON, E.; OZDOGAN, O.; LARSSON, E. G. Intelligent Reflecting Surface Versus Decode-and-Forward: How Large Surfaces are Needed to Beat Relaying? *IEEE Wireless Communications Letters*, v. 9, n. 2, p. 244–248, 2020.

Björnson e Sanguinetti 2020 BJöRNSON, E.; SANGUINETTI, L. Power Scaling Laws and Near-Field Behaviors of Massive MIMO and Intelligent Reflecting Surfaces. *IEEE Open Journal of the Communications Society*, v. 1, p. 1306–1324, 2020.

Boyd, Boyd e Vandenberghe 2004 BOYD, S.; BOYD, S. P.; VANDENBERGHE, L. *Convex optimization*. [S.l.]: Cambridge university press, 2004.

Carvalho et al. 2020 CARVALHO, E. D. et al. Non-Stationarities in Extra-Large-Scale Massive MIMO. *IEEE Wireless Communications*, v. 27, n. 4, p. 74–80, 2020.

Carvalho et al. 2017 CARVALHO, E. D. et al. Random Access Protocols for Massive MIMO. *IEEE Communications Magazine*, v. 55, n. 5, p. 216–222, 2017.

Cui et al. 2022 CUI, M. et al. Near-field Communications for 6G: Fundamentals, Challenges, Potentials, and Future Directions. *IEEE Communications Magazine*, IEEE, 2022.

Ding et al. 2022 DING, Z. et al. A State-of-the-Art Survey on Reconfigurable Intelligent Surface-Assisted Non-Orthogonal Multiple Access Networks. *Proceedings of the IEEE*, v. 110, n. 9, p. 1358–1379, 2022.

Dinkelbach 1967 DINKELBACH, W. On Nonlinear Fractional Programming. *Management science*, INFORMS, v. 13, n. 7, p. 492–498, 1967.

Goldsmith 2005 GOLDSMITH, A. *Wireless Communications*. [S.l.]: Cambridge University Press, 2005.

Gong et al. 2020 GONG, S. et al. Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey. *IEEE Communications Surveys and Tutorials*, v. 22, n. 4, p. 2283–2314, 2020.

Guo et al. 2021 GUO, F. et al. Enabling Massive IoT Toward 6G: A Comprehensive Survey. *IEEE Internet of Things Journal*, v. 8, n. 15, p. 11891–11915, 2021.

Guo et al. 2020 GUO, H. et al. Weighted Sum-Rate Maximization for Reconfigurable Intelligent Surface Aided Wireless Networks. *IEEE Trans. Wireless Commun.*, v. 19, n. 5, p. 3064–3076, 2020. ISSN 1558-2248.

Huang et al. 2019 HUANG, T. et al. A Survey on Green 6G Network: Architecture and Technologies. *IEEE Access*, v. 7, p. 175758–175768, 2019.

Jiang et al. 2021 JIANG, W. et al. The Road Towards 6G: A Comprehensive Survey. *IEEE Open Journal of the Communications Society*, v. 2, p. 334–366, 2021.

Jong 2012 JONG, Y. An Efficient Global Optimization Algorithm for Nonlinear Sum-of-Ratios Problem. In: . [S.l.: s.n.], 2012.

Kang, You e Zhang 2022 KANG, Z.; YOU, C.; ZHANG, R. *Active-Passive IRS aided Wireless Communication: New Hybrid Architecture and Elements Allocation Optimization.* [S.l.]: arXiv, 2022.

Kobayashi, Mark e Turin 2011 KOBAYASHI, H.; MARK, B. L.; TURIN, W. *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance.* [S.l.]: Cambridge University Press, 2011.

Le et al. 2019 LE, M. T. P. et al. What is the Benefit of Code-domain NOMA in Massive MIMO? In: *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC).* [S.l.: s.n.], 2019. p. 1–5.

Li et al. 2018 LI, H. et al. Joint Antenna Selection and Power Allocation for an Energy-Efficient Massive MIMO System. *IEEE Wireless Communications Letters*, IEEE, v. 8, n. 1, p. 257–260, 2018.

Liu et al. 2016 LIU, Y. et al. Cooperative Non-orthogonal Multiple Access With Simultaneous Wireless Information and Power Transfer. *IEEE Journal on Selected Areas in Communications*, v. 34, n. 4, p. 938–953, 2016.

Liu et al. 2021 LIU, Y. et al. Reconfigurable Intelligent Surfaces: Principles and Opportunities. *IEEE Communications Surveys and Tutorials*, v. 23, n. 3, p. 1546–1577, 2021.

Liu et al. 2021 LIU, Y. et al. STAR: Simultaneous Transmission and Reflection for 360° Coverage by Intelligent Surfaces. *IEEE Wireless Communications*, v. 28, n. 6, p. 102–109, 2021.

Long et al. 2021 LONG, R. et al. Active Reconfigurable Intelligent Surface-Aided Wireless Communications. *IEEE Transactions on Wireless Communications*, v. 20, n. 8, p. 4962–4975, 2021.

Marinello et al. 2020 MARINELLO, J. C. et al. Antenna Selection for Improving Energy Efficiency in XL-MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 69, n. 11, p. 13305–13318, 2020.

Marzetta et al. 2016 MARZETTA, T. et al. *Fundamentals of Massive MIMO.* [S.l.]: Cambridge University Press, 2016. ISBN 9781316813089.

Miridakis, Tsiftsis e Yao 2023 MIRIDAKIS, N. I.; TSIFTSIS, T. A.; YAO, R. Zero Forcing Uplink Detection Through Large-Scale RIS: System Performance and Phase Shift Design. *IEEE Transactions on Communications*, v. 71, n. 1, p. 569–579, 2023.

Muirhead e J 2009 MUIRHEAD; J, R. *Aspects of multivariate statistical theory*. [S.l.]: John Wiley & Sons, 2009.

Pan et al. 2021 PAN, C. et al. Reconfigurable Intelligent Surfaces for 6G Systems: Principles, Applications, and Research Directions. *IEEE Communications Magazine*, v. 59, n. 6, p. 14–20, 2021.

Petersen, Pedersen et al. 2008 PETERSEN, K. B.; PEDERSEN, M. S. et al. The Matrix Cookbook. *Technical University of Denmark*, v. 7, n. 15, p. 510, 2008.

Ren et al. 2022 REN, J. et al. RIS-Assisted Cooperative NOMA with SWIPT. *IEEE Wireless Communications Letters*, p. 1–1, 2022.

Shaheen e Soleymani 2022 SHAHEEN, E. M.; SOLEYMANI, M. R. Performance Analyses of SWIPT-NOMA Enabled IoT Relay Networks. In: *2022 International Symposium on Networks, Computers and Communications (ISNCC)*. [S.l.: s.n.], 2022. p. 1–6.

Shen e Yu 2018 SHEN, K.; YU, W. Fractional programming for communication systems—Part I: Power control and beamforming. *IEEE Transactions on Signal Processing*, IEEE, v. 66, n. 10, p. 2616–2630, 2018.

Shen e Yu 2018 SHEN, K.; YU, W. Fractional Programming for Communication Systems—Part II: Uplink Scheduling via Matching. *IEEE Transactions on Signal Processing*, v. 66, n. 10, p. 2631–2644, 2018.

Siriteanu et al. 2014 SIRITEANU, C. et al. Schur Complement Based Analysis of MIMO Zero-Forcing for Rician Fading. *IEEE Transactions on Wireless Communications*, IEEE, v. 14, n. 4, p. 1757–1771, 2014.

Siriteanu et al. 2012 SIRITEANU, C. et al. MIMO Zero-Forcing Detection Analysis for Correlated and Estimated Rician Fading. *IEEE transactions on Vehicular Technology*, IEEE, v. 61, n. 7, p. 3087–3099, 2012.

Taha, Alrabeiah e Alkhateeb 2021 TAHA, A.; ALRABEIAH, M.; ALKHATEEB, A. Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning. *IEEE Access*, v. 9, p. 44304–44321, 2021.

Tahir, Schwarz e Rupp 2020 TAHIR, B.; SCHWARZ, S.; RUPP, M. Analysis of uplink IRS-assisted NOMA Under Nakagami-m Fading via Moments Matching. *IEEE Wireless Communications Letters*, IEEE, v. 10, n. 3, p. 624–628, 2020.

Tahir, Schwarz e Rupp 2021 TAHIR, B.; SCHWARZ, S.; RUPP, M. Outage Analysis of Uplink IRS-assisted NOMA Under Elements Splitting. In: IEEE. *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. [S.l.], 2021. p. 1–5.

Tang et al. 2020 TANG, J. et al. Decoupling or learning: Joint Power Splitting and Allocation in MC-NOMA with SWIPT. *IEEE Transactions on Communications*, IEEE, v. 68, n. 9, p. 5834–5848, 2020.

Vaezi et al. 2019 VAEZI, M. et al. Non-Orthogonal Multiple Access: Common Myths and Critical Questions. *IEEE Wireless Communications*, v. 26, n. 5, p. 174–180, 2019.

Wu e Zhang 2020 WU, Q.; ZHANG, R. Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network. *IEEE Communications Magazine*, v. 58, n. 1, p. 106–112, 2020.

Wu e Dai 2023 WU, Z.; DAI, L. Location Division Multiple Access for Near-Field Communications. *arXiv preprint arXiv:2301.09082*, 2023.

Xie et al. 2022 XIE, Z. et al. STAR-RIS aided NOMA in multicell networks: A General Analytical Framework With Gamma Distributed Channel Modeling. *IEEE Transactions on Communications*, IEEE, v. 70, n. 8, p. 5629–5644, 2022.

Xu et al. 2021 XU, J. et al. STAR-RISs: Simultaneous Transmitting and Reflecting Reconfigurable Intelligent Surfaces. *IEEE Communications Letters*, v. 25, n. 9, p. 3134–3138, 2021.

Zeng et al. 2020 ZENG, M. et al. Cooperative NOMA: State of the Art, Key Techniques, and Open Challenges. *IEEE Network*, v. 34, n. 5, p. 205–211, 2020.

Zeng et al. 2020 ZENG, M. et al. Cooperative NOMA: State of the Art, Key Techniques, and Open Challenges. *IEEE Network*, IEEE, v. 34, n. 5, p. 205–211, 2020.

Zhang et al. 2022 ZHANG, G. et al. Hybrid Time-Switching and Power-Splitting EH Relaying for RIS-NOMA Downlink. *IEEE Transactions on Cognitive Communications and Networking*, p. 1–1, 2022.

Zheng, You e Zhang 2021 ZHENG, B.; YOU, C.; ZHANG, R. Double-IRS Assisted Multi-User MIMO: Cooperative Passive Beamforming Design. *IEEE Transactions on Wireless Communications*, v. 20, n. 7, p. 4513–4526, 2021.

Zhi et al. 2022 ZHI, K. et al. Ergodic Rate Analysis of Reconfigurable Intelligent Surface-Aided Massive MIMO Systems With ZF Detectors. *IEEE Communications Letters*, v. 26, n. 2, p. 264–268, 2022.

Özyurt et al. 2022 ÖZYURT, S. et al. A Survey on Multiuser SWIPT Communications for 5G+. *IEEE Access*, v. 10, p. 109814–109849, 2022.

Şahin e Arslan 2020 ŞAHIN, M. M.; ARSLAN, H. Waveform-Domain NOMA: The Future of Multiple Access. In: *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. [S.l.: s.n.], 2020. p. 1–6.

Appendix

APPENDIX A – RIS-aided Cooperative FD-SWIPT-NOMA Outage Performance in Nakagami-m Channels

RIS-aided Cooperative FD-SWIPT-NOMA Outage Performance in Nakagami-m Channels. Full paper. *Submitted to IEEE Transactions on Cognitive Communications and Networking on Jul. 2023* (IF: 8.6, Eng. IV Qualis-CAPES: A1)

RIS-aided Cooperative FD-SWIPT-NOMA Performance over Nakagami-m Channels

Wilson de Souza Junior and Taufik Abrão

***Abstract*—In this work, we investigate reconfigurable intelligent surfaces (RIS)-assisted cooperative non-orthogonal multiple access (C-NOMA) consisting of two paired devices. The device with better channel conditions (D_1) is designated to act as a full-duplex (FD) relay to assist the device with poor channel conditions (D_2). D_1 does not use its battery energy to cooperate but harvests energy using simultaneous wireless information power transfer (SWIPT). A practical non-linear energy harvesting model is considered. We assume that all devices' links undergo Nakagami-m channel fading. We first approximate the harvested power as Gamma random variables via the moment matching (MM) technique. This allows us to derive analytical Outage Probability (OP) Ergodic Rate (ER) expressions that are simple to compute yet accurate for a wide range of RIS passive elements configurations, energy harvesting (EH) coefficients, and residual self-interference (SI) levels, being extensively validated by numerical simulations. The OP expressions reveal how important it is to mitigate the SI in the FD relay mode since, for reasonable values of residual SI coefficient, it is notable its detrimental effect on the system performance. Also, numerical results reveal that increasing the number of RIS elements can benefit the cooperative system much more than the non-cooperative one.**

***Index Terms*—RIS, NOMA, Outage Probability, Ergodic Rate, Cooperative, SWIPT, Nakagami-m, Self-Interference, Energy Harvest.**

I. Introduction

The necessity for providing new technologies in order to fulfill the demands of the next generations of communication systems is extremely vital since numerous highly important services are expected to be present in the fifth-generation (5G) communication systems and the beyond 5G (B5G) communications systems. Among them, massive machine-type communication (mMTC) and ultra-reliable low latency communications (URRLC) are use case scenarios widely targeted [1]. Since wireless networks have become denser, one of the challenges is to support heavy traffic [2]. In recent studies in the literature, it has been proven that *non-orthogonal multiple access* (NOMA) can be superior to the multiple access schemes as deployed in the past generations of cellular networks [3], including frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA), which possibly may not be able to scale to meet such new use case demands. On the other hand, NOMA can be regarded as an exciting candidate to overcome such challenges. The NOMA technology, allows multiple devices to transmit in the same resource block (RB) leading to higher spectral efficiency (SE) and energy efficiency (EE), than the usual *orthogonal multiple access* (OMA). Adopting NOMA, the successive interference cancellation (SIC) is required to distinguish the respective device's signal, being essential for NOMA to work suitably [4].

In cellular systems, the individual performance of each device is often affected by its specific geographical position; hence, to enhance fairness, the device with poor channel conditions device needs to be allocated a large number of resources, which can harm the QoS of the good channel conditions. Nonetheless, according to the NOMA protocol, the information of the device with poor channel conditions is previously known by the device with good channel conditions [5]. Therefore, leveraging this fact, the deployment of cooperative communication with the device with good channel conditions acting as a relay can be a potential solution for avoiding this fairness issue. Such a scenario is highly significant since it characterizes device-to-device (D2D) communication, which is an important technology for the next generations of communication systems [6], [7]. Specifically, cooperative communications can be performed under two distinct modes, named half-duplex (HD) and full-duplex (FD). The HD mode is known for sub-dividing the transmission time block, on which can degrade the SE, in the meanwhile the FD mode can be performed simultaneously at the cost of *self-interference* (SI) aggregation. The authors in [8], [9] concluded that by employing the device with good channel conditions as a relay, the cooperative NOMA system can enhance significantly the system performance of the system, in addition, [10]–[12] also show that the outage performance of cooperative NOMA can outperform the non-cooperative NOMA counterpart, even with imperfect channel state information (CSI) knowledge and imperfect SIC.

Although attractive performance gains could be achieved with cooperative communications, the deployment of such technology can be detrimental to the battery lifetime of the device that acts can a relay since inevitably it drains its battery energy. To overcome this issue, *simultaneous wireless information and power transfer* (SWIPT) is a promising and sustainable technology. Moreover, the SWIPT technique allows the user-relay to harvest energy from the ambient radio-frequency (RF) sources and reuse it to cooperate with the device with poor channel conditions. Precisely, SWIPT also can be implemented in two ways through its protocols, named, *power-splitting* (PS) and *time-switching* (TS) [13]. The PS-SWIPT protocol splits the received signal power into two parts,

This work was supported in part by the CAPES (Financial Code 001) and National Council for Scientific and Technological Development (CNPq) of Brazil under Grant 310681/2019-7.

W. Junior and T. Abrão are with the Department of Electrical Engineering (DEEL). State University of Londrina (UEL). Po.Box 10.011, CEP:86057-970, Londrina, PR, Brazil. Email: wilsoonjr98@gmail.com; taufik@uel.br

the first one is dedicated for *information decoding* (ID), and the second one is for *energy harvesting* (EH) purposes. On the other hand, the TS-SWIPT protocol necessarily splits the time slot to perform the ID and EH processes separately. [14]–[17] investigate the system performance of the SWIPT-assisted C-NOMA system over FD/HD mode, carrying out extensive analyses on the OP/ER. Moreover, [18] proposes an alternative optimization-based algorithm to jointly optimize the power allocation, power splitting, receiver filter and transmit beamforming in a SWIPT-assisted C-NOMA system.

Recently, *reflecting intelligent surface* (RIS) has attracted remarkable research attention due to its capability of changing and customizing the wireless propagation environment supporting high throughput, and being useful in indoor/outdoor scenarios where dense obstacles arise. The RIS is composed of scattering elements, *i.e.*, artificial meta-material structures composed of adaptive composite material layers, which can reflect incident electromagnetic waves. Moreover, the RIS elements can be configured in order to increase the reflected signal level in a specific direction for a priority device. The RIS is calling attention mainly due to its features: the capacity to be sustainable, low-power consumption, enhancing communications metrics, facility of implementation/ installation, compatibility, and low cost. In [19], a two-device RIS-assisted non-C-NOMA system has been analyzed and the authors concluded that RIS-aided communication can substantially improve the system OP. In [20], the authors show that the RIS-aided system can outperform the relay-aided system. Furthermore, the performance of the RIS-aided non-C-NOMA system has been analyzed in [21] from the perspective of imperfect SIC effects, where the authors have shown the potential of RIS to counter this limitation. On the other hand, in [22], the authors assess the impact of RIS phase shift design on the OP, ER, and bit error rate (BER) through two phase-shift configurations: random phase, and coherent phase-shifting under Nakagami-m fading. Recently, [23] provided valuable analytical results on the OP for downlink RIS-assisted backscatter communications with NOMA. In [24], the RIS-NOMA system OP under hardware impairments has been analyzed analytically: an accurate closed-form for the OP was developed. Very recently, the end-to-end channel statistics for both weakest and strongest devices in a *non*-cooperative RIS-aided NOMA system under Nakagami-m was derived in [25]. In [26], a RIS-aided cooperative NOMA scheme was analyzed from the perspective of power consumption, has been shown that the RIS can potentially reduce the consumption of the transmitted power.

A. Motivation and Contributions

To further improve the throughput, reliability, and fairness of mobile devices, the integration of different technologies such as NOMA and the cooperative system with SWIPT and RIS is very promising; furthermore, there are few works in the literature dealing with RIS-aided C-NOMA-SWIPT systems [27]–[29]. A hybrid TS and PS EH relaying scheme for RIS-NOMA system with transmit antenna selection is proposed in [27], while in [29] the authors seek to minimize the transmit power at both the BS and at the user-cooperating relay. Differently, in [28], the authors proposed an algorithm to jointly optimize the beamforming and the power splitting coefficient. However, to the best of our knowledge, so far, no studies have provided a solid investigation on the OP/ER performance of RIS-aided C-NOMA-SWIPT communications assuming a non-linear energy harvesting model operating under generalized Nakagami-m fading channels.

Motivated by *improving the fairness between the devices, extending the battery lifetime of the user-relay, and providing better reliability and throughput to the devices*, in this work we aim to investigate the *potential benefits of combining RIS, NOMA, and cooperative communications with SWIPT and non-linear EH circuit for energy harvesting*. Herein, in order to suitably unveil the system OP and ER performance, we focus on a two-device scenario with perfect SIC and perfect CSI at the BS. In light of the above motivations and challenges, the main ***contributions*** of this work are fourfold and can be summarized as follows:

- We analyze a cooperative RIS-aided NOMA system scenario under Nakagami-m fading channels to capture the characteristics of both non-line-of-sight (NLoS) and line-of-sight (LoS) propagation, combined with the SWIPT technique adopting a non-linear energy harvesting model, where the device D_1 cooperates with the device D_2 without harming itself in terms of battery lifetime. Besides, we investigate the use of RIS, aiming to identify its benefits and drawbacks operating under the considered scenario.
- We derive novel general expressions for the analyzed system's OP and a tight upper bound for ER. Since such expression is parameterized w.r.t. the system and channel parameters, the effect of each parameter can be effectively scanned as a function of the number of RIS elements.
- The methodology is deployed to assess the RIS-aided C-FD-SWIPT-NOMA performance analytically; hence, we characterize the impact of system parameters, such as residual self-interference, number of reflective elements, maximum harvested power, and energy harvest coefficient on the performance of the communication system.
- Comprehensive numerical results simulations for the OP, ER, and device rate corroborating the effectiveness and accuracy of the proposed analytical performance expressions, as well as the provided insights of the studied cooperative system.

Notation: $\Gamma(\cdot)$ is the Gamma function; $\gamma(\cdot,\cdot)$ is the lower incomplete Gamma function; $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function; $X \sim \mathcal{CN}(\mu, \sigma^2)$ denotes a random variable X following a Complex Normal distribution with mean μ and variance σ^2; $X \sim \text{Gamma}(k, \theta)$ denotes a random variable following a Gamma distribution with shape parameter k and scale parameter δ; $X \sim \text{Exponential}(\lambda)$ denotes a random variable following a Exponential distribution with rate parameter λ; $X \sim \text{Rayleigh}(\sigma)$ denotes a random variable following a Rayleigh distribution with scale parameter σ; the magnitude of a complex number z is expressed by $|z|$; $\arg(\cdot)$ denotes the argument of a complex number; $\text{diag}(\cdot)$ denotes the diagonal operator; vectors and matrices are represented by bold-face

letters; F_X denotes the cumulative density function (CDF) of X; f_X denotes the probability density function (PDF); and $\Pr(\cdot)$ expresses probability.

II. System Model

Let us consider a RIS-aided C-FD-SWIPT-NOMA downlink where a source (S) equipped with a single antenna, serves simultaneously two devices. Let us denote the device with good channel conditions as D_1, while D_2 is the device with poor channel conditions as represented in Fig. 1. The transmission process is assisted by an N-elements RIS.

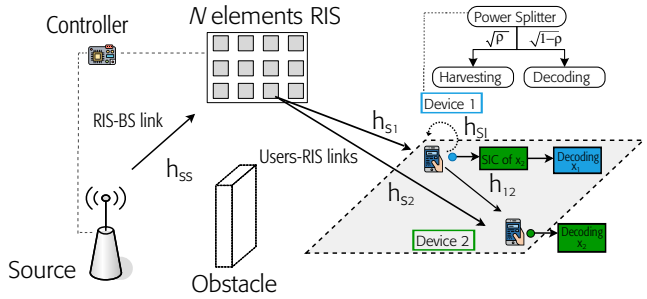


Fig. 1. Downlink RIS-aided C-FD-SWIPT-NOMA system model, consisting of a source (S) and two paired devices, D_1 and D_2. To improve the performance of D_2, the device D_1 acts as a full-duplex relay under the self-interference (SI) effect.

The system model depicted in Figure 1 assumes that the direct link between the source and the devices is fully blocked or obstructed. This system model is specifically designed for application scenarios related to IoT systems, taking into consideration environments such as offices and residences where D2D communication can be advantageous. Additionally, the proposed setup has significant potential for applications in industries, enabling communication and automation between industrial robots.

To ensure improved QoS to the system, D_1 can act as a cooperative relay using the decode-and-forward (DF) protocol [30], [31]. In order to achieve ultra-low latency in the cooperative framework, a crucial requirement in the 5G and B5G communication systems, we assume that the cooperative device (D_1) is equipped with two antennas. The first antenna is responsible for receiving the signal while the second antenna is used for relaying [32]. This enables the cooperative process to occur in FD mode [5], [16], [18], [33], [34], allowing simultaneous transmission and reception. However, this FD mode operation is subject to the self-interference (SI) or loop-interference (LI), which refers to the signal that is transmitted by the transmitter antenna and looped back to the receiver antenna at the same node [12].

To ensure the battery lifetime of D_1 is not compromised, D_1 can leverage the SWIPT technique by adopting the PS architecture [17], [35]. This architecture allows D_1 to divide the received power into two fractions, where a fraction ρ of the received power is allocated for energy harvesting and the remaining fraction $(1-\rho)$ is used for information decoding. By utilizing the harvested energy, D_1 can effectively relay the reconstructed message of D_2, thus enabling efficient utilization of available energy resources.

The operation of the RIS is encapsulated by its phase shift matrix, which represents the phase shifts applied to the incoming signals. The RIS is assumed to be an ideal passive and lossless surface. The phase shift matrix is given as

$$\mathbf{\Phi} = \text{diag}([\nu_1 e^{j\phi_1}, \nu_2 e^{j\phi_2}, \ldots, \nu_n e^{j\phi_N}]^T) \tag{1}$$

where $\phi_n \in (0, 2\pi]$ is the phase shift applied to the n-th element of the RIS and ν_n is the amplitude coefficient of the n-th element which is assumed to be one for all elements, *i.e.*, $\nu_n = 1$, $n = 1, \ldots, N$. We consider that the RIS is programmed by a dedicated controller connected to the S via a high-speed backhaul [36], whose main aim is to update systematically the RIS phases shift at each coherence-time since the phase-shift variables are directly related to the CSI, which is assumed to be perfectly known at the S.

We assume that all link channels experience quasi-static flat fading over a coherence time and vary independently from one coherence time to another. The small-scale fading channel from S to the RIS is denoted by $\mathbf{h}_{ss} \in \mathbb{C}^N$, while the small-scale fading channel between the RIS and the ℓ-th device is denoted by $\mathbf{h}_{s\ell} \in \mathbb{C}^N$ $\forall \ell \in \{1, 2\}$. It is assumed that $|[\mathbf{h}_{ss}]_n| = |h_{ss,n}|$ and $|[\mathbf{h}_{s\ell}]_n| = |h_{s\ell,n}|$ are independent random variables following Nakagami-m fading,

$$\begin{aligned} |h_{ss,n}| &\sim \text{Nakagami}(m_{ss}, 1), \quad \forall n \in \{1, \ldots, N\} \\ |h_{s\ell,n}| &\sim \text{Nakagami}(m_{s\ell}, 1), \quad \forall \ell \in \{1, 2\}, \end{aligned} \tag{2}$$

Accordingly, the cascaded channel between the S, RIS and D_ℓ can be written as

$$h_\ell = \sqrt{\beta_{ss}\beta_{s\ell}}\mathbf{h}_{ss}^H \mathbf{\Phi} \mathbf{h}_{s\ell}, \qquad \forall \ell \in \{1, 2\} \tag{3}$$

where β_{ss} and $\beta_{s\ell}$ are the path-loss of $S \rightarrow$ RIS and RIS $\rightarrow D_\ell$ links, respectively. In this study, our main objective is to improve the communication conditions for D2. However, we configure the RIS to maximize the received power at the D1. This enables D1 to benefit from better channel conditions, making it the strong device in the NOMA protocol, while D2 represents the weak device with poorer channel conditions. The phase shift matrix of the RIS for this particular configuration is given as follows

$$\phi_n = -\angle h_{ss,n} - \angle h_{s1,n}, \qquad \forall n = \{1, \ldots, N\}, \tag{4}$$

Here, with the phase-shift matrix defined as (4), the cascaded channel for D_1, $\mathbf{h}_{ss}\mathbf{\Phi}\mathbf{h}_{s1} = \sum_{n=1}^{N} |h_{ss,n}||h_{s1,n}|$, is a real-number[1] and $\mathbf{h}_{ss}\mathbf{\Phi}\mathbf{h}_{s2} = \sum_{n=1}^{N} e^{j\phi_n} h_{ss,n} h_{s2,n}$, a complex number since the phases of RIS appear random for D_2.

Besides, for the $D_1 \rightarrow D_2$ communication link (D2D), the h_{12} follows a complex Gaussian distribution, *i.e.*, $h_{12} \sim \mathcal{CN}(0, \beta_{12})$ $\forall n \in \{1, \ldots, N\}$, where β_{12} is the path loss of D2D communication.

[1]Sum of product between two Nakagami-m random variables, since the RIS is programmed to cancel the phase of $h_{ss,n}$ and $h_{s1,n}$

A. Signal Model

In NOMA, S transmits a superimposed signal $x(t)$ which propagates in direction to the devices through the RIS, with $x = \sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2$, where $\alpha_\ell \in (0,1)$ denotes the power allocation coefficients, with $\alpha_1 + \alpha_2 = 1$, $\alpha_2 > \alpha_1$ and x_ℓ, with $\mathbb{E}[|x_\ell|^2] = 1$, is the message of ℓ-th device, $\ell \in \{1,2\}$.

1) Device 1: The observation at the D_1 which will be designated for ID can be written as follows

$$y_1^{\text{ID}}(t) = \underbrace{h_1\sqrt{P_t(1-\rho)}x(t)}_{\substack{\text{Superimposed}\\ \text{information}}} + \underbrace{h_{SI}\sqrt{P_H}\hat{x}_2(t-\tau)}_{\text{Residual Self-interference}} + \underbrace{n_1(t)}_{\text{Noise}}, \quad (5)$$

where $0 \leq \rho \leq 1$ represents the fraction of received power utilized for energy harvesting (EH factor), P_t denotes the transmit power, and the SI term arises from the adoption of FD mode at the user-relay. In this work, we assume that the channel coefficient associated with the SI is not affected by fading due to the short distance. In addition, we consider imperfect self-interference cancellation, where h_{SI} denotes the residual self-interference and is modeled as $\mathcal{CN}(0,\omega)$ [14], [15], [37], [38]. Besides, $\hat{x}_2(t-\tau)$ is the rebuilt message of D_2, retransmitted by D_1, τ denotes the processing delay at the user-relay caused by FD mode, which is assumed lower than the coherence time. The additive white Gaussian noise (AWGN) at the D_1 is modeled as $n_1 \sim \mathcal{CN}(0,\sigma_1^2)$.

2) Device 2: The observation at the D_2 can be written as follows

$$y_2(t) = \underbrace{h_2\sqrt{P_t}x(t)}_{\substack{\text{Superimposed}\\ \text{information}}} + \underbrace{h_{12}\sqrt{P_H}\hat{x}_2(t-\tau)}_{\substack{\text{Cooperative}\\ \text{transmission}}} + \underbrace{n_2(t)}_{\text{noise}}, \quad (6)$$

where $n_2 \sim \mathcal{CN}(0,\sigma_2^2)$. is the AWGN at the D_2.

Non-Linear EH Model. For the EH process, we employ a practical non-linear model [39]; thus, the power utilized in the relaying step at D_1 (P_H) can be mathematically expressed as:

$$P_H(\rho, P_{in}) = \frac{P_{th}\left(\frac{1}{1+e^{-a(\rho P_{in}-b)}} - \frac{1}{1+e^{ab}}\right)}{1 - \frac{1}{1+e^{ab}}}, \quad (7)$$

where P_{th} is the threshold harvested power in saturation and a and b are constants related to the EH circuits as capacitance, resistance, and diode turn-on voltage. The adopted non-linear EH model from [39] closely matches experimental/practical EH circuit results for both the low (nW) and high (mW) wireless power harvested regime. The RF input power in the EH circuit at D_1 is defined as:

$$P_{in} = P_t|h_1|^2. \quad (8)$$

B. Signal-Interference-to-Noise-Ratio

D_1 receives a superimposed message y_1 from the $S \rightarrow RIS \rightarrow D_1$, so, according to NOMA the message of D_2 is detected firstly, and the corresponding signal-to-interference-plus-noise (SINR) is given by

$$\text{SINR}_{D_1}^{x_2} = \frac{|h_1|^2(1-\rho)P_t\alpha_2}{|h_1|^2(1-\rho)P_t\alpha_1 + |h_{SI}|^2P_H + \sigma^2}, \quad (9)$$

where we consider $\sigma_1^2 = \sigma_2^2 = \sigma^2$. After message x_2 is detected, it is eliminated from the received signal (5) by performing the SIC process[2], thus, the SINR in D_1 for detecting x_1 is given by

$$\text{SINR}_{D_1}^{x_1} = \frac{|h_1|^2(1-\rho)P_t\alpha_1}{|h_{SI}|^2P_H + \sigma^2}. \quad (10)$$

At D_2, the received SINR to detect x_2 from $S \rightarrow RIS \rightarrow D_2$ link, and the received SNR to detect x_2 from $D_1 \rightarrow D_2$ link are respectively expressed as

$$\text{SINR}_{D_2,S}^{x_2} = \frac{|h_2|^2P_t\alpha_2}{|h_2|^2P_t\alpha_1 + \sigma^2}, \quad (11)$$

and

$$\text{SNR}_{D_2,D_1}^{x_2} = \frac{P_H|h_{12}|^2}{\sigma^2}. \quad (12)$$

By adopting the maximum ratio combining (MRC) rule, the overall SINR in D_2 is equivalent to the sum of SINR from $S \rightarrow RIS \rightarrow D_2$ link and SNR from $D_1 \rightarrow D_2$ link, which can be expressed as [18], [38], [40]

$$\begin{aligned}\text{SINR}_{D_2}^{x_2} &= \text{SINR}_{D_2,S}^{x_2} + \text{SNR}_{D_2,D_1}^{x_2} \\ &= \frac{|h_2|^2P_t\alpha_2}{|h_2|^2P_t\alpha_1 + \sigma^2} + \frac{P_H|h_{12}|^2}{\sigma^2}. \end{aligned} \quad (13)$$

III. Statistics of the Channel and Harvested Power

Our objective is to derive new expressions that characterize the OP and ER for both devices in the RIS-aided C-FD-SWIPT-NOMA. To achieve this, we statistically analyze the cascaded channel for both devices and the harvested power at D_1 in the two subsequent subsections.

A. Statistical Channel Characterization

Before proceeding to the OP and ER derivations, it is of paramount importance to characterize the channel statistically. For that reason, let us define the following variables to facilitate our analysis $X_1 \triangleq \frac{|h_1|}{\sqrt{\beta_{ss}\beta_{s1}}} = \sum_{n=1}^{N}|h_{ss,n}||h_{s1,n}|$, and $X_2 \triangleq \frac{|h_2|}{\sqrt{\beta_{ss}\beta_{s2}}} = \left|\sum_{n=1}^{N} e^{j\phi_n}h_{ss,n}h_{s2,n}\right|$. According to Lemma 2 of [41], the distribution of X_1 and X_2, can be approximated as

$$X_1 \overset{\text{approx}}{\sim} \text{Gamma}\left(N\frac{\mu_{ss}^2\mu_{s1}^2}{1-\mu_{ss}^2\mu_{s1}^2}, \frac{1-\mu_{ss}^2\mu_{s1}^2}{\mu_{ss}\mu_{s1}}\right), \quad (14)$$

$$X_2 \overset{\text{approx}}{\sim} \text{Rayleigh}\left(\sqrt{\frac{N}{2}}\right), \quad (15)$$

where μ_{ss} and μ_{s1} are the mean of a Nakagami-m random variable given as [42, Table 5.2]

$$\mu_{ss} = \frac{\Gamma(m_{ss}+1/2)}{\Gamma(m_{ss})\sqrt{m_{ss}}}, \quad \mu_{s1} = \frac{\Gamma(m_{s1}+1/2)}{\Gamma(m_{s1})\sqrt{m_{s1}}}. \quad (16)$$

[2]Here we consider that the SIC process is performed perfectly, *i.e.*, we do not take into account an eventual residual error from this process.

Proceeding, Lemma 1 of [41] states that the distribution of X_1^2 and X_2^2, can be approximated respectively as

$$X_1^2 \overset{\text{approx}}{\sim} \text{Gamma}\left(k_1, \theta_1\right), \tag{17}$$

$$X_2^2 \overset{\text{approx}}{\sim} \text{Exponential}\left(N\right), \tag{18}$$

where

$$k_1 = \frac{\left(\mu_{X_1}^{(2)}\right)^2}{\mu_{X_1}^{(4)} - \left(\mu_{X_1}^{(2)}\right)^2}, \qquad \theta_1 = \frac{\mu_{X_1}^{(4)} - \left(\mu_{X_1}^{(2)}\right)^2}{\mu_{X_1}^{(2)}}, \tag{19}$$

with $\mu_{X_1}^{(m)}$ being the m-th moment of a Gamma random variable given as

$$\mu_{X_1}^{(m)} = \frac{\left(\frac{1-\mu_{ss}^2\mu_{s1}^2}{\mu_{ss}\mu_{s1}}\right)^m \Gamma\left(N\frac{\mu_{ss}^2\mu_{s1}^2}{1-\mu_{ss}^2\mu_{s1}^2} + m\right)}{\Gamma\left(N\frac{\mu_{ss}^2\mu_{s1}^2}{1-\mu_{ss}^2\mu_{s1}^2}\right)}. \tag{20}$$

Substituting Eq. (20) in Eq. (19), we obtain the following useful approximations

$$k_1 = \frac{\frac{N\mu_{ss}^2\mu_{s1}^2}{1-\mu_{ss}^2\mu_{s1}^2}\left(\frac{N\mu_{ss}^2\mu_{s1}^2}{1-\mu_{ss}^2\mu_{s1}^2}+1\right)}{4\left(\frac{N\mu_{ss}^2\mu_{s1}^2}{1-\mu_{ss}^2\mu_{s1}^2}\right)+6} \approx \frac{N\mu_{ss}^2\mu_{s1}^2}{4\left(1-\mu_{ss}^2\mu_{s1}^2\right)}, \tag{21}$$

$$\theta_1 = \left(\frac{1-\mu_{ss}^2\mu_{s1}^2}{\mu_{ss}\mu_{s1}}\right)^2\left(\frac{4N\mu_{ss}^2\mu_{s1}^2}{1-\mu_{ss}^2\mu_{s1}^2}+6\right) \approx 4N\left(1-\mu_{ss}^2\mu_{s1}^2\right). \tag{22}$$

The PDF and CDF of X_1^2 and X_2^2 can be written respectively as

$$f_{X_1^2}(x) = \frac{x^{k1-1}e^{\frac{-x}{\theta_1}}}{\Gamma(k_1)\theta_1^{k_1}}, \qquad f_{X_2^2}(x) = \frac{1}{N}e^{-\frac{x}{N}}, \tag{23}$$

$$F_{X_1^2}(x) = \frac{\gamma\left(k1, \frac{x}{\theta_1}\right)}{\Gamma(k_1)}, \qquad F_{X_2^2}(x) = 1 - e^{-\frac{x}{N}}. \tag{24}$$

B. Statistical Harvested Power Characterization

In order to understand the proposed system model, it is crucial to examine the distribution of the harvested power in D_1 when it acts as a relay ($\rho \neq 0$). To shed light on this aspect, we introduce the following lemma, which provides insights into the distribution of harvested power at D_1, crucially impacting its role as a relay in the system.

Lemma 1. *Let $\chi \triangleq \theta_1 a\rho P_t\beta_{ss}\beta_{s1}$, the distribution of the harvested power by the non-linear EH model in D_1, P_H, given by Eq. (7), can be approximated as a Gamma random variable, $P_H \overset{\text{approx}}{\sim} \text{Gamma}\left(k_{P_H}, \frac{P_{th}}{1-\frac{1}{1+e^{ab}}}\theta_{P_H}\right)$, where k_{P_H} and θ_{P_H} are given respectively by*

$$k_{P_H} = \frac{\left(\frac{(1+\chi)^{k_1}}{(1+\chi)^{k_1}+e^{ab}} - \frac{1}{1+e^{ab}}\right)^2}{\frac{1}{1+\frac{2e^{ab}}{(1+\chi)^{k_1}}+\frac{e^{2ab}}{(1+2\chi)^{k_1}}} - \left(\frac{(1+\chi)^{k_1}}{(1+\chi)^{k_1}+e^{ab}1} - \frac{1}{1+e^{ab}}\right)^2}, \tag{25}$$

$$\theta_{P_H} = \frac{\frac{1}{1+\frac{2e^{ab}}{(1+\chi)^{k_1}}+\frac{e^{2ab}}{(1+2\chi)^{k_1}}} - \left(\frac{(1+\chi)^{k_1}}{(1+\chi)^{k_1}+e^{ab}} - \frac{1}{1+e^{ab}}\right)^2}{\frac{(1+\chi)^{k_1}}{(1+\chi)^{k_1}+e^{ab}1} - \frac{1}{1+e^{ab}}}. \tag{26}$$

Proof. The proof is available in Appendix A. ∎

Theorem 1. *Using Lemma 1, we can directly compute the average harvested power in closed form for the non-linear EH model as follows*

$$\widetilde{P}_H = k_{P_H}\theta_{P_H} = \frac{P_{th}}{1-\frac{1}{1+e^{ab}}}\left(\frac{(1+\chi)^{k_1}}{(1+\chi)^{k_1}+e^{ab}} - \frac{1}{1+e^{ab}}\right). \tag{27}$$

Proof. The proof is straightforward, as it follows directly from the properties of the Gamma distribution. The mean of the Gamma distribution is given by the product of its shape parameter (k_{P_H}) and scale parameter (θ_{P_H}) [42, Table 5.2]. ∎

Corollary 1. *The total number of reflective elements required on the RIS to attain the threshold power P_{th} with an accuracy factor ϵ can be determined as follows*

$$N^{\text{sat}} \approx \sqrt{\frac{ab - \ln\epsilon}{a\rho P_t\beta_{ss}\beta_{s1}\mu_{ss}^2\mu_{s1}^2}}. \tag{28}$$

Proof. According to Eq. (27), the saturation $\widetilde{P}_H = P_{th}$ occurs when $\frac{(1+\chi)^{k_1}}{(1+\chi)^{k_1}+e^{ab}} = 1$; thus, assuming an error ϵ we have:

$$\begin{aligned} \frac{e^{ab}}{(1+\theta_1 a\rho P_t\beta_{ss}\beta_{s1})^{k_1}} &\leq \epsilon \\ ab - \ln\epsilon &\leq k_1\ln\left(1+\theta_1 a\rho P_t\beta_{ss}\beta_{s1}\right) \\ ab - \ln\epsilon &\approx k_1\theta_1 a\rho P_t\beta_{ss}\beta_{s1}, \end{aligned} \tag{29}$$

where we utilized $\ln(1+x) \approx x$, for low values of x. substituting Eq. (21) and Eq. (22) in Eq. (29) this completes the proof. ∎

Corollary 2. *The transmit power required on the BS to attain the threshold power P_{th} with an accuracy factor ϵ can be determined as follows*

$$P_t^{\text{sat}} \approx \frac{ab - \ln\epsilon}{a\rho\beta_{ss}\beta_{s1}N^2(\mu_{ss}\mu_{s1})^2}. \tag{30}$$

Proof. The proof is obtained similarly to the proof of Corollary 1. ∎

Remark 1. We can see that both N^{sat} and P_t^{sat} are inversely proportional to the cascaded channel conditions of D_1. Moreover, both are parameterized in an accuracy factor ϵ, which value will be further explored in Section V.

IV. System Performance Analysis

In order to explore the advantages in RIS-aided C-FD-SWIPT-NOMA scenarios, in this section new expressions for the OP and SE of D_1 and D_2 in the RIS-aided C-FD-SWIPT-NOMA system considering generalized Nakagami-m

fading channels and non-linear energy harvesting model are presented, where we consider that target SINRs are determined by the devices' QoS requirements. The OP metric is crucial for assessing the reliability of transmissions in 5G and B5G systems, especially in the URLLC mode.

A. Device 1 Outage Probability

Particularly, the outage behavior for D_1 occurs: since D_1 cannot detect effectively x_2 and consequently its own message; or, when x_2 is detected successfully but an error occurs to decode x_1; *i.e.*, the outage occurs except for the case that both D_1 and D_2 messages are decoded successfully. Mathematically it can be formulated by

$$P_{\text{out}}^{D_1} = 1 - \Pr(\text{SINR}_{D_1}^{x_2} \geq \gamma_{th_2}, \text{SINR}_{D_1}^{x_1} \geq \gamma_{th_1}), \quad (31)$$

where $\gamma_{th_2} = 2^{\hat{R}_2} - 1$ and $\gamma_{th_1} = 2^{\hat{R}_1} - 1$, with $\hat{R}_1$ and $\hat{R}_2$ being the target rates of the D_1 and D_2, respectively. The following theorem provides the OP of D_1 for the RIS-aided C-FD-SWIPT-NOMA system.

Theorem 2. *Let* $\xi = \max\left(\frac{\gamma_{th}}{P_t(1-\rho)\alpha_1}, \frac{\gamma_{th2}}{P_t(1-\rho)(\alpha_2-\alpha_1\gamma_{th2})}\right)$, *if* $\frac{\alpha_2}{\alpha_1} < \gamma_{th_2}$ *or* $\frac{k_1\theta_1\beta_{ss}\beta_{s1}}{\xi_1} < \omega\widetilde{P}_H$, D_1 *is in outage,* $P_{\text{out}}^{D_1} = 1$, *otherwise the closed-form expression for the OP of* D_1 *under Nakagami-*m *fading is given by* (32) *at bottom of this page*

Proof. The proof is available in Appendix B. ∎

B. Device 2 Outage Probability

The outage behavior for the D_2 can occur in two distinct ways: *1)* D_1 detects effectively the D_2' message however the sum of the SINRs after MRC in D_2 is lower than the SINR of threshold **or** *2)* D_1 cannot detect effectively the D_2' message and the SINR becoming of the $S \rightarrow RIS \rightarrow D_2$ is lower than the SINR threshold. Therefore, the OP of D_2 can be formulated as

$$\begin{aligned} P_{\text{out}}^{D_2} &= \Pr(\text{SINR}_{D_1}^{x_2} < \gamma_{th_2}, \text{SINR}_{D_2,S}^{x_2} < \gamma_{th_2}) \\ &\quad + \Pr(\text{SINR}_{D_1}^{x_2} \geq \gamma_{th_2}, \text{SINR}_{D_2}^{x_2} < \gamma_{th_2}). \end{aligned} \quad (33)$$

Theorem 3. *For the case when* $\alpha_2 < \alpha_1\gamma_{th_2}$, *then* $P_{\text{out}}^{\text{U}_2} = 1$, *otherwise, then the closed-form expression for the OP of* D_2 *under Nakagami-*m *fading is given by Eq.* (34)

Proof. Please refer to Appendix C. ∎

C. Device 1 Ergodic Rate

In this subsection, in order to understand better the RIS-aided C-FD-SWIPT-NOMA system with a non-linear EH model, we proposed an upper bound for the ER of D_1.

Theorem 4. *Assuming the* D_1 *can successfully detect the message of* D_2 *and itself message, then the upper bound spectral efficiency for* D_1 *under Nakagami-*m *fading can be calculated as Eq.* (35) *at the bottom of next page*

Proof. The Proof is available in Appendix D. ∎

D. Device 2 Ergodic Rate

An approximation for the ER of D_2 is proposed in the following theorem.

Theorem 5. *Assuming the* D_2 *decoded successfully the message of* x_2 *from* $S \rightarrow RIS \rightarrow D_2$ *as well as the rebuilt message* $\hat{x}_2$ *from the cooperative link (since* D_1 *decoded successfully* x_2 *in order to relay it), the upper bound for ER of* D_2 *is given as Eq.* (36)

Proof. The Proof is available in Appendix E. ∎

Corollary 3. *When* $P_t \rightarrow \infty$ *or* $N \rightarrow \infty$, *the asymptotic ER of* D_2, R_2^{∞} *can be computed as*

$$R_2^{\infty} = \underbrace{\log_2\left(1 + \frac{\alpha_2}{\alpha_1}\right)}_{\substack{\text{Non-Cooperative} \\ \text{Asymptotic Rate}}} + \underbrace{\frac{e^{\frac{\sigma^2\left(1+\frac{\alpha_2}{\alpha_1}\right)}{P_{th}\beta_{12}}}}{\log(2)}\Gamma\left(0, \frac{\sigma^2\left(1+\frac{\alpha_2}{\alpha_1}\right)}{P_{th}\beta_{12}}\right)}_{\substack{\text{Cooperative} \\ \text{Asymptotic Rate}}} \quad (37)$$

Remark 2. It is worth noting that the asymptotic rate of D_2 in RIS-aided cooperative systems with NOMA comprises two distinct terms: the non-cooperative asymptotic rate and the cooperative asymptotic rate. Surprisingly, the number of RIS elements, denoted by N, does not influence the asymptotic rate of D_2. However, this does not imply that the RIS cannot have an impact on cooperative systems.

V. Simulation Results

In this section, we aim to confirm through Monte-Carlo simulations (MCs) the accuracy of our previous mathematical analysis and illustrate the achievable enhanced performance of the RIS-aided C-FD-SWIPT-NOMA system. The simulation results are averaged over 10^6 realizations. Unless stated otherwise, the parameter values adopted in this section are presented in Table I. In the following numerical results, the Monte-Carlo simulations curves are labeled as "MCs", and the derived analytical expression-based curves are labeled as "Analytical". We use blue color lines to represent the D_1 performance, while D_2 is represented by green color lines.

$$P_{\text{out}}^{D_1} \approx \begin{cases} \dfrac{\gamma\left(k_1, \frac{\xi\sigma^2}{\theta_1\beta_{ss}\beta_{s1}P_t}\right)}{\Gamma(k_1)} & \rho = 0 \\ \dfrac{\gamma\left(k_1, \frac{\xi\sigma^2}{\theta_1\beta_{ss}\beta_{s1}P_t(1-\rho)}\right)}{\Gamma(k_1)} + \dfrac{e^{\frac{\sigma^2}{\omega\widetilde{P}_H}}}{\Gamma(k_1)} \dfrac{\Gamma\left(k_1, \frac{\xi\sigma^2}{\theta_1\beta_{ss}\beta_{s1}P_t(1-\rho)} + \frac{\sigma^2}{\xi\omega\widetilde{P}_H}\right)}{\left(1 + \frac{\theta_1\beta_{ss}\beta_{s1}P_t(1-\rho)}{\xi\omega\widetilde{P}_H}\right)^{k_1}} & \rho \neq 0 \end{cases} \quad (32)$$

$$P_{\text{out}}^{D_2} \approx \begin{cases} 1 - e^{\frac{-\gamma_{th2}\sigma^2}{NP_t\beta_{ss}\beta_{s2}(\alpha_2-\alpha_1\gamma_{th2})}} & \rho = 0 \\ \frac{\gamma\left(k_1, \frac{\gamma_{th2}\sigma^2}{\theta_1 P_t(1-\rho)\beta_{ss}\beta_{s1}(\alpha_2-\alpha_1\gamma_{th2})}\right)}{\Gamma(k_1)} \left[- e^{\frac{-\gamma_{th2}\sigma^2}{NP_t\beta_{ss}\beta_{s2}(\alpha_2-\alpha_1\gamma_{th2})}} + e^{-\frac{\gamma_{th2}\sigma^2}{\beta_{12}\widetilde{P}_H}} \right. & \\ \left. + \frac{\sigma^2}{\beta_{12}\widetilde{P}_H} \left(\frac{e^{-\frac{\gamma_{th2}\sigma^2}{\beta_{12}\widetilde{P}_H}} - e^{\frac{-\gamma_{th2}\sigma^2}{P_t\beta_{ss}\beta_{s2}N\alpha_2}}}{\left(\frac{\sigma^2}{P_tN\beta_{ss}\beta_{s2}\alpha_2} - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H}\right)} \right) \right] + 1 - e^{-\frac{\gamma_{th2}\sigma^2}{\beta_{12}\widetilde{P}_H}} - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H} \left(\frac{e^{-\frac{\gamma_{th2}\sigma^2}{\beta_{12}\widetilde{P}_H}} - e^{\frac{-\gamma_{th2}\sigma^2}{P_t\beta_{ss}\beta_{s2}N\alpha_2}}}{\left(\frac{\sigma^2}{P_tN\beta_{ss}\beta_{s2}\alpha_2} - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H}\right)} \right) & \rho \neq 0 \end{cases} \quad (34)$$

To gain a comprehensive understanding of the studied system model, assess the performance of the transmission scheme, and demonstrate the accuracy of the derived analytical expressions and obtained insights, we adopt two distinct scenarios: I), where D_1 encounters poor channel conditions compared to the BS and RIS, with $\beta_{ss} = -45$ dB and $\beta_{s1} = -65$ dB, resulting in low input power and consequently low harvested power. On the other hand, we also assessed the opposing case, considered in scenario II), where favorable channel conditions for D_1 concerning the RIS and BS are assumed, with $\beta_{ss} = -30$ dB and $\beta_{s1} = -30$ dB, leading to high input power and consequently high harvested power.

Fig. 2 illustrates the average harvested power as a function of the transmit power at the base station. We can firstly observe the strong agreement between the analytical result obtained from Eq. (27) and the MCs result. This comparison provides convincing evidence for the accuracy of the proposed Theorem 1, validating its effectiveness across a wide range of transmit power (P_t) values and the number of reflective elements (N). It emphasizes the significant benefits of using this lemma in characterizing the relationship between harvested power, transmitted power, and the number of reflective elements while offering valuable guidance for applications in the RIS-aided C-FD-SWIPT-NOMA system.

Additionally, Fig. 2 highlights the non-linear EH model's behavior, which is paramount to understanding the subsequent analyses. In scenario I, we can see that the average harvested power increases with P_t and N, however, due to the poor conditions of channels of D_1, the average harvested power does not achieve saturation. Alternatively, for scenario II, we can see that the harvested power in D_1 reaches its maximum value, denoted as P_{th}, for the computed values of P_t^{sat} according to the Corollary 2, for the three values of N adopted.

TABLE I
ADOPTED SIMULATION PARAMETERS.

Parameter	Value
RIS-aided C-FD-SWIPT-NOMA system	
Transmit power	$P_t = [-20, 45]$ dBm
Noise power	$\sigma^2 = -95$ dBm
Power Allocation coefficients	α_1 = 0.1, α_2 = 0.9
Residual self-interference	$\omega = -35$ dB;
	$\omega = [-90; -35]$ dB (Fig. 6)
Minimum rate	I) $R_1 = 6$, $R_2 = 1$ bps/Hz
	II) $R_1 = 18$, $R_2 = 3$ bps/Hz
Accuracy factor	$\epsilon = 0.05$
Non-Linear Energy Harvesting Parameters	
EH coefficient	$\rho \in [0, 1]$
EH model constants	$a = 150$; $b = 0.014$ [43], [44]
Max. RF (harvested) power	$P_{th} = 250$ μW;
	$P_{th} = 500$ μW, 5 mW (Fig. 6)
Channel Parameters	
Channel Model S-RIS/RIS-Ds	Nakagami-m
Shape Parameter S-RIS	m_{ss} = 5
Shape Parameter RIS-Ds	$m_{s1} = 4$ $m_{s2} = 1.5$
Channel Model D2D	$h_{12} \sim \mathcal{CN}(0,\beta_{12})$
Channel Gains	$\beta_{s2} = -90$; $\beta_{12} = -40$ [dB]
	I) $\beta_{ss} = -45$ $\beta_{s1} = -65$ [dB]
	II) $\beta_{ss} = -30$ $\beta_{s1} = -30$ [dB]

$$R_1 \leq \begin{cases} \log_2\left(1 + \frac{N^2\beta_{ss}\beta_{s1}\mu_{ss}^2\mu_{s1}^2 P_t\alpha_1}{\sigma^2}\right) & \rho = 0 \\ \log_2\left(1 + \frac{N^2\beta_{ss}\beta_{s1}\mu_{ss}^2\mu_{s1}^2(1-\rho)P_t\alpha_1}{\sigma^2}\right) - \frac{e^{\frac{\sigma^2}{\widetilde{P}_H\omega}}}{\log(2)}\Gamma\left(0, \frac{\sigma^2}{\widetilde{P}_H\omega}\right) & \rho \neq 0 \end{cases} \quad (35)$$

$$R_2 \leq \begin{cases} \log_2\left(1 + \frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1 + \sigma^2}\right) & \rho = 0 \\ \log_2\left(1 + \frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1 + \sigma^2}\right) + \frac{e^{\frac{\sigma^2\left(1+\frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1+\sigma^2}\right)}{\widetilde{P}_H\beta_{12}}}}{\log(2)}\Gamma\left(0, \frac{\sigma^2\left(1 + \frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1+\sigma^2}\right)}{\widetilde{P}_H\beta_{12}}\right) & \rho \neq 0 \end{cases} \quad (36)$$

A noteworthy observation is that increasing N leads to a significant boost in harvested power, as confirmed by Eq. (27), achieving a saturation point at lower values of P_t^{sat}, confirmed by Eq. (30) as well.

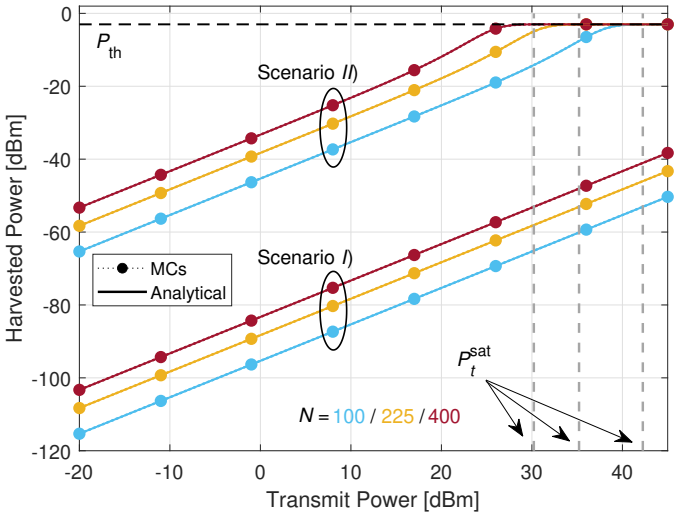


Fig. 2. Average Harvested Power ($\widetilde{P}_H$) *vs* Transmit Power (P_t) for two different scenarios: I) low input power; II) high input power. We also consider $N = \{100, 225, 400\}$ elements.

Fig. 3 and 4 depicts the ER and OP of D_1 and D_2 *vs* transmit power (P_t), respectively, considering $N = 20^2 = 400$ RIS elements, while the energy harvested factor is adopted as $\rho = 0$ (non-cooperative mode) and $\rho = 0.4$ (cooperative mode). Firstly, it is noticeable that the derived OP expression given by Eq. (32) and Eq. (34) are very accurate as well as the derived upper bound expressions provided by Eq. (35) Eq. (36) are very tight, closely matching the obtained MCs results.

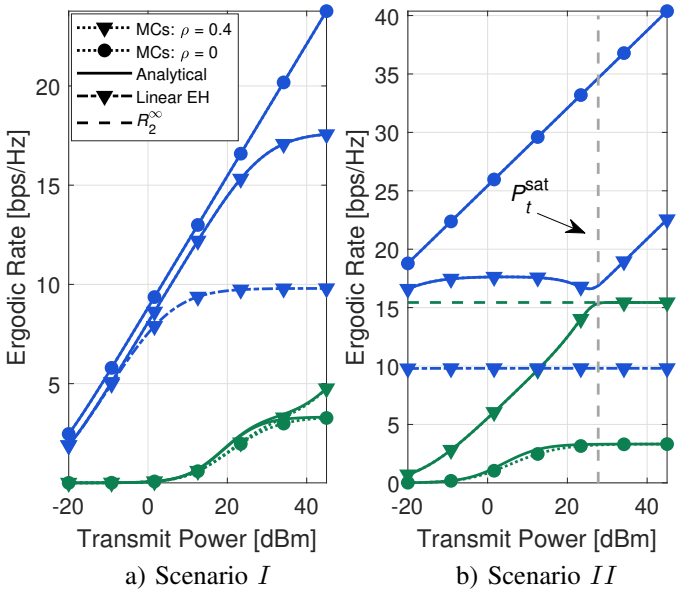


Fig. 3. ER for D_1 and D_2, denoted as blue and green color, respectively. $N = 20^2 = 400$ elements; energy harvested factor $\rho = 0$ and $\rho = 0.4$

Analysis of Fig. 3 and 4 reveals a clear trade-off between the performance of the devices. When D_1 acts as a relay, its own performance is inevitably compromised to enhance the performance of D_2. Parameters such as residual self-interference (ω), number of reflective elements (N), maximum harvested power (P_{th}), and energy harvest coefficient (ρ) impact the system performance and should all be thoroughly examined in the sequel to gain a comprehensive understanding of their real impacts on the overall system performance.

In Fig. 3, we observe distinct ergodic rate behaviors for scenarios I and II. For scenario I, D_2 exhibits a higher ergodic rate in the cooperative mode compared to the non-cooperative mode ($\rho = 0$); such a difference becomes more pronounced for transmit power values above 30 dBm. However, in D_1, the ergodic rate for the cooperative mode results lower than in the non-cooperative mode; such a gap constantly increases, mainly for high P_t values. This constant difference in D_1 for low power regime, $P_t \leq 15$dBm, is due to D_1 utilizing only $1 - \rho = 60\%$ of the received power for information detection, impacting its performance regardless of the transmit power. Moreover, when $P_t \geq 22$ dBm, the impact on the rate of D_1 in the cooperative scenario is further affected. In cooperative mode, the residual self-interference justifies the ER degradation since the transmit power increasing leads to higher harvested power and consequently to higher RSI, increasing interference, and decreasing SINR of D_1, as predicted by Eq. (10). The same degradation mechanism occurs in scenario II, but it is more pronounced since the rate of D_1 is severely degraded w.r.t. the non-cooperative curve for any P_t value, once the harvested power is very higher in this scenario than the former.

In scenario II, we observe a further enhancement in the rate of D_2, demonstrating the potential of cooperative communication in scenarios where D_1 has high input power. However, it is worth noting that even with substantial rate increments, D_2 reaches a maximum limit when the transmit power increases considerably. We refer to this asymptotic behavior as R_2^{∞}, as stated in Corollary 3. This observation is insightful, as it confirms that the cooperative system with RIS cannot provide an unlimited rate for D_2, even in the asymptotic regime. Additionally, this result shows that N does not influence the value of R_2^{∞}, indicating that the RIS cannot impact the asymptotic ergodic rate of D_2.

Finally, in Fig. 3 and 4, we also plotted the widely adopted *linear* EH model, defined as $P_H = \eta P_{in}$, with an efficiency parameter, adopted herein as $\eta = 0.9$. In scenario I, for $P_t > 15$ dBm, and in scenario II, for any value of $P_t > -20$ dBm, we observe that the rate of D_1 saturates at a maximum value. This phenomenon occurs because the linear EH model provides misleading insights regarding high-power regimes. According to this model, the harvested power has no limit and can achieve unlimited values, which is physically impractical. Therefore, based on Eq. (10), the maximum rate of D_1 for the linear EH model can be expressed as $\log_2\left(1 + \frac{(1-\rho)\alpha_1}{\eta\omega}\right)$. However, this is not true, as the non-linear EH model introduces limitations on the interference in the denominator of Eq. (10) justifying the behavior of the rate of D_1 in Fig. 3 b), which reaches a minimum value at the saturation point and then begins to increase again. This observation highlights the importance of considering realistic non-linear EH models, as

they accurately represent the system's behavior, particularly at high power levels. The adopted non-linear EH model takes into account the limitations of the energy harvesting process, leading to more reliable rate predictions and aiding in the design of practical and efficient wireless communication systems.

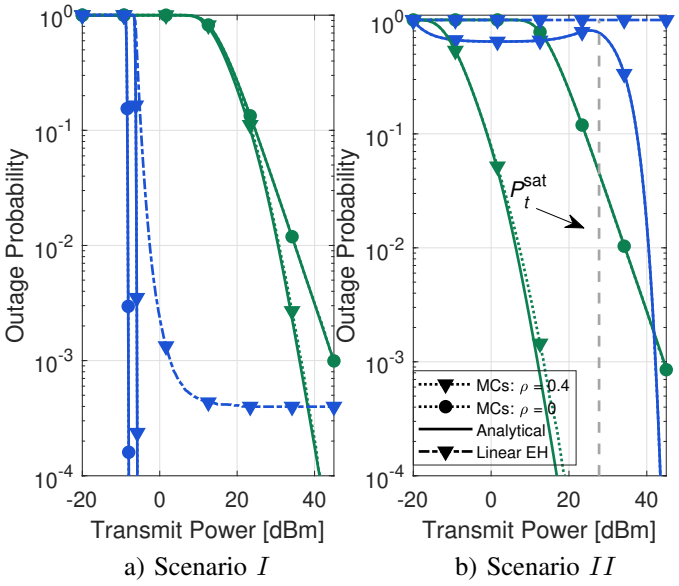


Fig. 4. OP for D_1 and D_2, denoted as blue and green color, respectively; $N = 400$ elements; $\rho = 0$ (non-cooperative) and $\rho = 0.4$ (cooperative).

The outage probability in Fig. 4 again reveals distinct behaviors for low and high power regimes, scenarios I and II, respectively. In scenario I, as P_t increases, the cooperative mode becomes beneficial, reducing the outage probability of D_2 for $P_t > 30$ dBm. Interestingly, the impact of the cooperative mode on D_1 is minimal, indicating that it does not significantly affect its performance. In contrast, for scenario II, we observe an interesting reduction in outage probability for D_2 across a wide range of P_t. This reduction is particularly noteworthy as it comes without any additional power consumption at the source, making it a promising gain in terms of sustainability. However, the cooperative mode increases outage probability for D_1, severely degrading its performance. Nevertheless, it is essential to note that despite the significant increment in OP for D_1, it does not reach an outage floor. This is a vital observation as D_1 achieves an outage limit for the linear EH model in both power regimes, scenarios I and II. These findings highlight the benefits and trade-offs of RIS-aided C-FD-SWIPT-NOMA systems. The cooperative mode substantially improves outage probability for D_2, promoting sustainability without additional energy consumption. However, it may negatively impact the outage probability for D_1, requiring careful consideration in system design, mainly over ω, to balance performance gains and potential drawbacks.

Figure 5 illustrates the ER of D_1 as the energy harvesting coefficient (ρ) increases, considering four different values of residual self-interference (ω). As the values of ρ increase, the performance of D_1 decreases, which is expected due to its utilization of lower power levels for information decoding.

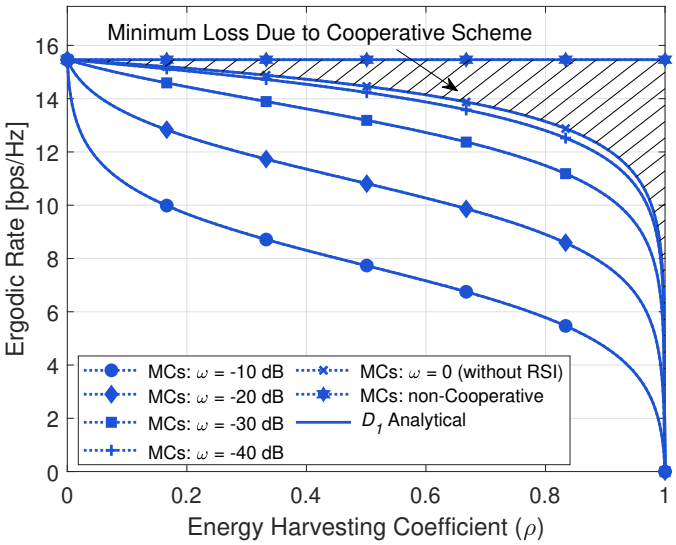


Fig. 5. Ergodic rate of D_1 *vs* energy harvest coefficient (ρ) for scenario I, with $N = 400$ and $P_t = 20$ dBm and different values of residual self-interference (ω).

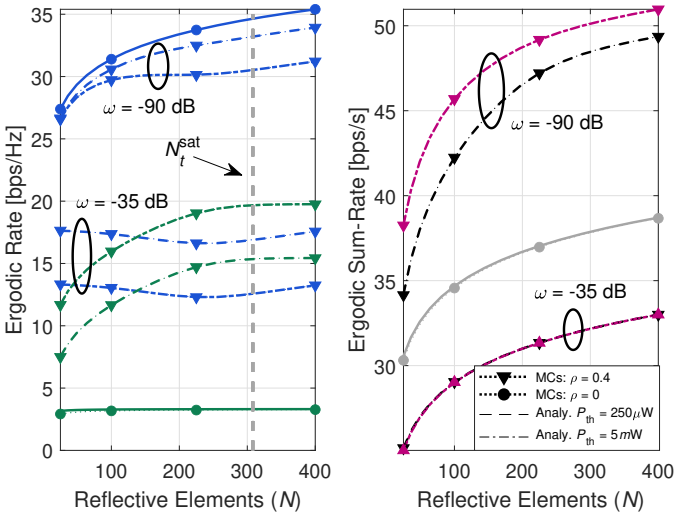
This leads to a reduced SINR and, consequently, worsened performance.

Furthermore, when ω decreases, particularly for $\omega \leq -40$ dB, the system operates closer to the ideal scenario with no residual self-interference ($\omega = 0$). In other words, the system approaches the minimum loss region, as indicated by the hatched area in Fig. 5. Specifically, at $\omega = -40$ dB and $\rho = 0.4$, R_1 achieves a rate of 15.7 bps/Hz, whereas at $\omega = -10$ dB, maintaining the same $\rho = 0.4$, R_1 decreases to 8.6 bps/Hz. This demonstrates that reducing the residual self-interference is crucial, even in the low power regime (scenario I), for the RIS-aided C-FD-SWIPT-NOMA system to operate in good conditions concerning D_1, which can save approximately 50% of its rate.

Figure 6 illustrates the impact of the number of RIS elements on the individual rate (R_i, $i = 1, 2$) and sum rate ($R_1 + R_2$) of NOMA devices with fixed parameters $P_t = 30$ dBm, $\rho = 0.4$, $\omega = \{-90, -35\}$ dB, considering $P_{th} = 5$ mW and $P_{th} = 250$ μW in scenario II. First, we can observe two distinct behaviors corresponding to different levels of residual self-interference: one with very low self-interference ($\omega = -90$ dB) and the other with a considerable level of residual self-interference ($\omega = -35$ dB).

For the scenarios with very low self-interference, the RIS-aided C-FD-SWIPT-NOMA system remarkably outperforms the RIS-aided non-C-FD-SWIPT-NOMA system in the range of RIS elements. This improvement in system performance is evident in the substantially increased sum rate, which is attributed to D_1 acting as a relay and enhancing the rate of D_2. At $N = 400$ elements, the cooperative system exhibits an enhancement in sum rate of about 10 and 13 bps/s, *i.e.*, 31% and 34% increasing w.r.t. the non-cooperative mode for both maximum harvested power $P_{th} = 250\mu$W (low) and 5 mW (high), respectively, resulting in remarkable gains. Conversely, when the residual self-interference increases substantially to $\omega = -35$ dB, the non-cooperative system ($\rho = 0$) outperforms

the cooperative system. This is due to the increased level of residual self-interference, which leads to a rate loss more significant in device D_1 than the obtained rate gain with D_2, deteriorating the overall system performance. Thus, this observation emphasizes the importance of efficient EH circuits, which contribute to attaining improved system performance and higher overall rate gains. Moreover, the results indicate that the residual self-interference should be minimized as the power in the EH process increases.



a) ER for D_1 (blue) and D_2 (green) solid lines: analytical

b) Sum-rate solid gray line: non-cooperative

Fig. 6. ER *vs* N where $\rho = \{0; 0.4\}$ for $\omega = -\{90; 35\}$ dB, $P_t = 30$ dBm and $P_{th} = 5$ mW and $P_{th} = 250$ μW over scenario II.

Additionally, the figure highlights the significant impact of the maximum harvested power (P_{th}) on the system performance. It is shown that by increasing the maximum harvested power, D_1 has the capacity to harvest more power, consequently exhibiting a greater impact on the D_2 performance. Therefore, this finding reveals that a more robust design of EH circuits allows a more satisfactory cooperative system performance enhancement.

Finally, we can see how the Corollary 1 can be applied in practice, where the number of reflective elements to achieve the saturation (N^{sat}) can be obtained.

VI. Conclusions

This work has investigated the downlink communication reliability and throughput of two devices in a RIS-aided C-FD-SWIPT-NOMA system operating under Nakagami-m fading; the device with better channel conditions (D_1) leverages of SWIPT technique equipped with a non-linear EH circuit to act as an FD-DF relay for assisting the device with poor channel conditions (D_2). It was shown that the harvested power in D_1 can be approximated as a Gamma random variable, allowing finding a closed-form solution for the average harvested power and consequently flexible modeling of the system. We have derived tractable closed-form expressions for the performance of both types of users, including ***a***) upper bound for ER of a D_1/D_2 devices; ***b***) tight expression for OP of both devices; and ***c***) the ergodic sum rate, and the ergodic rate that D_2 can attain in the (non)-cooperative modes. In addition, we assessed the impact of the RIS dimension on both adopted metrics. Moreover, we find closed-form expressions for the transmitted power (P_t^{sat}) and RIS elements (N^{sat}) for the non-linear EH model to achieve saturation. Furthermore, it has been demonstrated that the RIS can significantly enhance the system sum-rate when D_1 is acting as a cooperative relay. However, this gain is achieved only when the residual self-interference is effectively suppressed, underscoring the sensitivity of cooperative systems to residual self-interference. Simulation results validate the correctness and effectiveness of the developed theoretical analysis while demonstrating the advantages of the analyzed RIS-aided C-FD-SWIPT-NOMA system over the non-cooperative RIS-aided system.

Appendix A
Proof of Lemma 1

Our objective is to approximate the power utilized in the relaying step, P_H, defined by in Eq. (7), as a Gamma random variable using the method of moments. For this reason, let us define $Y = \frac{1}{1+e^{-a(\rho P_{in}-b)}}$, thus the first moment of Y can be lower bounded as

$$\mathbb{E}[Y] \geq \frac{1}{1+\mathbb{E}\left[e^{-a(\rho P_{in}-b)}\right]} = \frac{1}{1+\frac{e^{ab}}{(1+\theta_1 a\rho P_t\beta_{ss}\beta_{s1})^{k_1}}} \tag{38}$$

where, utilizing Eq. (8), we obtain

$$\begin{aligned}\mathbb{E}\left[e^{-a(\rho P_{in}-b)}\right] &= e^{ab}\int_0^\infty e^{-a\rho P_t\beta_{ss}\beta_{s1}x} f_{X_1^2}(x)dx \\ &\overset{(i)}{=} \frac{e^{ab}}{\Gamma(k_1)\theta_1^{k_1}}\int_0^\infty x^{k1-1}e^{-x\left(\frac{1}{\theta_1}+a\rho P_t\beta_{ss}\beta_{s1}\right)} \\ &= \frac{e^{ab}}{\left(1+\theta_1 a\rho P_t\beta_{ss}\beta_{s1}\right)^{k_1}}\end{aligned} \tag{39}$$

where [45, 3.351.3] is utilized in (i). By proceeding with the same steps above, similarly, the second moment of Y can be given by

$$\mathbb{E}[Y^2] \geq \frac{1}{1+\frac{2e^{ab}}{(1+\theta_1 a\rho P_t\beta_{ss}\beta_{s1})^{k_1}}+\frac{e^{2ab}}{(1+2\theta_1 a\rho P_t\beta_{ss}\beta_{s1})^{k_1}}}. \tag{40}$$

The shape k_{P_H} and scale θ_{P_H} parameters of P_H, which will be approximated as Gamma random variable, are given by [41]

$$k_{P_H} = \frac{\left(\mathbb{E}[Y]-\frac{1}{1+e^{ab}}\right)^2}{\mathbb{E}[Y^2]-\mathbb{E}[Y]^2}, \tag{41}$$

$$\theta_{P_H} = \frac{\mathbb{E}[Y^2]-\mathbb{E}[Y]^2}{\mathbb{E}[Y]-\frac{1}{1+e^{ab}}}. \tag{42}$$

By defining, $\chi \triangleq \theta_1 a\rho P_t\beta_{ss}\beta_{s1}$ and plugging (38) and (40) in (41) and (42); (25) (26) are obtained and this completes the proof.

APPENDIX B
PROOF OF THEOREM 2

After performing some mathematical manipulations, Eq. (31) can be written as follows

$$P_{\text{out}}^{D_1} = \Pr\left(X_1^2 < \frac{\xi(|h_{SI}|^2 P_H + \sigma^2)}{P_t(1-\rho)\beta_{ss}\beta_{s1}}\right), \tag{43}$$

where $\xi = \max\left(\frac{\gamma_{th}}{\alpha_1}, \frac{\gamma_{th2}}{(\alpha_2 - \alpha_1\gamma_{th2})}\right)$. We observe that when $\rho = 0$ or $\omega = 0$, it leads to $P_H = 0$ or $|h_{SI}|^2 = 0$, respectively. In both of these particular cases, we can compute the OP of D_1 as the following

$$P_{\text{out}}^{D_1} = \frac{1}{\Gamma(k_1)}\gamma\left(k_1, \frac{\xi\sigma^2}{\theta_1 P_t(1-\rho)\beta_{ss}\beta_{s1}}\right), \tag{44}$$

where we utilized the CDF of X_1^2 as defined in (24).

When we have $\rho \neq 0$ and $\omega \neq 0$, (43) should be further analyzed. Firstly, we should notice that according to Lemma 1, P_H is a Gamma random variable with shape and scale parameters given by k_{PH} and θ_{PH} respectively. Thus (43) can be written as

$$P_{\text{out}}^{D_1} = 1 - \Pr\left(\underbrace{\frac{|h_{SI}|^2\xi P_H}{P_t(1-\rho)}}_{V} - \underbrace{X_1^2\beta_{ss}\beta_{s1}}_{Z} < -\frac{\xi\sigma^2}{P_t(1-\rho)}\right), \tag{45}$$

let us define $V \triangleq \frac{|h_{SI}|^2\xi P_H}{P_t(1-\rho)}$ and $Z \triangleq X_1^2\beta_{ss}\beta_{s1}$. Since P_H range from low values lower than nano-Watts to P_{th}, which assumes values till mili-Watts, its values are lower than $|h_{SI}|^2$, which is parameterized in ω and undergo an Exponential distribution, due to this, here, we propose to approximate V as an exponential random variable, *i.e.*, $V \overset{\text{approx}}{\sim} \text{Exponential}\left(\frac{\omega\xi\widetilde{P}_H}{P_t(1-\rho)}\right)$. Furthermore, since Z is the product of X_1 with a constant, utilizing the scale property of Gamma distribution, $Z \sim \text{Gamma}(k_1, \theta_1\beta_{ss}\beta_{s1})$, hence, (45) can be written as

$$\begin{aligned}
P_{\text{out}}^{D_1} &= 1 - \int_{\frac{\xi\sigma^2}{P_t(1-\rho)}}^{\infty} F_V\left(z - \frac{\xi\sigma^2}{P_t(1-\rho)}\right) f_Z(z)dz \\
&\overset{(i)}{=} 1 - \int_{\frac{\xi\sigma^2}{P_t(1-\rho)}}^{\infty} f_Z(z)dz + e^{\frac{\sigma^2}{\omega\widetilde{P}_H}} \int_{\frac{\xi\sigma^2}{P_t(1-\rho)}}^{\infty} e^{-\frac{zP_t(1-\rho)}{\omega\xi\widetilde{P}_H}} f_Z(z) \\
&= \frac{\gamma\left(k_1, \frac{\xi\sigma^2}{\theta_1\beta_{ss}\beta_{s1}P_t(1-\rho)}\right)}{\Gamma(k_1)} \\
&\quad + \frac{e^{\frac{\sigma^2}{\omega\widetilde{P}_H}}}{\Gamma(k_1)} \frac{\Gamma\left(k_1, \frac{\xi\sigma^2}{P_t(1-\rho)\theta_1\beta_{ss}\beta_{s1}} + \frac{\sigma^2}{\omega\widetilde{P}_H}\right)}{\left(\frac{P_t(1-\rho)\theta_1\beta_{ss}\beta_{s1}}{\omega\widetilde{P}_H\xi} + 1\right)^{k_1}},
\end{aligned} \tag{46}$$

where in (i) we utilized the CDF of exponential random variable V, given by $F_V(z) = 1 - e^{-\frac{zP_t(1-\rho)}{\omega\xi\widetilde{P}_H}}$, for first integral, and we apply [45, 3.381.9] for solving the second integral.

APPENDIX C
PROOF OF THEOREM 3

To obtain the OP of D_2, it is reasonable to consider two different cases: *I)* when the D_1 does not operate as relay ($\rho = 0$); *II)* when D_1 operates as a relay ($\rho \neq 0$).

I) $\rho = 0$ *(D_1 does not act as a relay)*

By recalling the established definition in subsection III-A, we have that $|h_2|^2 = \beta_{ss}\beta_{s2}X_2^2$, after some manipulations, we can written Eq. (33) as

$$P_{\text{out}}^{D_2} = \Pr\left(X_2^2 < \frac{\sigma^2\gamma_{th2}}{\beta_{ss}\beta_{s2}P_t(\alpha_2 - \alpha_1\gamma_{th2})}\right), \tag{47}$$

according to subsection III-A, X_2 follows a Rayleigh random variable with scale parameter $\sqrt{\frac{N}{2}}$, thus, its squared magnitude follows an exponential distribution with rate parameter N. By using Eq. (24), we obtain the following expression for the OP of D_2 when D_1 does not operate as a relay

$$P_{\text{out}}^{D_2} = 1 - e^{\frac{-\sigma^2\gamma_{th2}}{NP_t\beta_{ss}\beta_{s2}(\alpha_2 - \alpha_1\gamma_{th2})}}. \tag{48}$$

II) $\rho \neq 0$ *(D_1 operates as a relay)*

By recalling the established definition in subsection III-A, we have that $|h_1|^2 = \beta_{ss}\beta_{s1}X_1^2$, similarly to the subsection *I)* of this Appendix, after some manipulations, we can write Eq. (33) as

$$\begin{aligned}
P_{\text{out}}^{D_2} &= \Pr\left(X_1^2 < \frac{\gamma_{th2}(|h_{SI}|^2 P_H + \sigma^2)}{\beta_{ss}\beta_{s1}P_t(1-\rho)(\alpha_2 - \alpha_1\gamma_{th2})}, \right. \\
&\qquad \left. X_2^2 < \frac{\sigma^2\gamma_{th2}}{\beta_{ss}\beta_{s2}P_t(\alpha_2 - \alpha_1\gamma_{th2})}\right) \\
&\quad + \Pr\left(X_1^2 \geq \frac{\gamma_{th2}(|h_{SI}|^2 P_H + \sigma^2)}{\beta_{ss}\beta_{s1}P_t(1-\rho)(\alpha_2 - \alpha_1\gamma_{th2})}, \right. \\
&\qquad \left. \frac{P_t X_2^2\beta_{ss}\beta_{s2}}{\sigma^2} < \frac{\gamma_{th2} - \frac{P_H}{\sigma^2}|h_{1,2}|^2}{\alpha_2 - \alpha_1(\gamma_{th2} - \frac{P_H}{\sigma^2}|h_{1,2}|^2)}\right)
\end{aligned} \tag{49}$$

We can straightforwardly observe that the first term given by $\Pr\left(X_1^2 < \frac{\gamma_{th2}(|h_{SI}|^2 P_H + \sigma^2)}{\beta_{ss}\beta_{s1}P_t(1-\rho)(\alpha_2 - \alpha_1\gamma_{th2})}\right)$ has already been derived in the Appendix B. Similarly, the second term given by $\Pr\left(X_2^2 < \frac{\sigma^2\gamma_{th2}}{\beta_{ss}\beta_{s2}P_t(\alpha_2 - \alpha_1\gamma_{th2})}\right)$, has been previously derived earlier in subsection *I)* of this Appendix.

Therefore, here, we should focus on computing the following term $\Pr\left(\frac{P_t X_2^2\beta_{ss}\beta_{s2}}{\sigma^2} < \frac{\gamma_{th2} - \frac{P_H}{\sigma^2}|h_{1,2}|^2}{\alpha_2 - \alpha_1(\gamma_{th2} - \frac{P_H}{\sigma^2}|h_{1,2}|^2)}\right)$. For easiness, let us define the following variables, $Q \triangleq \frac{P_t X_2^2\beta_{ss}\beta_{s2}}{\sigma^2}$ and $W \triangleq \frac{P_H}{\sigma^2}|h_{1,2}|^2$. Since Q is X_2 scaled by constants, its distribution is given as $Q \sim \exp\left(\frac{NPt\beta_{ss}\beta_{s2}}{\sigma^2}\right)$. Moreover, It can be observed once again that the condition $P_{th} < \beta_{12}$, leads to P_H values generally lower than $|h_{12}|^2$. In other words, the harvested power is highly likely to be lower than path-loss in D2D communication, as D_1 is typically located in close proximity to D_2. Therefore, as $|h_{12}|^2$ follows an exponential distribution, we again proposed to approximate W

as exponential random variable, $W \sim \exp\left(\frac{\widetilde{P}_H \beta_{12}}{\sigma^2}\right)$, hence we must compute the following integral

$$\begin{aligned}
&\Pr\left(Q < \frac{\gamma_{th2} - W}{\alpha_2 - \alpha_1(\gamma_{th2} - W)}\right) \\
&= \int_0^{\gamma_{th2}} F_Q\left(\frac{\gamma_{th2} - w}{\alpha_2 - \alpha_1(\gamma_{th2} - w)}\right) f_W(w) dw \\
&= F_W(\gamma_{th2}) \\
&\quad - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H} \underbrace{\int_0^{\gamma_{th2}} e^{\left(\frac{-\sigma^2}{P_t N \beta_{ss}\beta_{s2}}\right)\left(\frac{\gamma_{th2}-w}{\alpha_2 - \alpha_1(\gamma_{th2}-w)}\right) - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H} w} dw}_{I},
\end{aligned} \tag{50}$$

Since D_2 assume low rate values, and $\alpha_2 > \alpha_1$ due to implementation of NOMA, we have $\alpha_2 \gg \alpha_1 \gamma_{th2}$, thus, the integral in I can be approximated as

$$I \approx e^{\frac{-\gamma_{th2}\sigma^2}{P_t \beta_{ss}\beta_{s2} N \alpha_2}} \int_0^{\gamma_{th2}} e^{y\left(\frac{\sigma^2}{P_t N \beta_{ss}\beta_{s2}\alpha_2} - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H}\right)} dy, \tag{51}$$

whose solution is found trivially. Substituting the solution of Eq. (51) in Eq. (50) we obtain

$$\begin{aligned}
&\Pr\left(\frac{P_t X_2^2 \beta_{ss}\beta_{s2}}{\sigma^2} < \frac{\gamma_{th2} - \frac{P_H}{\sigma^2}|h_{1,2}|^2}{\alpha_2 - \alpha_1(\gamma_{th2} - \frac{P_H}{\sigma^2}|h_{1,2}|^2)}\right) \\
&\approx 1 - e^{-\frac{\gamma_{th2}\sigma^2}{\beta_{12}\widetilde{P}_H}} - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H}\left(\frac{e^{-\frac{\gamma_{th2}\sigma^2}{\beta_{12}\widetilde{P}_H}} - e^{\frac{-\gamma_{th2}\sigma^2}{P_t\beta_{ss}\beta_{s2}N\alpha_2}}}{\left(\frac{\sigma^2}{P_t N\beta_{ss}\beta_{s2}\alpha_2} - \frac{\sigma^2}{\beta_{12}\widetilde{P}_H}\right)}\right)
\end{aligned} \tag{52}$$

Hence, by utilizing Eq. (52) and the remaining analytical results derived for Eq. (49) in the previous Appendix, Eq. (34) is obtained. This completes the proof.

Appendix D
Proof of Theorem 4

We have that the ergodic rate of D_1 can be computed as

$$\begin{aligned}
R_1 &= \mathbb{E}\left[\log_2\left(1 + \frac{|h_1|^2(1-\rho)P_t\alpha_1}{|h_{SI}|^2 P_H + \sigma^2}\right)\right] \\
&\overset{(i)}{\leq} \log_2\left(\mathbb{E}\left[1 + \frac{|h_1|^2(1-\rho)P_t\alpha_1}{|h_{SI}|^2 P_H + \sigma^2}\right]\right),
\end{aligned} \tag{53}$$

where Jensen's inequality has been utilized in (i). Here will take the conditional expectation of Eq. (53) with respect to $|h_1|^2$, therefore, Eq. (53) can be written as

$$R_1 \leq \log_2\left(1 + \frac{N^2\beta_{ss}\beta_{s1}\mu_{ss}^2\mu_{s1}^2(1-\rho)P_t\alpha_1}{|h_{SI}|^2 P_H + \sigma^2}\right). \tag{54}$$

By rewriting Eq. (54) we have

$$\begin{aligned}
R_1 \leq \log_2&\left(\frac{N^2\beta_{ss}\beta_{s1}\mu_{ss}^2\mu_{s1}^2(1-\rho)P_t\alpha 1 + \sigma^2}{\widetilde{P}_H} + |h_{\mathrm{SI}}|^2\right) \\
&- \log_2\left(\frac{\sigma^2}{\widetilde{P}_H} + |h_{\mathrm{SI}}|^2\right),
\end{aligned} \tag{55}$$

defining $A \triangleq \frac{N^2\beta_{ss}\beta_{s1}\mu_{ss}^2\mu_{s1}^2(1-\rho)P_t\alpha 1 + \sigma^2}{\widetilde{P}_H}$ and $B \triangleq \frac{\sigma^2}{\widetilde{P}_H}$, we should computing the expectation with respect the variable $|h_{SI}|^2$, therefore we have that

$$\begin{aligned}
\mathbb{E}_{|h_{SI}|^2}[R_1] &\leq \int_0^\infty \log_2(A+x) f_{|h_{SI}|^2}(x) dx \\
&\quad - \int_0^\infty \log_2(B+x) f_{|h_{SI}|^2}(x) dx, \\
&= \frac{\ln\left(\frac{A}{B}\right) + e^{\frac{A}{\omega}}\Gamma\left(0, \frac{A}{\omega}\right) - e^{\frac{B}{\omega}}\Gamma\left(0, \frac{B}{\omega}\right)}{\log(2)},
\end{aligned} \tag{56}$$

It is essential to note that

$$\lim_{x\to\infty} e^x \Gamma(0,x) = \lim_{x\to\infty} \frac{\Gamma(0,x)}{\frac{1}{e^x}} \overset{L'H}{=} \frac{-e^{-x}x^{-1}}{-\frac{1}{e^x}} = \frac{1}{x} = 0, \tag{57}$$

where the L'Hospital's rule has been utilized, with [46, 06.06.20.0003.01]. Since the term $\frac{A}{\omega}$ assumes high values, we neglect it, therefore, the ergodic rate of D_1 can be given as

$$R_1 \leq \log_2\left(\frac{A}{B}\right) - \frac{e^{\frac{B}{\omega}}\Gamma\left(0, \frac{B}{\omega}\right)}{\log(2)}, \tag{58}$$

substituting the values of A and B in Eq. (58), Eq. (35) is obtained and this completes the proof.

Appendix E
Proof of Theorem 5

The Ergodic Rate of D_2 is given as

$$R_2 = \mathbb{E}\left[\log_2\left(1 + \mathrm{SINR}_{D_2,S}^{x_2} + \mathrm{SNR}_{D_2,D_1}^{x_2}\right)\right], \tag{59}$$

Let us focus firstly when $\rho = 0$, then by utilizing Eq. (11), we can rewritten Eq. (59) as

$$R_2 = \mathbb{E}\left[\log_2\left(1 + \frac{|h_2|^2 P_t\alpha_2}{|h_2|^2 P_t\alpha_1 + \sigma^2}\right)\right], \tag{60}$$

By utilizing Jensen's inequality, we can obtain an upper bound for the ergodic rate of D_2, given as

$$\begin{aligned}
R_2 &\leq \log_2\left(1 + \mathbb{E}\left[\frac{|h_2|^2 P_t\alpha_2}{|h_2|^2 P_t\alpha_1 + \sigma^2}\right]\right) \\
&\leq \log_2\left(1 + \frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1 + \sigma^2}\right).
\end{aligned} \tag{61}$$

When $\rho \neq 0$, we should consider the term $\mathrm{SNR}_{D_2,D_1}^{x_2}$ in Eq. (59). By taking the conditional expectation with respect to the random variable P_H, we obtain

$$\begin{aligned}
R_2 &\leq \mathbb{E}_{P_H}\left[\log_2\left(1 + \frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1 + \sigma^2} + \frac{P_H|h_{12}|^2}{\sigma^2}\right)\right] \\
&\leq \log_2\left(1 + \frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1 + \sigma^2} + \frac{\mathbb{E}[P_H]\,|h_{12}|^2}{\sigma^2}\right) \\
&= \log_2\left(1 + \frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1 + \sigma^2} + \frac{\widetilde{P}_H|h_{12}|^2}{\sigma^2}\right),
\end{aligned} \tag{62}$$

Now, we turn our focus on calculating the expectation of Eq. (62), with respect to the variable $|h_{12}|^2$, rewriting Eq. (62) in a conventional way, we wave

$$R_2 \leq \mathbb{E}\left[\log_2\left(\frac{\sigma^2}{\widetilde{P}_H}\left(1+\frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1+\sigma^2}\right)+|h_{12}|^2\right)+\log_2\left(\frac{\widetilde{P}_H}{\sigma^2}\right)\right], \tag{63}$$

let us define $A \triangleq \frac{\sigma^2}{\widetilde{P}_H}\left(1+\frac{N\beta_{ss}\beta_{s2}P_t\alpha_2}{N\beta_{ss}\beta_{s2}P_t\alpha_1+\sigma^2}\right)$. According to Section II, h_{12} is a complex Gaussian random variable of zero mean and variance β_{12}. Therefore, the square magnitude of h_{12}, follows an exponential distribution whose PDF is given as $f_{|h_{12}|^2} = \frac{e^{-\frac{x}{\beta_{12}}}}{\beta_{12}}$. Based on this, we should compute the following integral

$$\int_0^\infty \log_2(A+x) f_{|h_{12}|^2}(x)dx = \frac{1}{\beta_{12}}\int_0^\infty \log_2(A+x)e^{-\frac{x}{\beta_{12}}}dx$$
$$\overset{(i)}{=} \log_2(A) + \frac{e^{\frac{A}{\beta_{12}}}}{\log(2)}\Gamma\left(0,\frac{A}{\beta_{12}}\right),$$

where in (i), [45, 4.337.1] and [45, 8.359.1] have been utilized. Finally, we conclude the proof by substituting the value of A in Eq. (64) and the result into Eq. (63) and performing a few basic mathematical manipulations.

REFERENCES

[1] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, and M. Liyanage, "Survey on 6G Frontiers: Trends, Applications, Requirements, Technologies and Future Research," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 836–886, 2021.

[2] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling Massive IoT Toward 6G: A Comprehensive Survey," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 11 891–11 915, 2021.

[3] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, "Non-Orthogonal Multiple Access: Common Myths and Critical Questions," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 174–180, 2019.

[4] Z. Ding, L. Lv, F. Fang, O. A. Dobre, G. K. Karagiannidis, N. Al-Dhahir, R. Schober, and H. V. Poor, "A State-of-the-Art Survey on Reconfigurable Intelligent Surface-Assisted Non-Orthogonal Multiple Access Networks," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1358–1379, 2022.

[5] M. Zeng, W. Hao, O. A. Dobre, and Z. Ding, "Cooperative NOMA: State of the Art, Key Techniques, and Open Challenges," *IEEE Network*, vol. 34, no. 5, pp. 205–211, 2020.

[6] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[7] N. H. Mahmood, H. Alves, O. A. López, M. Shehab, D. P. M. Osorio, and M. Latva-Aho, "Six Key Features of Machine Type Communication in 6G," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[8] J. So and Y. Sung, "Improving Non-Orthogonal Multiple Access by Forming Relaying Broadcast Channels," *IEEE Communications Letters*, vol. 20, no. 9, pp. 1816–1819, 2016.

[9] Z. Ding, M. Peng, and H. V. Poor, "Cooperative Non-Orthogonal Multiple Access in 5G Systems," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1462–1465, 2015.

[10] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual Transceiver Hardware Impairments on Cooperative NOMA Networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 680–695, 2020.

[11] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Exploiting Full/Half-Duplex User Relaying in NOMA Systems," *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 560–575, 2018.

[12] Z. Zhang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Full-duplex device-to-device-aided cooperative nonorthogonal multiple access," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4467–4471, 2016.

[13] N. Ashraf, S. A. Sheikh, S. A. Khan, I. Shayea, and M. Jalal, "Simultaneous Wireless Information and Power Transfer With Cooperative Relaying for Next-Generation Wireless Networks: A Review," *IEEE Access*, vol. 9, pp. 71 482–71 504, 2021.

[14] C. Liu, L. Zhang, Z. Chen, and S. Li, "Outage Probability Analysis in Downlink SWIPT-Assisted Cooperative NOMA Systems," *Journal of Communications and Information Networks*, vol. 7, no. 1, pp. 72–87, 2022.

[15] Z. Liu, Y. Ye, G. Lu, and R. Q. Hu, "System Outage Performance of SWIPT Enabled Full-Duplex Two-Way Relaying With Residual Hardware Impairments and Self-Interference," *IEEE Systems Journal*, 2022.

[16] T.-N. Tran, T. P. Vo, P. Fazio, and M. Voznak, "SWIPT Model Adopting a PS Framework to Aid IoT Networks Inspired by the Emerging Cooperative NOMA Technique," *IEEE Access*, vol. 9, pp. 61 489–61 512, 2021.

[17] T. N. Do and B. An, "Optimal Sum-Throughput Analysis for Downlink Cooperative SWIPT NOMA systems," in *2018 2nd International Conference on Recent Advances in Signal Processing, Telecommunications Computing (SigTelCom)*, 2018, pp. 85–90.

[18] W. Wu, X. Yin, P. Deng, T. Guo, and B. Wang, "Transceiver Design for Downlink SWIPT NOMA Systems With Cooperative Full-Duplex Relaying," *IEEE Access*, vol. 7, pp. 33 464–33 472, 2019.

[19] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, and H. V. Poor, "Downlink and uplink intelligent reflecting surface aided networks: NOMA and OMA," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3988–4000, 2021.

[20] A.-T. Le, N.-D. X. Ha, D.-T. Do, A. Silva, and S. Yadav, "Enabling User Grouping and Fixed Power Allocation Scheme for Reconfigurable Intelligent Surfaces-Aided Wireless Systems," *IEEE Access*, vol. 9, pp. 92 263–92 275, 2021.

[21] X. Yue and Y. Liu, "Performance Analysis of Intelligent Reflecting Surface Assisted NOMA Networks," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2021.

[22] D. Selimis, K. P. Peppas, G. C. Alexandropoulos, and F. I. Lazarakis, "On the Performance Analysis of RIS-Empowered Communications Over Nakagami-m Fading," *IEEE Communications Letters*, vol. 25, no. 7, pp. 2191–2195, 2021.

[23] S. Li, L. Bariah, S. Muhaidat, A. Wang, and J. Liang, "Outage Analysis of NOMA-Enabled Backscatter Communications with Intelligent Reflecting Surfaces," *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[24] A. Hemanth, K. Umamaheswari, A. C. Pogaku, D.-T. Do, and B. M. Lee, "Outage Performance Analysis of Reconfigurable Intelligent Surfaces-Aided NOMA Under Presence of Hardware Impairment," *IEEE Access*, vol. 8, pp. 212 156–212 165, 2020.

[25] B. Tahir, S. Schwarz, and M. Rupp, "Analysis of Uplink IRS-Assisted NOMA Under Nakagami-m Fading via Moments Matching," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 624–628, 2021.

[26] J. Zuo, Y. Liu, and N. Al-Dhahir, "Reconfigurable Intelligent Surface Assisted Cooperative Non-orthogonal Multiple Access Systems," *IEEE Transactions on Communications*, pp. 1–1, 2021.

[27] G. Zhang, X. Gu, W. Duan, M. Wen, J. Choi, F. Gao, and P.-H. Ho, "Hybrid Time-Switching and Power-Splitting EH Relaying for RIS-NOMA Downlink," *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, 2022.

[28] Q. Liu, M. Lu, N. Li, M. Li, F. Li, and Z. Zhang, "Joint Beamforming and Power Splitting Optimization for RIS-Assited Cooperative SWIPT NOMA Systems," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 351–356.

[29] M. Elhattab, M. A. Arfaoui, C. Assi, and A. Ghrayeb, "Reconfigurable Intelligent Surface Enabled Full-duplex/half-duplex Cooperative non-orthogonal Multiple Access," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3349–3364, 2021.

[30] D. Wan, M. Wen, F. Ji, Y. Liu, and Y. Huang, "Cooperative NOMA Systems With Partial Channel State Information Over Nakagami-m Fading Channels," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 947–958, 2017.

[31] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Enable Device-to-Device Communications Underlaying Cellular Networks: Challenges and Research Aspects," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 90–96, 2014.

[32] Y. Alsaba, C. Y. Leow, and S. K. A. Rahim, "Full-duplex cooperative non-orthogonal multiple access with beamforming and energy harvesting," *IEEE Access*, vol. 6, pp. 19 726–19 738, 2018.

[33] H. Huang and M. Zhu, "Energy efficiency maximization design for full-duplex cooperative NOMA systems with SWIPT," *IEEE Access*, vol. 7, pp. 20 442–20 451, 2019.

[34] Q. Y. Liau and C. Y. Leow, "Cooperative NOMA System with Virtual Full Duplex User Relaying," *IEEE Access*, vol. 7, pp. 2502–2511, 2018.
[35] K. M. Rabie, A. Salem, E. Alsusa, and M.-S. Alouini, "Energy-harvesting in cooperative AF relaying networks over log-normal fading channels," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–7.
[36] T. N. Do, G. Kaddoum, T. L. Nguyen, D. B. da Costa, and Z. J. Haas, "Multi-RIS-aided wireless systems: Statistical characterization and performance analysis," *arXiv preprint arXiv:2104.01912*, 2021.
[37] Y. Jin, R. Guo, L. Zhou, and Z. Hu, "Secure Beamforming for IRS-Assisted Nonlinear SWIPT Systems with Full-Duplex User," *IEEE Communications Letters*, 2022.
[38] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Outage Performance of full/half-duplex User Relaying in NOMA Systems," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.
[39] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2082–2085, 2015.
[40] X. Liang, X. Gong, Y. Wu, D. W. K. Ng, and T. Hong, "Analysis of Outage Probabilities for Cooperative NOMA Users with Imperfect CSI," in *2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2018, pp. 1617–1623.
[41] B. Tahir, S. Schwarz, and M. Rupp, "Outage analysis of uplink IRS-assisted NOMA under elements splitting," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. IEEE, 2021, pp. 1–5.
[42] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. Boston: McGraw Hill, 2002.
[43] Z. Zhu, M. Ma, G. Sun, W. Hao, P. Liu, Z. Chu, and I. Lee, "Secrecy Rate Optimization in NonLinear Energy Harvesting Model-Based mmWave IoT Systems With SWIPT," *IEEE Systems Journal*, pp. 1–11, 2022.
[44] S. Gao, K. Xiong, R. Jiang, L. Zhou, and H. Tang, "Outage Performance of Wireless-Powered SWIPT Networks with Non-linear EH Model in Nakagami-m Fading," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, 2018, pp. 668–671.
[45] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic press, 2014.
[46] The Wolfram Functions Site. [Online]. Available: https://functions.wolfram.com

APPENDIX B – Energy Efficiency Maximization for Intelligent Surfaces Aided Massive MIMO with Zero Forcing

Energy Efficiency Maximization for Intelligent Surfaces Aided Massive MIMO with Zero Forcing. Full paper. *Submitted to IEEE Transactions on Green Communications and Networking on Jun. 2023* (IF: 4.8, Eng. IV Qualis-CAPES: A3)

Energy Efficiency Maximization for Intelligent Surfaces Aided Massive MIMO with Zero Forcing

Wilson de Souza Junior and Taufik Abrão

***Abstract*—In this work, we address the energy efficiency (EE) maximization problem in a downlink communication system utilizing reconfigurable intelligent surfaces (RIS) in a multi-user massive multiple-input multiple-output (mMIMO) setup with Zero-Forcing (ZF) precoding. The channel between the BS and RIS operates under a Rician fading with Rician factor K_1. Since systematically optimizing the RIS angle phase shift in each channel coherence time interval is challenging and burdensome, we employ the statistical CSI-based optimization strategy to alleviate this overhead. By treating the RIS phase shift matrix as a constant over multiple channel coherence time intervals, we can reduce the computational complexity while maintaining an interesting performance. Based on an ergodic rate (ER) lower bound closed-form, the EE optimization problem is formulated. Such a problem is non-convex and challenging to tackle due to the coupled variables. To circumvent such an obstacle, we explore the sequential optimization approach where the power allocation vector $\mathbf{p}$, the number of antennas M, and the RIS angle phase shift $\mathbf{v}$ are separate and sequentially solved iteratively until convergence. With the help of the Lagrangian dual method and fractional programming techniques, insightful closed-form expressions for all the optimization variables are derived. Simulation results validate the effectiveness of the proposed method across different generalized channel scenarios, including NLoS ($K_1 = 0$) and partially LoS ($K_1 \neq 0$) conditions. Our approach offers significant performance gains once it can reduce the power consumption and the number of active antennas t BS, demonstrating the potential of RIS deployment and optimization in mMIMO systems with ZF.**

***Index Terms*—statistical CSI, massive multiple-input multiple-output (mMIMO), Reconfigurable intelligent surfaces (RIS), energy-efficiency (EE), zero-forcing (ZF), Lagrangian.**

I. Introduction

Reconfigurable intelligent surfaces (RIS) is one of the most promising recent techniques which has the potential to integrate future communications systems (beyond fifth generation, 6G, etc.) due to its potential to improve the transmission by creating a "virtual" reconfigurable communication link [1]. Specifically, the RIS is composed of several individually scattering elements which are artificial meta-material structures that can reflect incident electromagnetic waves and can be configured to increase the signal level in a specific direction [2].

On the other hand, to enhance the spectral efficiency (SE) and to serve a plurality of users using the same physical layer resources, massive multiple-input multiple-output (mMIMO) is paramount; however, this technology resorts to hundreds of antennas operating at the base station (BS) [3], which can significantly increase the consumption.

RIS and mMIMO are promising for future systems. Since the integration of these technologies is expected to be present in the emerging wireless networks, which impose highly demanding data rate requirements and a massive number of connected users [4], [5], many serious concerns about energy consumption have been raised. Consequently, bit-per-Joule efficiency is a key performance indicator to ensure a green and sustainable communication system.

An extensive number of works such as in [6]–[13] have studied the RIS deployment impact in mMIMO systems. In [6], the authors derived a tight closed-form expression for the ergodic rate (ER) of a RIS-aided mMIMO system with zero-forcing (ZF) precoding over Rician fading. In [7], the authors exploit the long-term statistical CSI to optimize the phase-shift matrix of RIS aiming to maximize the achievable rate with maximum ratio combiner (MRC), where a genetic algorithm is deployed. In [9], the authors investigated the EE problem in downlink RIS-aided mMIMO systems by proposing two algorithms based on alternating maximization, gradient descent, and sequential fractional programming. In [10], the EE problem in uplink RIS mMIMO is approached, where the authors proposed a joint optimization of the resources through the block coordinate descent algorithm. A resembling setup is defined in [11], and the power minimization problem subject to the QoS constraint is tackled by an alternating optimization-based method.

All the works mentioned above, [9]–[12], [14] concerning the EE problem require the transmitter to access accurate instantaneous channel state information (CSI), to optimize the phase-shift matrix of the RIS. In practice, acquiring accurate instantaneous CSI in RIS-aided mMIMO is challenging. Furthermore, the instantaneous channel estimation process can be impracticable in scenarios with high mobility and short channel coherence time. Hence, the long-term statistical CSI-based optimization strategy could be more practical and feasible [7]. To corroborate the necessity and usefulness of the statistical CSI-based channel estimation approach, Fig. 1 reveals the excessive time overhead by adopting instantaneous CSI estimates compared to statistical CSI-based estimation transmission schemes. In the latter methods, the statistical channel estimation is deployed for DL and UL data transmission across different channel coherence times, reducing the overhead and the signaling, so being practical in the challenging RIS-aided mMIMO system scenarios.

To our best knowledge, the EE optimization problem with statistical CSI under Rician fading for RIS-aided mMIMO

This work was supported by the National Council for Scientific and Technological Development (CNPq) of Brazil, Grant: 310681/2019-7, by CAPES, Grant: FC001, and by Londrina State University (UEL), Brazil.

W. Junior and T. Abrão are with the Department of Electrical Engineering (DEEL). State University of Londrina (UEL). Po.Box 10.011, CEP:86057-970, Londrina, PR, Brazil. Email: wilsoonjr98@gmail.com; taufik@uel.br

has not been investigated yet. Thus, motivated by the lack of robust and effective EE optimization methods for recent RIS mMIMO systems, in our work, different of [9] and [10], the EE optimization problem is approached in RIS-aided mMIMO with statistical CSI operating in generalized Rician fading channels. Moreover, we also extended the adopted problem in [15] for more challenging generalized channel and system scenarios, deriving closed-form expressions for three optimization variables, *i.e.,* the number of BS antennas (M), the RIS angles phase-shift ($\mathbf{v}$), and the transmit power allocation ($\mathbf{p}$).

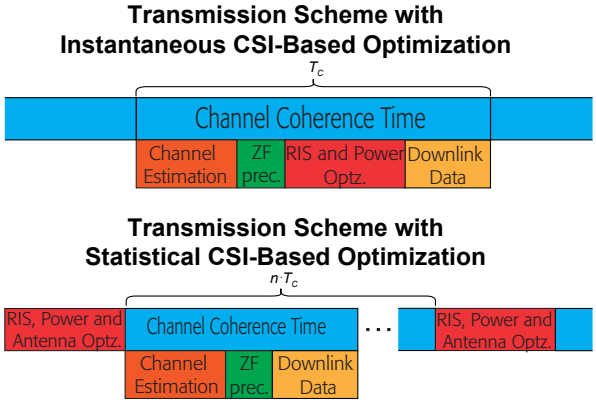


Fig. 1. Illustration of the advantages by adopting Statistical CSI-based parameters optimization, where the RIS phase-shift vector ($\mathbf{v}$), transmit power allocation ($\mathbf{p}$), and the number of BS antennas (M) remain fixed for the users over multiple channel coherence times (nT_c); differently, for the Instantaneous CSI-based parameter optimization approach, where the channel estimation, precoding, RIS phase shift, and transmit power must be updated at each T_c

The main ***contributions*** of this work are fourfold:

- Different from the previously works [9], [10], [16], we proposed an effective, low-complexity solution for the EE maximization problem subject to a minimum user rate in RIS-aided mMIMO systems with ZF, considering different LoS and NLoS channels, and by relying only on the statistical CSI estimates, which can be beneficial in high mobility and short channel coherence time wireless scenarios; indeed, such an approach does not need precise instantaneous CSI estimates in view of optimizing the RIS angles phase shift and power allocation.
- Based on a tight closed-form expression of the ER of a single cell RIS-aided mMIMO system, we formulate a joint optimization problem concerning the users transmit power $\mathbf{p}$, the number of active BS antennas M, and RIS phase shift angles $\mathbf{v}$ variables under the constraints of overall power budget, ZF precoding feasibility, quality of service (QoS), and unity module requirements.
- To deal with this problem, we apply the Lagrangian dual method with some fractional programming techniques in order to obtain closed-form solutions for all optimization variables. The complexity of the proposed antenna (M), power allocation ($\mathbf{p}$), and RIS phase shift ($\mathbf{v}$) optimization algorithm has been accurately investigated.
- Extensive numerical results corroborate the effectiveness and efficiency of the proposed analytical EE optimization method for RIS-aided mMIMO systems with ZF operating under generalized Rician channels.

The remainder of this paper is organized as follows. Section II describes the RIS-aided mMIMO system model. The optimization problem and the overall proposed analytical optimization methods and algorithms are developed and discussed in Section III. We present numerical results corroborating our findings in Section IV. Final remarks are offered in Section V.

Notations: Scalars, vectors, and matrices are denoted by the lower-case, bold-face lower-case, and bold-face upper-case letters, respectively; $\mathbb{C}^{A\times B}$ denotes the space of A rows and B columns complex matrices; $|\cdot|$ denote the absolute operator; $\mathrm{diag}(\cdot)$ denotes the diagonal operator; $\mathrm{Tr}(\cdot)$ is trace operator; $[x]^+$ denotes $\max(x,0)$; $[x]_b^a$ denotes $\min(\max(x,b),a)$; $[\cdot]^*$ denote the conjugate operator; $\angle[\cdot]$ denotes the phase of the complex argument; $j \triangleq \sqrt{-1}$; $\mathbb{E}[\cdot]$ is the expetation operator; $[\cdot]^T$ is the transpose operator; $[\cdot]^H$ is the conjugate transpose (Hermitian) operator; $\mathbf{I}_K$ stands for the $K\times K$ identity matrix; $\mathbf{X}^{-1}$ is the inverse of X; $\det(\cdot)$ is the determinant of a given matrix; $\mathrm{Re}(\cdot)$ is the real part of a complex value; $\mathcal{CN}(\mu,\sigma^2)$ defines a complex-valued Gaussian random variable with mean μ and variance σ^2; $x \sim \mathcal{U}[a,b]$ defines a uniform random variable from range of a to b; $\lceil\cdot\rceil$ denotes the ceiling function; $[\mathbf{X}]_{k,l}$ stands for the entry at k-th row and l-th column of matrix $\mathbf{X}$; $[\mathbf{X}]_{k,:}$ denotes the k-th row vector of matrix $\mathbf{X}$; $[\mathbf{X}]_{:,k}$ denotes the k-th column vector of matrix $\mathbf{X}$; $[\mathbf{X}]_{a:b,c:d}$ represents the sub-matrix of $\mathbf{X}$ consisting of rows from a to b and columns from c to d; The notation $\mathcal{O}(\cdot)$ denotes the computational complexity.

II. System Model

We consider a single-cell multiuser RIS-aided mMIMO system operating in downlink mode, where the BS is equipped with a uniform linear array (ULA) with $M_{\max}$ antennas, where M antennas will be activated in order to serve K single-antenna users (UEs). The single RIS panel is equipped with N elements and deployed to improve the network communication's reliability and coverage. Fig. 2 illustrates the general RIS mMIMO system and channel scenarios investigated.

Let us denote $\mathbf{D} \in \mathbb{C}^{M\times K}$ as the direct link communication channel between the UEs and the BS, given by

$$\mathbf{D} = \tilde{\mathbf{D}}\boldsymbol{\alpha}_D^{1/2}, \tag{1}$$

where $\boldsymbol{\alpha}_D = \mathrm{diag}([\alpha_{D,1},\ldots,\alpha_{D,K}])$ and $\alpha_{D,k}$ denotes the large-scale fading term between the BS and the k-th user, the matrix $\tilde{\mathbf{D}} \in \mathbb{C}^{M\times K}$ is composed by independent and identically distributed (i.i.d) complex Gaussian random variables, whose mean is zero and variance is unit, *i.e.* $d_{m,k} \sim \mathcal{CN}(0,1)$, $\forall k = 1,\ldots,K$ and $m = 1,\ldots,M$.

We define the channel between the RIS and the BS as $\mathbf{G} \in \mathbb{C}^{M\times N}$ and the channel between the users and the RIS as $\mathbf{F} \in \mathbb{C}^{N\times K}$. The channel matrix $\mathbf{G}$ is assumed to be a Rician fading channel and can be given as [6]

$$\mathbf{G} = \sqrt{\alpha_G \frac{1}{K_1+1}}\mathbf{G}_{\mathrm{NLoS}} + \sqrt{\alpha_G \frac{K_1}{K_1+1}}\mathbf{G}_{\mathrm{LoS}}, \tag{2}$$

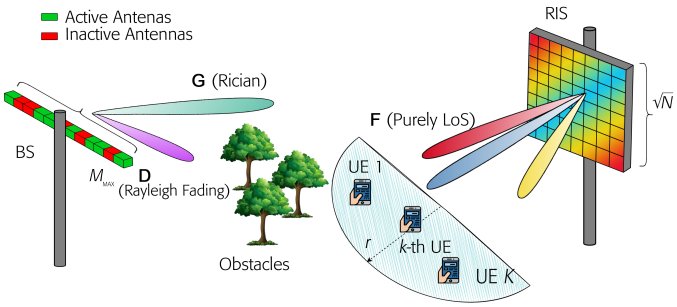


Fig. 2. RIS-aided mMIMO multi-user investigated scenarios. Particularly, this setup is illustrated with $N = 64$ reflective elements at RIS and $M_{\max} = 12$ BS antennas; in this specific configuration, there are $M = 7$ actives elements at BS.

where α_G is the large-scale fading (path loss) between the RIS and the BS and K_1 denotes the Rician factor. Besides, the fading channel matrix $\mathbf{F}$ is assumed to be purely Line of Sight (LoS) and expressed as

$$\mathbf{F} = \mathbf{F}_{\text{LoS}}\boldsymbol{\alpha}_F^{1/2} = [\sqrt{\alpha_{F,1}}\mathbf{f}_{\text{LoS}}^1, \ldots, \sqrt{\alpha_{F,K}}\mathbf{f}_{\text{LoS}}^K], \quad (3)$$

where $\boldsymbol{\alpha}_F = \text{diag}([\alpha_{F,1}, \ldots, \alpha_{F,K}])$ and $\alpha_{F,k}$ denotes the large-scale fading factor between the k-th user and the BS. Besides, the LoS channel for RIS adopted herein is based on the two-dimensional uniform squared planar array (USPA) and given as [17]

$$\mathbf{a}_N(\varphi,\phi) = \Big[1, \ldots, e^{j2\pi\frac{d}{\lambda}(x\sin\phi\sin\varphi + y\cos\phi)}, \ldots, e^{j2\pi\frac{d}{\lambda}((X-1)\sin\phi\sin\varphi + (Y-1)\cos\phi)}\Big]^T, \quad (4)$$

where $0 \le x, y < X$ with $X = Y \in \mathbb{Z}_+$ being the square side of the USPA [18], [19], d denotes the element spacing, λ denotes the wavelength, φ and ϕ are azimuth and elevation angles respectively. Let us denote $\varphi_{\text{AoD}}^{bs}$ the angle of departure (AoD) from the BS towards the RIS, $\varphi_{\text{AoD}}^{r,k}$ and $\phi_{\text{AoD}}^{r,k}$ the AoD from RIS towards the k-th user and φ_{AoA}^{r} and ϕ_{AoA}^{r} the angles of arrival (AoA) at the RIS from the BS respectively. Moreover, the LoS channel for BS is based on the Uniform Linear Array (ULA), given by [20]

$$\mathbf{a}_M(\varphi) = \Big[1, \ldots, e^{j2\pi\frac{d}{\lambda}x\sin\varphi}, e^{j2\pi\frac{d}{\lambda}((X-1)\sin\varphi)}\Big]^T. \quad (5)$$

Under this condition, the LoS channel matrix $\mathbf{G}_{\text{LoS}}$ is expressed as

$$\mathbf{G}_{\text{LoS}} = \mathbf{a}_M(\varphi_{\text{AoD}}^{bs})\mathbf{a}_N(\varphi_{\text{AoA}}^{r}, \phi_{\text{AoA}}^{r}) \triangleq \mathbf{a}_M\mathbf{a}_N^H, \quad (6)$$

and the LoS component $\mathbf{f}_{\text{LoS}}^k$ is

$$\mathbf{f}_{\text{LoS}}^k = \mathbf{a}_N(\varphi_{\text{AoD}}^{r,k}, \phi_{\text{AoD}}^{r,k}). \quad (7)$$

The angle phase shift matrix of the RIS can be defined as $\boldsymbol{\Phi} = \text{diag}(\mathbf{v})$, where $\mathbf{v} = [e^{-j\theta_1}, \ldots, e^{-j\theta_N}]^H$ and θ_n is the phase shift of the n-th RIS element. In this way, the cascaded channel can be written as $\mathbf{H} = \mathbf{D} + \mathbf{G}\boldsymbol{\Phi}\mathbf{F}$.

The received signal y_k at the k-th UE is

$$y_k = \sqrt{p_k}\mathbf{h}_k^H\mathbf{w}_k s_k + \sum_{j=1, j\neq k}^{K}\sqrt{p_j}\mathbf{h}_k^H\mathbf{w}_j s_j + n_k, \quad (8)$$

where $\mathbf{h}_k$ and $\mathbf{w}_k$ is the k-th column of $\mathbf{H}$ and $\mathbf{W}$ respectively with $\mathbf{W}$ being the precoding matrix, n_k is the i.i.d. additive white Gaussian noise with zero mean and unit variance, $n_k \sim \mathcal{CN}(0,1)$, $p_k \in \mathbf{p}$ is the downlink transmit power for the k-th user, where $\mathbf{p} = [p_1, \ldots, p_K]^T$ denotes the set of transmit powers, and s_k is the transmitted information symbol of the k-th user with $\mathbb{E}[|s_k|^2] = 1$. In this work, we adopt the linear ZF precoding, given by the Moore–Penrose pseudoinverse $\mathbf{W} = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$.

III. Energy Efficiency Optimization Problem Formulation with Statistical CSI

Since channels might be fast time-varying in practical wireless communications scenarios, the $\boldsymbol{\Phi}$ matrix optimization could be difficult to be implemented every channel coherence time since accurate CSI should be available; thus, utilizing the statistical CSI to optimize the EE in RIS-aided mMIMO can be a cost-efficient alternative way. Therefore, we formulate and investigate the EE optimization problem in this section by exploiting the statistical CSI estimates. For this reason, we should adopt the ergodic achievable rate with ZF precoder in RIS-aided mMIMO systems, which can be written as [21, Eq. (18)]:

$$R_k = \mathbb{E}\left\{\log_2\left(1 + \frac{p_k}{\sigma^2\left[(\mathbf{H}^H\mathbf{H})^{-1}\right]_{k,k}}\right)\right\}. \quad (9)$$

In [6], Kangda Zhi, *et. al.*, derived a closed-form lower bound expression for the ergodic rate as given in Eq. (9), in a RIS-aided massive MIMO system with ZF. This expression can be represented as [6, Eq. (22)]

$$R_k \ge \widetilde{R}_k = \log_2\left(1 + \frac{p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right), \quad (10)$$

where the constants $\mathbf{B}$ and $\mathbf{A}_k$ $\forall k \in \{1, \ldots, K\}$ are given as

$$\mathbf{B} \triangleq \frac{1}{N}\mathbf{I}_N + \alpha_G K_1\,\text{diag}(\mathbf{a}_N^H)\mathbf{F}\boldsymbol{\Lambda}^{-1}\mathbf{F}^H\text{diag}(\mathbf{a}_N) \quad (11)$$

$$\mathbf{A}_k \triangleq \left[\boldsymbol{\Lambda}^{-1}\right]_{k,k}\mathbf{B} - K_1\alpha_G\mathbf{s}_k\mathbf{s}_k^H, \quad (12)$$

where the variables $\boldsymbol{\Lambda} \triangleq \alpha_G\mathbf{F}^H\mathbf{F} + (K_1+1)\text{diag}(\boldsymbol{\alpha}_D)$, and $\mathbf{s}_k^H \triangleq \left[\boldsymbol{\Lambda}^{-1}\mathbf{F}^H\text{diag}(\mathbf{a}_N)\right]_{k,:}$ are computed through of the path-loss and LoS component

Thus, the *maximum energy efficiency optimization* problem regarding the interest variables is described as follows.

$$\mathcal{P}_1: \underset{M,\mathbf{p},\mathbf{v}}{\text{maximize}}\ \eta_{\text{EE}} = \frac{\sum_{k=1}^{K}\log_2\left(1 + \frac{p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)}{\varrho\sum_{k=1}^{K}p_k + P_{\text{FIX}} + MP_{\text{BS}} + NP_{\text{RIS}}} \quad (13)$$

$$\text{subject to} \quad \sum_{k=1}^{K}p_k \le P_{\text{TX}}, \quad (13a)$$

$$\widetilde{R}_k \ge R_{\min}, \quad \forall k = 1, \ldots, K, \quad (13b)$$

$$M > K, \quad (13c)$$

$$|e^{j\theta_n}| = 1, \quad \forall n = 1, \ldots, N, \quad (13d)$$

where ϱ is the power amplifier (PA) inefficiency; P_{BS} is the constant radio frequency (RF) chain circuit power consumption per transmit antenna; P_{RIS} is a fixed power value necessary to activate one reflecting element on RIS; P_{FIX} is the fixed consumed power at the BS. Constraint (13a) is the total RF transmit power constraint at BS with P_{TX} being the *power budget* available at BS, $i.e.$, the maximum power available; constraint (13b) guarantees the minimum achievable downlink rate for all UEs, where $R_{\min}$ is a pre-defined parameter which depends on QoS requirements; constraint (13c) guarantees the applicability of the ZF in the ER lower bound equation [1], and constraint (13d) accounts for the fact that each RIS reflecting element can only redirect, without amplifying the incoming signal, $i.e.$, the RIS is assumed to operate exclusively only in passive mode.

We can see that the objective function of the problem given in Eq. (13) is a concave-linear fraction whose numerator and denominator are concave and linear functions w.r.t. the number of active antennas M and power allocation $\mathbf{p}$ respectively. Furthermore, one can notice that the constraint (13b) is a non-linear constraint; and the constraint (13d) is a non-convex constraint; thus, the formulated problem in (13) is an intricate non-convex NLP, hence, difficult to solve.

Due to the coupling of variables, it is hard to obtain the globally optimal solution of (13). Therefore, we proposed an iterative procedure to solve problem $\mathcal{P}_1$ sub-optimally and with low complexity. For this purpose, we leverage the classical *alternating optimization* strategy, which optimizes one variable while the others are kept fixed. The proposed alternating algorithm contains basically two fundamental steps, which are run sequentially. In the first step, we optimize the RIS phase shift $\mathbf{v}$ for a fixed number of active antennas M and power allocation $\mathbf{p}$; in the second step, we optimize the number of active antennas$M$ and power allocation $\mathbf{p}$ with given RIS phase shift $\mathbf{v}$.

A. Phase Optimization

In this subsection, we deal with the RIS phase shift optimization problem, where we design a new low-complexity algorithm based on an analytical approach. Specifically, we first apply the *Lagrangian Dual Transform* procedure [22], [23] to translate our sum-of-logarithm original problem into an equivalent sum-of-ratios problem. In addition, we utilized the strategy to deal with the sum-of-ratio proposed in [24] to reach a closed-form expression for $\theta_n\ \forall n \in \{1,\dots,N\}$.

1) Lagrangian Dual Transform: From the objective function of (13), we observe that when $\mathbf{p}$ and M are fixed, the EE maximization is equivalent to the sum-rate maximization; therefore, the optimal $\mathbf{v}$ is the one that maximizes the numerator of (13). One can notice that due to the sum-of-logarithm, the closed-form expression is challenging or even impracticable to obtain directly; to address this issue, we initially employed the *Lagrangian Dual Transform* proposed in [22] to decouple the argument of the logarithm function argument, enabling us to derive the equivalent optimization problem, given as:

$$\begin{aligned}\mathcal{P}_2:\ \underset{\mathbf{v},\boldsymbol{\gamma}}{\text{maximize}}\quad & \sum_{k=1}^{K}\Bigg(\log_2(1+\gamma_k)-\gamma_k \\ & +\frac{(1+\gamma_k)p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\left(\sigma^2(K_1+1)\mathbf{A}_k+p_k(M-K)\mathbf{B}\right)\mathbf{v}}\Bigg), \\ \text{subject to}\quad & \gamma_k\geq 0,\quad \forall k=1,\dots,K, \\ & |e^{j\theta_n}|=1,\quad \forall n=1,\dots,N.\end{aligned} \tag{14}$$

where $\boldsymbol{\gamma}=[\gamma_1,\dots,\gamma_K]^T$ is an auxiliary variable introduced for each argument of the logarithm function $\frac{p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}$ in the numerator of Eq. (13). Here, the idea is optimize $\boldsymbol{\gamma}$ and $\mathbf{v}$ in an alternating, interative way, $i.e.$, we optimize $\boldsymbol{\gamma}$ for given $\mathbf{v}$, while $\mathbf{v}$ is updated for values of $\boldsymbol{\gamma}$ at their last updated. The optimal values of $\boldsymbol{\gamma}$ can be straightly obtained in closed-form according to the Karush-Kuhn-Tucker (KKT) conditions applied to Eq. (14), resulting:

$$\overset{\star}{\gamma}_k=\left[\left(\frac{\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}+p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\log(2)\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right)-1\right]^{+}. \tag{15}$$

Given $\overset{\star}{\boldsymbol{\gamma}}$, one may focus on the following fractional problem

$$\max_{\mathbf{v}}\quad \sum_{k=1}^{K}\frac{(1+\overset{\star}{\gamma}_k)p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\left(\sigma^2(K_1+1)\mathbf{A}_k+p_k(M-K)\mathbf{B}\right)\mathbf{v}}. \tag{16}$$

2) Fractional Problem Sum-of-Ratio: Since the problem in (16) is a sum-of-ratios, classical transforms for fractional programming problems, as *Charnes-Cooper Transform* and *Dinkelbach's Transform* can not be generalized to multiple-ratio problems [25]. In addition, although the *Quadract Transform* proposed in [25] works well in many multiple-ratio problems, specifically, in the phase-shift angles optimization problem, it can impose difficulty to obtain a closed-form expression since it introduces a square root in the optimization variable. In order to circumvent this problem, herein we adopt the strategy proposed in [24]. The key idea is introducing auxiliary variables $\boldsymbol{\beta}=[\beta_1,\dots,\beta_K]^T$, and the problem represented in Eq. (16) is equivalently translated to

$$\begin{aligned}\mathcal{P}_3:\underset{\mathbf{v},\boldsymbol{\beta}}{\text{maximize}}\ & \sum_{k=1}^{K}u_k\Big((1+\gamma_k^{\star})p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v} \\ & -\beta_k\mathbf{v}^H\left(\sigma^2(K_1+1)\mathbf{A}_k+p_k(M-K)\mathbf{B}\right)\mathbf{v}\Big) \\ \text{subject to}\quad & \beta_k\geq 0,\quad \forall k=1,\dots,K, \\ & |e^{j\theta_n}|=1,\quad \forall n=1,\dots,N.\end{aligned} \tag{17}$$

where

$$u_k=\frac{1}{\mathbf{v}^H\left(\sigma^2(K_1+1)\mathbf{A}_k+p_k(M-K)\mathbf{B}\right)\mathbf{v}}, \tag{18}$$

is defined as the denominator of the objective function given in Eq. (16), $\forall k=1,\dots,K$, and is set constant being updated in each interaction after $\mathbf{v}$ is optimized. Besides, according to [24] the optimal values of $\boldsymbol{\beta}$ can be obtained as:

[1]In the practice, it is important to note that $M\geq K$ for ZF feasibility, however, when considering the ER defined in (10), the number of antennas should exceed than the number of users, $i.e.$, $M>K$, as the rate becomes null when the equality condition is met.

$$\overset{\star}{\beta}_k = \frac{(1+\overset{\star}{\gamma}_k)p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\mathbf{v}^H\left(\sigma^2(K_1+1)\mathbf{A}_k + p_k(M-K)\mathbf{B}\right)\mathbf{v}}. \tag{19}$$

Notice that for given $\overset{\star}{\boldsymbol{\beta}}$, we can rewrite Eq. (17) as

$$\begin{aligned}\mathcal{P}_3: \quad &\underset{\mathbf{v}}{\text{maximize}} \quad \mathbf{v}^H\mathbf{C}\mathbf{v}, \\ &\text{subject to} \quad |e^{j\theta_n}| = 1, \quad \forall n = 1,\ldots,N\end{aligned} \tag{20}$$

where $\mathbf{C} \in \mathbb{C}^{N\times N}$ is a Hermitian matrix defined as

$$\begin{aligned}\mathbf{C} = &\sum_{k=1}^{K} u_k \Big((1+\overset{\star}{\gamma}_k)p_k(M-K)\mathbf{B} \\ &-\overset{\star}{\beta}_k\left(\sigma^2(K_1+1)\mathbf{A}_k + p_k(M-K)\mathbf{B}\right)\Big).\end{aligned} \tag{21}$$

The problem $\mathcal{P}_3$ is known in the literature, and initially, the first proposed solution has been the *Semidefinite Relation* (SDR) [26]. However, due to the high complexity for the cases where $\text{rank}\,(\mathbf{v}\mathbf{v}^H) \neq 1$, this solution can not be adequate. Furthermore, many authors also adopt the *Sequential Fractional Programming* (SFP) method, also known as the *Majorization-Minimization* strategy. Herein, we proposed an alternative way to optimize $\mathcal{P}_3$ with low complexity and based on closed-form expression, according to the following Lemma.

Lemma 1. *Let the optimization problem given in Eq.* (20)*, when $v_\ell = e^{j\theta_\ell}$ is fixed, $\forall \ell \in \{1,\ldots,N\} \setminus n$, the optimal solution of $[\overset{\star}{\mathbf{v}}]_n$ can be given in closed-form expression by*

$$\begin{aligned}\overset{\star}{\theta}_n = \angle\Bigg[&\text{Tr}\left(\left([\mathbf{C}]_{1:(n-1),n}\right)^H[\mathbf{v}]_{1:(n-1)}\right) \\ &+ \text{Tr}\left([\mathbf{v}]_{(n+1):N}[\mathbf{C}]_{n,(n+1):N}\right)\Bigg], \\ &\forall n \in \{1,\ldots,N\}\end{aligned} \tag{23}$$

Proof. The proof is Available in Appendix A. ■

The proposed EE maximization algorithm w.r.t. RIS phase-shift, based on statistical-CSI for RIS-aided mMIMO, is summarized in Algorithm 1. The RIS phase-shift $\mathbf{v}$, is optimized when the user's power allocation $\mathbf{p}$ and the number of transmit antennas at BS M are fixed. Additionally, we present two distinct solutions for $\mathcal{P}_3$. The first solution is based on the SFP methodology, thoroughly explained in subsection III.A.2 of [9]. The second solution is the optimal proposed analytical solution, given by (23). For the SFP solution, $\lambda_{\min}$ represents the minimum eigenvalue of matrix $\mathbf{C}$.

Algorithm 1 RIS Phase-shift Optimization

Input: M, $\mathbf{p}$, $\mathbf{v}$, σ^2, K_1, $\mathbf{B}$, $\mathbf{A}_k$, $\forall k \in \{1,\ldots,K\}$
repeat
  Step 1: Update $\boldsymbol{\gamma}$ by (15);
  repeat
    Step 2: Compute $\mathbf{u}$ by (18);
    Step 3: Update $\boldsymbol{\beta}$ by (19);
    Step 4: Compute $\mathbf{C}$ by (21);
    Set $\ell = 1$;
    repeat
      Step 5:
      1) SFP Approach:
        Set $\lambda_{\min}$ as the minimum eigenvalue of $\mathbf{C}$;
        $\mathbf{v}^{(\ell)} = e^{j\angle(\lambda_{\min}\mathbf{I}_N - \mathbf{C})\mathbf{v}^{(\ell-1)}}$;
      2) Analytical Approach:
        $[\mathbf{v}]_n^{(\ell)} = e^{j\overset{\star}{\theta}_n}$ based on (23), $n = 1,\ldots,N$;
    until $||\mathbf{v}^{(\ell)} - \mathbf{v}^{(\ell-1)}|| < \epsilon$
  until the objective function in (16) converge
until the objective function in (14) converge
Output: $\overset{\star}{\mathbf{v}} = [e^{-j\overset{\star}{\theta}_1},\ldots,e^{-j\overset{\star}{\theta}_N}]^H$

B. Active Antennas and Power Allocation Optimization

The formulated original problem $\mathcal{P}_1$ is a fractional programming problem, whereas Dinkelbach's transform is a classical method that can deal with this kind of problem. By invoking the Dilkelbach's transform, problem $\mathcal{P}_1$ can be transformed to

$$\begin{aligned}\mathcal{P}_4: \underset{M,\mathbf{p}}{\text{maximize}} &\sum_{k=1}^{K}\log_2\left(1+\frac{p_k(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}{\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}\right) \\ &-\eta_{\text{EE}}\left(\varrho\sum_{k=1}^{K}p_k + P_{\text{FIX}} + MP_{\text{BS}} + NP_{\text{RIS}}\right). \\ \text{subject to} \quad &\text{(13a), (13b).}\end{aligned} \tag{24}$$

Notice that the existence of logarithmic functions imposes a serious difficulty deriving a closed-form solution for M. Intending to propose a low complexity algorithm for optimization variable M, we proceed similarly to [27]. Hence, by recalling Eq. (10) and after some mathematical manipulations we can define

$$p_k = \frac{(2^{\widetilde{R}_k}-1)\sigma^2(K_1+1)\mathbf{v}\mathbf{A}_k\mathbf{v}}{(M-K)\mathbf{v}\mathbf{B}\mathbf{v}} \tag{25}$$

Substituting (10) and (25) into $\mathcal{P}_4$ we can further convert

$$\begin{aligned}\mathcal{L}\left(M,\widetilde{\mathbf{R}},\boldsymbol{\mu},\vartheta\right) = &\sum_{k=1}^{K}\widetilde{R}_k - \eta_{\text{EE}}\left(\varrho\sum_{k=1}^{K}\frac{(2^{\widetilde{R}_k}-1)\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}{(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}} + P_{\text{FIX}} + MP_{\text{BS}} + NP_{\text{RIS}}\right) + \\ &+\sum_{k=1}^{K}\mu_k\left(\widetilde{R}_k - R_{\min}\right) + \vartheta\left(P_{\text{TX}} - \sum_{k=1}^{K}\frac{(2^{\widetilde{R}_k}-1)\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}{(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}\right).\end{aligned} \tag{22}$$

$$\overset{\star}{M} = \left[\frac{\sum_{k=1}^{K}(\mu_k+1) + \sqrt{\left(\sum_{k=1}^{K}(\mu_k+1)\right)^2 - 4\eta_{\mathrm{EE}}P_{\mathrm{BS}}\log(2)^2\sigma^2(K_1+1)(\eta_{\mathrm{EE}}\varrho+\vartheta)\sum_{k=1}^{K}\frac{\mathbf{v}\mathbf{A}_k\mathbf{v}}{\mathbf{v}\mathbf{B}\mathbf{v}}}}{2\eta_{\mathrm{EE}}P_{\mathrm{BS}}\log(2)} + K \right]_{K+1}^{M_{\max}}, \tag{31}$$

this problem to the following equivalent problem:

$$\mathcal{P}_5 : \underset{M,\widetilde{\mathbf{R}}}{\text{maximize}} \sum_{k=1}^{K}\widetilde{R}_k - \eta_{\mathrm{EE}}\left(P_{\mathrm{FIX}} + MP_{\mathrm{BS}} + NP_{\mathrm{RIS}} \right.$$
$$\left. \varrho\sum_{k=1}^{K}\frac{(2^{\widetilde{R}_k}-1)\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}{(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}}\right). \tag{26}$$
$$\text{subject to} \sum_{k=1}^{K}\frac{(2^{\widetilde{R}_k}-1)\sigma^2(K_1+1)\mathbf{v}^H\mathbf{A}_k\mathbf{v}}{(M-K)\mathbf{v}^H\mathbf{B}\mathbf{v}} \leq P_{\mathrm{TX}}, \tag{26a}$$
$$\widetilde{R}_k \geq R_{\min}, \quad \forall k = 1,\ldots,K \tag{13b}$$

Since $\mathbf{v}^H\mathbf{A}_k\mathbf{v} > 0$ and $\mathbf{v}^H\mathbf{B}\mathbf{v} > 0$, $\mathcal{P}_5$ is convex, one can utilize the Lagrangian Dual method [28] to solve this problem and reach near-optimal low-complexity solution regarding the original problem $\mathcal{P}_1$.

The Lagrangian Dual problem can be modeled by [28], [29]

$$\underset{\boldsymbol{\mu},\vartheta}{\text{minimize}} \quad \underset{M,\widetilde{\mathbf{R}}}{\text{maximize}} \quad \mathcal{L}\left(M,\widetilde{\mathbf{R}},\boldsymbol{\mu},\vartheta\right), \tag{27}$$
$$\text{subject to} \quad \boldsymbol{\mu} \succeq \mathbf{0},\ \vartheta \geq 0, \tag{28}$$

where the corresponding Lagrangian function can be expressed as (22) , in which the vector $\boldsymbol{\mu}$ and the scalar ϑ are non-negative Lagrange multipliers.

We can obtain the optimal number of transmit antennas $\overset{\star}{M}$ by satisfying the KKT conditions, which can be directly computed as:

$$\frac{\partial\mathcal{L}(M,\widetilde{\mathbf{R}},\boldsymbol{\mu},\vartheta)}{\partial M} = 0, \quad \frac{\partial\mathcal{L}(M,\widetilde{\mathbf{R}},\boldsymbol{\mu},\vartheta)}{\partial\widetilde{R}_k} = 0. \tag{29}$$

After some mathematical manipulations from the computed derivatives via Eq. (29), the following second-order equation can be obtained:

$$-(M-K)^2\eta_{\mathrm{EE}}P_{\mathrm{BS}}\log(2) + (M-K)\sum_{k=1}^{K}(\mu_k+1)$$
$$-(\eta_{\mathrm{EE}}\varrho+\vartheta)\sigma^2(K_1+1)\log(2)\sum_{k=1}^{K}\frac{\mathbf{v}^H\mathbf{A}_k\mathbf{v}}{\mathbf{v}^H\mathbf{B}\mathbf{v}} = 0 \tag{30}$$

Finally, the optimal number of transmit antennas $\overset{\star}{M}$ in closed-form is given at the top of this page by Eq. (31).

Notice that for the proposed solution be feasible in practice, $\overset{\star}{M}$ must be limited in the range of $K+1$ and $M_{\max}$ since the constraint of the ZF precoding, (13c), should be attained and M should be lower than the maximum number of antennas available at the BS, $M_{\max}$.

Proceeding by substituting the calculated derivatives, Eq. (29), into the power equation p_k, Eq. (25), we obtain the optimal power allocation closed-form expression:

$$\overset{\star}{p}_k = \left[\frac{1+\mu_k}{\log(2)(\eta_{\mathrm{EE}}\varrho+\vartheta)} - \frac{\sigma^2(K_1+1)}{M-K}\frac{\mathbf{v}^H\mathbf{A}_k\mathbf{v}}{\mathbf{v}^H\mathbf{B}\mathbf{v}}\right]^+. \tag{32}$$

Lagrange multipliers Updating. Although the analytical expressions for the optimal number of transmit antenna, $\overset{\star}{M}$, and power allocation, $\overset{\star}{p}_k$, have been derived, these expressions are related to the Lagrangian multipliers. Herein, we apply the widely-used sub-gradient method to update the Lagrangian multipliers; it means that μ_k and ϑ should decrease if the gradients are positives, *i.e.*, $\nabla_{\mu_k}\mathcal{L} > 0$ and $\nabla_{\vartheta}\mathcal{L} > 0$, and vice versa. The values of μ_k and ϑ at the ℓ-th iteration will be updated according to

$$\begin{aligned}\mu_k^{(\ell)} &= \left[\mu_k^{(\ell-1)} - \frac{(R_k - R_{\min})}{10}\right]^+ \\ \vartheta^{(\ell)} &= \left[\vartheta^{(\ell-1)} - \frac{\sqrt{\ell}}{10}\left(P_{\mathrm{TX}} - \sum_{k=1}^{K}p_k\right)\right]^+,\end{aligned} \tag{33}$$

Algorithm 2 Proposed Solution for M and $\mathbf{p}$ Optimization

Input: M, σ^2, K_1, $\mathbf{p}$, $\mathbf{v}$, $\mathbf{B}$, $\mathbf{A}_k$, $\forall k \in \{1,\ldots,K\}$
repeat
  Step 0: Compute η_{EE} by (13);
  Step 1: Compute M by (31);
  Step 2: Compute $\mathbf{p}$ by (32);
  Step 3: Update the Lagrangian Multipliers by (33);
until η_{EE} converge
Outputs: $\overset{\star}{M}$ and $\overset{\star}{\mathbf{p}}$

Algorithm 3 EE Maximization Complete Solution Algorithm

Input: N, P_{TX}, σ^2, K_1, α_G, $\boldsymbol{\alpha}_F$, $\boldsymbol{\alpha}_D$, $\mathbf{a}_N$, $\mathbf{F}$
Step 0: Initilize M, $\mathbf{p}$, and $\mathbf{v}$ to feasible values;
Step 1: Compute $\mathbf{B}$ and $\mathbf{A}_k$ as (11) and (12) respectively;
repeat
  Step 2: Compute $\mathbf{v}$ according to Algorithm 1;
  Step 3: Update $\mathbf{A}_k$ as (12);
  Step 4: Compute $\mathbf{p}$ and M by Algorithm 2;
until the objective function in (13) converge
Outputs: $\overset{\star}{M}$, $\overset{\star}{\mathbf{p}}$ and $\overset{\star}{\mathbf{v}}$

C. Complexity

We provide the computational complexity analysis of the proposed Algorithms 1, 2, and 3. Concerning to the Algorithm 1, it is known that the complexity to update $\boldsymbol{\gamma}$, $\mathbf{u}$, and $\boldsymbol{\beta}$ results

in the same order, given by $\mathcal{O}(KN^2)$. Additionally, solve $\mathcal{P}_3$ by Eq. (23) has the complexity of $C^{\text{ANA}} = \mathcal{O}(N(N-1))$, while solve $\mathcal{P}_3$ with SFP metodology has complexity of $C^{\text{SFP}} = \mathcal{O}(N)$. Furthermore, Algorithm 2 has a complexity of $\mathcal{O}(3I_2KN^2)$. Therefore, the complexity of the proposed RIS phase-shift optimization $\mathbf{v}$ procedure, as well as for the number of active antennas M and power allocation $\mathbf{p}$ optimization, and the complete proposed solution procedure are given in Table I.

TABLE I
COMPLEXITY FOR THE OPTIMIZATION METHOD

Alg.	Complexity
1	$\mathcal{O}(I_1(KN^2 + \hat{I}_1(2KN^2 + \bar{I}_1 C^{\{i\}})))$
2	$\mathcal{O}(3I_2KN^2)$
3	$\mathcal{O}(I_3((I_1(KN^2 + \hat{I}_1(2KN^2 + \bar{I}_1 C^{\{i\}}))) + 3I_2KN^2))$

where I_1 and $\hat{I}_1$ represent the number of iterations for the inner and outer layers of Algorithm 1, while $\bar{I}_1$ is the number of iterations for optimizing $\mathbf{v}$, and $i \in \{\text{SFP}, \text{ANA}\}$ holds for SFP and Analytical approach, respectively; I_2 is the number of iterations for Algorithm 2 and I_3 is the number of iterations for Algorithm 3.

IV. SIMULATION RESULTS

In this section, we aim to investigate the performance of the multi-user RIS-aided mMIMO system. We denote a system configuration setup where the users' localization is fixed concerning their positions, AoA, and AoD for azimuth/elevation as illustrated in Fig. 2. For one single system configuration setup, the instantaneous-CSI optimization-based approach is averaged over $\mathcal{T} = 500$ realizations of $\mathbf{D}$ and $\mathbf{G}_{\text{NLoS}}$. Moreover, we assume the users are uniformly distributed in a circular area with the center in $(100;\, 50)\, m$ and a radius of $r = 15\, m$, while the RIS is located at $(100;\, 0)\, m$ and the BS is located at $(0;\, 0)\, m$. In each setup, the AoD angles were uniformally distributed as: $\phi^{r,k}_{\text{AoD}} \sim \mathcal{U}[0, \frac{\pi}{2}]$, $\varphi^{r,k}_{\text{AoD}} \sim \mathcal{U}[-\frac{\pi}{3}, \frac{\pi}{3}]$, and $\varphi^{bs}_{\text{AoD}} \sim \mathcal{U}[-\frac{\pi}{2}, \frac{\pi}{2}]$; besides, the AoA angles $\phi^{r}_{\text{AoA}} \sim \mathcal{U}[0, \frac{\pi}{2}]$ and $\varphi^{r}_{\text{AoA}} \sim \mathcal{U}[-\frac{\pi}{2}, \frac{\pi}{2}]$. All presented results for statistical CSI-optimization have been averaged over $\mathcal{S} = 50$ different setups. Table II summarizes the adopted values for the main simulation parameters.

Aiming to highlight the proposed method, we simulate four different variable optimization strategies denoted as

1) **p optz**: only the power of the users, $\mathbf{p}$, is optimized;
2) **p, v optz**.: the power of the users, $\mathbf{p}$, and the RIS phase-shift, $\mathbf{v}$, are optimized;
3) **p,** M **optz**.: The power of the users, $\mathbf{p}$, and the number of antennas, M, are optimized;
4) **p, v,** M **optz**.: The power of the users, $\mathbf{p}$, the number of antennas M, and the RIS phase-shift $\mathbf{v}$ are optimized sequentially.

It should be noted that when M, $\mathbf{v}$ or $\mathbf{p}$ are not optimized, they are assigned random feasible values. In addition, for comparison purposes, we also consider the **random all** allocation approach, where all variables are assigned random feasible values. Specifically, $\mathbf{p}$ is assigned values that satisfy constraint (13a), M obeys the constraint (13c), and $\mathbf{v}$ adheres to constraint (13d).

TABLE II
ADOPTED SIMULATION PARAMETERS.

Parameter	Value
RIS-Aided Massive MIMO System	
Max. Power Budget at BS	$P_{\text{TX}} \in \{20;\, 50\}$ [dBm]
Noise variance	$\sigma^2 = -95$ dBm
Numbers of UEs	$K = 10$
Max. # Antennas at BS	$M_{\max} = 256$
# Reflecting meta-surfaces	$N = 100$
Target rate	$R_{\min} = 1$ bps/Hz
Power inefficiency	ϱ = 1.2
Fixed power consumption	P_{FIX} = 9 dBW
Power BS antenna activation	P_{BS} = 1 W
Power RIS element activation	P_{RIS} = 0.1 W
Channel Parameters	
Path-loss models	$\alpha_G = -25\log(d_g)$
	$\alpha_{F,k} = -10.6 - 20\log(d_{F,k})$
	$\alpha_{D,k} = -35.6 - 40\log(d_{D,k})$
Matrix $\mathbf{G}$ (BS-RIS)	Rician fading
Matrix $\mathbf{D}$ (BS-UEs)	Rayleigh fading
Matrix $\mathbf{F}$ (RIS-UEs)	Purely LoS
Rician coefficient	$K_1 = 3.5$
Monte-Carlo simulation (MCs)	$\mathcal{T} = 500$ realizations
Setups	$\mathcal{S} = 50$ setups

A. Attainable Energy Efficiency

Fig. 3 illustrates the achievable EE performance as a function of the transmit power budget at BS, P_{TX}. We evaluated the performance for the four variable optimizations strategies: 1) only $\mathbf{p}$; 2) $\mathbf{p}$ and $\mathbf{v}$; 3) $\mathbf{p}$ and M; and 4) $\mathbf{p}$, $\mathbf{v}$ and M optimization, denoted as yellow, red, green, and blue color respectively. Moreover, we adopt as benchmarks the random all ($\mathbf{p}$, $\mathbf{v}$, and M) strategy, and the instantaneous CSI-based optimization strategy, which has been implemented based on Algorithm 1 proposed in [30], denoted herein as black and purple color curves, respectively. We also plotted the average instantaneous EE obtained from the proposed statistical CSI-based optimization approach, evaluated from the MCs, and depicted in orange color.

First and foremost, it must be noticed the consistency of the EE computed utilizing the ER lower bound, Eq. (10), which undoubtedly proves that $R_k > \widetilde{R}_k$. This is evident as the obtained EE MCs curve is slightly higher than the EE evaluated from the ER equation, demonstrating the tightness of the derivation proposed in [6] and affirming the potential of the statistical CSI-based optimization for enhancing EE. Regarding the different strategies for statistical CSI-based optimization, it is evident that the proposed algorithm for strategy 4) can outperform all other strategies with statistical CSI optimization in terms of EE enhancement, achieving a significant improvement by increasing the EE from ≈ 0.58 bps/Hz/W with the random solution (in the maximum value) to ≈ 1.42 bps/Hz/W. This translates to a remarkable EE gain of approximately 145%. Since it optimizes all variables $\mathbf{p}$,

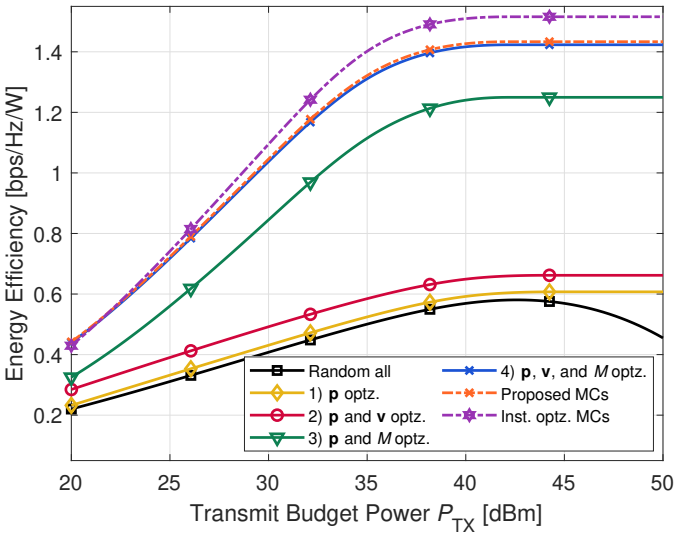


Fig. 3. Average EE *vs* the transmit power budget at the BS (P_{TX}). Performance evaluation of the proposed algorithm for four different approaches: 1) only $\mathbf{p}$; 2) $\mathbf{p}$ and $\mathbf{v}$; 3) $\mathbf{p}$ and M; and 4) $\mathbf{p}$, $\mathbf{v}$ and M optimization denoted as yellow, red, green, and blue color respectively. Random $\mathbf{p}$, $\mathbf{v}$, and M and the instantaneous optimization strategy are denoted as black and purple, respectively.

M, and $\mathbf{v}$ sequentially, such a result was expected. Besides, given the high value assumed by $M_{\max}$ as in practical RIS-aided mMIMO scenarios, there is a significant probability of obtaining a high value of M in the random realizations for both strategies 1) and 2) as depicted in Fig. 4, and analyzed in details in subsection IV-B. Thus, when M is high, the energy consumption is significantly increased implying poor EE. Accordingly, optimizing M becomes of paramount importance.

Moreover, in Fig. 3, one can infer that the impact of optimizing the phase shift RIS angles is not negligible on the attainable EE, even considering the statistical optimization. It is noteworthy that the achieved improvement in scenarios in which the number of active BS antennas M assumes low values (optimized M) is considerably greater compared to those scenarios where M assumes high values. This can be readily verified by examining the obtained gain in the EE curve of strategy 1) compared to the curve of strategy 2), as well as in the curve of strategy 3) with respect to the curve of strategy 4).

Based on Fig. 3, one can see the potential to achieve higher EE through the statistical CSI optimization approach since this methodology enables us to reduce the overhead associated with the phase-shift optimization during each channel coherence time while maintaining the fixed $\mathbf{v}$ over multiples channel coherence time with a minimal performance loss. As a result, the EE decreases from 1.510 to 1.433 bps/Hz/W, representing a loss of $\approx 5.3\%$, in high power regimes.

B. Average Number of Active Antennas

Fig. 4 shows the number of average active antennas M for the optimization strategies 1), 2), 3), and 4). It is noteworthy that strategies 1) and 2) exhibit a significantly higher average number of active antennas compared to the average number of active antennas in strategies 3) and 4). It occurs due to the random assignment for M in each setup $\mathcal{S}$, and since strategies 1) and 2) do not optimize M, there is a high probability for M in strategies 1) and 2) to be higher than that in strategies 3) and 4). This finding helps explain why strategies 1) and 2) attain poor EE as illustrated in Fig. 3, and higher rates as analyzed further ahead in Fig. 6.

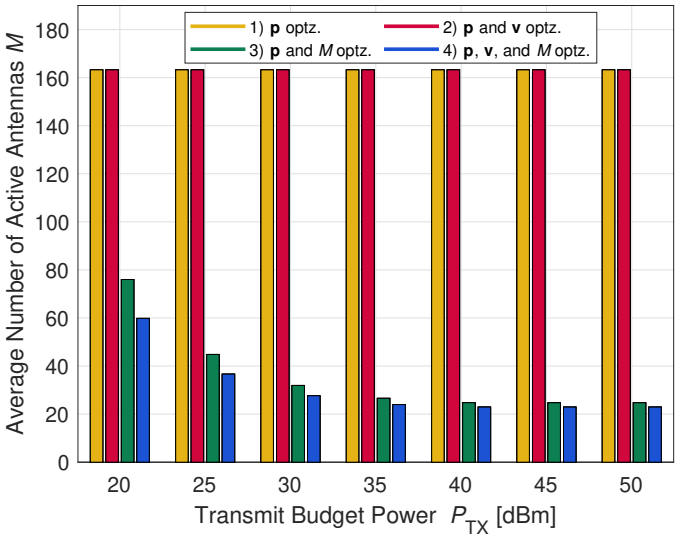


Fig. 4. Average number of antennas for four different strategies.

Furthermore, by optimizing the RIS phase-shift angles remarkably reduce the average number of active BS antennas, mainly in low-power regimes, *e.g.*, by comparing strategy 3) with strategy 4), at $P_{\mathrm{TX}} = 20$ dBm, the average number of antennas is reduced from 76 to 60. These results highlight the potential of RIS deployment and optimization.

C. Power Consumption

Fig. 5 depicts the average percentage of transmit power budget utilized for transmission versus the transmit power budget. In the low-budget power regime, all strategies deplete the available power budget; however, as the transmit power budget increases beyond 35dBm, the algorithm starts to minimize the consumption of the power utilized for transmission. This behavior is expected as in order to achieve the maximum EE, the available power should not be fully utilized.

Additionally, this strategy is directly accountable for the constant EE observed in high-power budget regimes, as presented in Fig. 3. Furthermore, it becomes apparent that optimizing the number of antennas M is of utmost importance in order to save a significant amount of power, reducing the power consumption substantially, from 100% (as in the "random all" case) to 84%, 26%, and 8% in the best strategy, for $P_{\mathrm{TX}} = 40$, 45, and 50 dBm respectively. Consequently, this strategy also implies high EE gain as presented in Fig. 3. Moreover, one can see the substantial impact on the reduction of power consumption caused by optimizing the RIS phase-shift angles. This power reduction corroborates the potential of RIS deployment and configuration, resulting in a remarkable energy efficiency gain that cannot be obtained in non-RIS scenarios.

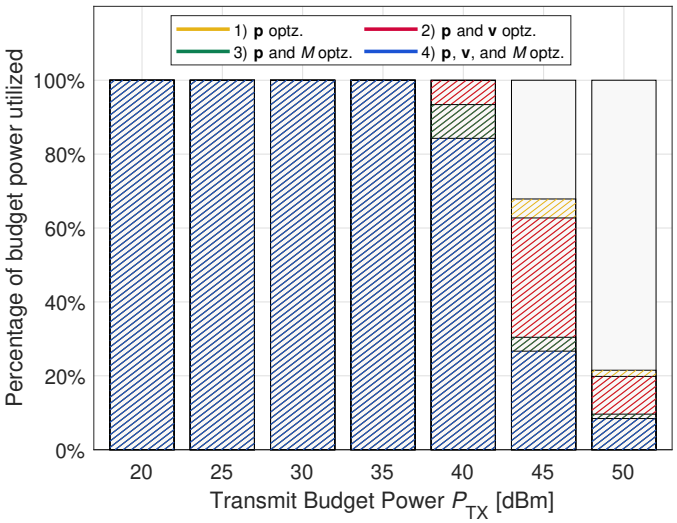


Fig. 5. Percentage of the transmit power budget utilized *vs* the transmit power budget at BS. The colors in the graph are associated with the strategy according to Fig. 3.

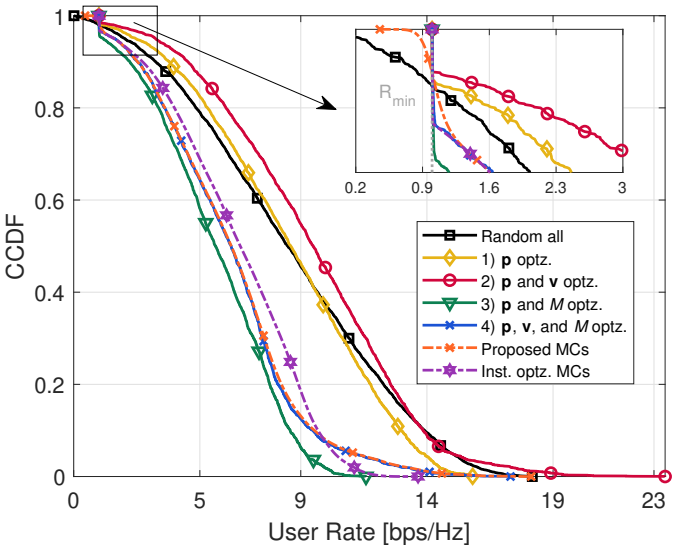


Fig. 6. CCDF of the ergodic user rate $\widetilde{R}_k$ for four different strategies with statistical CSI-based optimization (solid lines) and the instantaneous user rate R_k for instantaneous CSI-based optimization (dashed lines), with $R_{\min} = 1$ bps/Hz.

D. Distribution on the Attainable Data Rates

Fig. 6 reveals the complementary cumulative distribution function (CCDF) of the users' ergodic rate $\widetilde{R}_k$ over $\mathcal{S} = 50$ different realizations setups, represented by the solid lines. Besides, dashed lines indicate the computed instantaneous users' rate, R_k, over $\mathcal{T} = 500$ MCs for $\mathbf{D}$ and $\mathbf{G}_{\text{NLoS}}$, in the range of P_{TX} from 20 to 50 dBm. Notice that the proposed Statistical CSI-based optimization method mandatorily obeys the constraint (13b) for all strategies solutions. This fulfillment is primordial to guarantee appropriate QoS for all users. Moreover, one can see that strategies 1) and 2) are capable to provide higher ER than strategies 3) and 4). It can be justified owing to strategies 1) and 2), the number of antennas is not optimized, resulting higher number of BS antennas in the majority of setups than the optimized M's obtained in the strategy 3) and 4), as illustrated in Fig. 4, and consequently higher ER at cost of EE degradation.

Indeed, the consumed power in strategies 1) and 2) is higher than that one in strategies 3) and 4), as shown in Fig. 5, leading to higher rates. However, one can see that although strategy 4) has a much lower number of active antennas and lower power consumption than strategy 3), strategy 4) still has a feasible probability of attaining high rates than strategy 3). This is due to the RIS angle phase shift optimization, which can potentialize some users by maximizing the ER, as discussed in the subsection III-A. We can also notice from Fig. 6 that the instantaneous user rate R_k obtained through statistical CSI-based optimization (orange dashed curve) can achieve lower rates than $R_{\min}$, implying a violation of (13b), in contrast to the data rates obtained through the instantaneous CSI-based optimization (purple dashed curve), which are always higher or equal than $R_{\min}$.

E. Computional Complexity of Algorithm 1

In Fig. 7, we illustrate the ergodic rate *vs.* the average number of iterations (I_1) attainable with the proposed procedure for RIS phase shift optimization, Algorithm 1. Additionally, we include a benchmark utilizing the Gradient approach algorithm proposed in [6]. It is apparent that the proposed algorithm achieves a gradual convergence with significantly fewer iterations than the Gradient approach, *i.e.,* the proposed optimization method for both SFP and Analytical strategies needs $I_1 \approx 10$ iterations for convergence. Moreover, one can observe that the proposed algorithm outperforms the Gradient approach by achieving ≈ 146 bps/Hz, while the latter achieves around 141 bps/Hz after 50 iterations.

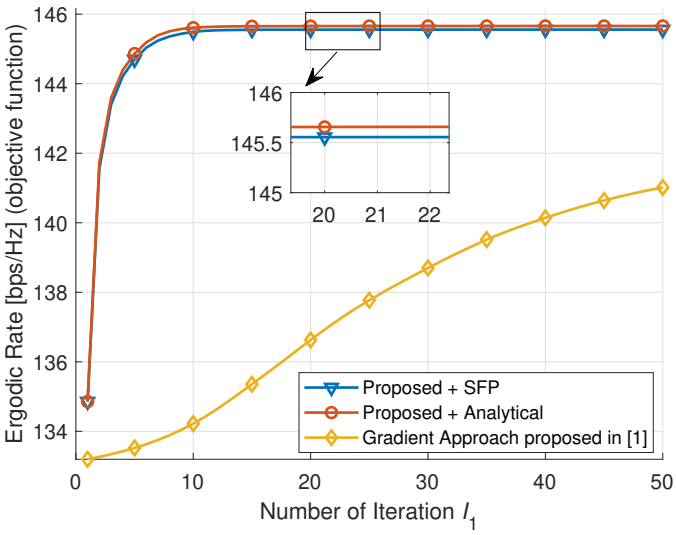


Fig. 7. The convergence behavior of the proposed Algorithm 1 for EE maximization and the convergence behavior of the Gradient approach benchmark proposed in [6] over $\mathcal{S} = 1000$ different setups. Here we adopt $M = M_{\max}$, $P_{\text{TX}} = 30$ dBm, $K = 20$ and $N = 100$.

We also can notice in Fig. 7 the potential of the proposed scheme with the SFP method solution since the performance gap is negligible regarding the optimal solution given ana-

lytically by Eq. (23), but resulting in a significantly lower complexity according to Table I and III. This clearly highlights the potential of the proposed solution. Furthermore, Table III confirms that the optimal Analytical proposed solution is more complex than the SFP approach, as the former has a complexity of $\mathcal{O}(N)$ while the latter has a complexity of $\mathcal{O}(N(N-1))$.

TABLE III
RUNNING TIME COMPARISON OF THE ALGORITHMS IN SECONDS

	Elapsed Time [s]
Proposed + SFP	0.2782
Proposed + Analytical	0.4575
Gradient Approach [6]	0.3075

V. CONCLUSIONS

This work addresses the EE maximization problem in a downlink RIS-aided mMIMO communication system operating under generalized Rician BS-RIS channels and ZF precoding. We proposed a statistical CSI-based optimization strategy that treats the RIS phase shift and power allocation as a constant over multiple channel coherence time intervals, reducing the overhead of the variables optimization in every channel coherence time. Based on an ergodic rate (ER) lower-bound closed-form expression, the EE optimization problem with power and QoS constraints is formulated. To circumvent the challenging coupled optimization variables and non-convex constraints, the proposed alternating-sequential optimization approach, based on the fractional programming techniques and dual Lagrangian method, optimizes the power allocation vector ($\mathbf{p}$), the number of BS antennas (M), and the RIS phase-shift vector ($\mathbf{v}$) in an iterative manner based on analytical closed-form expressions until convergence. Our approach offers an elegant methodology to gradually improve the system EE performance by optimizing step-by-step the system variables, demonstrating the potential of RIS deployment and optimization in mMIMO systems since the proposed scheme can result in high data rates for the users, lower power consumption and lower active antennas operating at BS. Our numerical results validate the effectiveness of the proposed Algorithm, as it substantially reduces the computational complexity while maintaining an exciting trade-off in terms of EE performance. Our approach incurs a performance loss of $\approx 5.3\%$, compared to instantaneous CSI-based optimization method. Besides, our proposed Algorithm can achieve better EE gains with lower complexity.

APPENDIX A
PROOF OF LEMMA 1

Initially, we should notice the $\mathcal{P}_3$ can be equivalently transformed into

$$\max_{\mathbf{v}} \ \mathbf{v}^H\mathbf{C}\mathbf{v} = \max_{\mathbf{v}} \ \mathrm{Tr}\left(\mathbf{C}\mathbf{V}\right), \tag{34}$$

with $\mathbf{V} = \mathbf{v}\mathbf{v}^H$, where the following constraints $\mathrm{rank}(\mathbf{V}) = 1$ $[\mathbf{V}]_{n,n} = 1$, $\forall n \in \{1,\ldots,N\}$, must be attined. Procceding, conveniently, we can rewritten $\mathbf{v}$ as the following manner

$$\mathbf{v} = [\mathbf{v}_1^H, e^{-j\theta_n}, \mathbf{v}_2^H]^H. \tag{35}$$

where $\mathbf{v}_1 \in \mathbb{C}^{n-1}$ and $\mathbf{v}_2 \in \mathbb{C}^{N-n}$ are respetively defined as $\mathbf{v}_1 \triangleq [\mathbf{v}]_{1:(n-1)}$ and $\mathbf{v}_2 \triangleq [\mathbf{v}]_{(n+1):N}$. This alternative notation enables us to look for the n-th element of $\mathbf{v}$, $e^{j\theta_n}$. Therefore, by utilizing Eq. (35), one can rewrite $\mathbf{V}$ as:

$$\mathbf{V} = \mathbf{v}\mathbf{v}^H = \begin{bmatrix} \mathbf{v}_1\mathbf{v}_1^H & \mathbf{v}_1 e^{-j\theta_n} & \mathbf{v}_1\mathbf{v}_2^H \\ e^{j\theta_n}\mathbf{v}_1^H & 1 & e^{j\theta_n}\mathbf{v}_2^H \\ \mathbf{v}_2\mathbf{v}_1^H & \mathbf{v}_2 e^{-j\theta_n} & \mathbf{v}_2\mathbf{v}_2^H \end{bmatrix}. \tag{36}$$

By defining the following variables

$$\begin{cases} \mathbf{C}_1 = [\mathbf{C}]_{1:(n-1),1:(n-1)} & \in \mathbb{C}^{(n-1)\times(n-1)} \\ \mathbf{c}_2 = [\mathbf{C}]_{1:(n-1),n} & \in \mathbb{C}^{(n-1)\times 1} \\ \mathbf{C}_3 = [\mathbf{C}]_{1:(n-1),(n+1):N} & \in \mathbb{C}^{(n-1)\times(N-n)} \\ C_4 = [\mathbf{C}]_{n,n} & \in \mathbb{C} \\ \mathbf{c}_5 = [\mathbf{C}]_{n,(n+1):N} & \in \mathbb{C}^{1\times(N-n)} \\ \mathbf{C}_6 = [\mathbf{C}]_{(n+1):N,(n+1):N} & \in \mathbb{C}^{(N-n)\times(N-n)}, \end{cases} \tag{37}$$

the matrix $\mathbf{C}$ can be rewritten in terms of these variables:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{c}_2 & \mathbf{C}_3 \\ \mathbf{c}_2^H & C_4 & \mathbf{c}_5 \\ \mathbf{C}_3^H & \mathbf{c}_5^H & \mathbf{C}_6 \end{bmatrix}, \tag{38}$$

Therefore, after some basic algebraic manipulations:

$$\begin{aligned} \mathrm{Tr}(\mathbf{C}\mathbf{V}) &= e^{-j\theta_n}(\mathrm{Tr}(\mathbf{c}_2^H\mathbf{v}_1) + \mathrm{Tr}(\mathbf{c}_5\mathbf{v}_2)) \\ &\quad + e^{j\theta_n}(\mathrm{Tr}(\mathbf{c}_2\mathbf{v}_1^H) + \mathrm{Tr}(\mathbf{c}_5^H\mathbf{v}_2^H)) + C_4 \\ &+\mathrm{Tr}\left(\mathbf{C}_3\mathbf{v}_2\mathbf{v}_1^H + \mathbf{C}_1\mathbf{v}_1\mathbf{v}_1^H + \mathbf{C}_3^H\mathbf{v}_1\mathbf{v}_2^H\right) + \mathrm{Tr}\left(\mathbf{C}_6\mathbf{v}_2\mathbf{v}_2^H\right), \end{aligned} \tag{39}$$

Interestingly, Eq. (39) leads us to see that maximize $\mathcal{P}_3$ w.r.t. θ_n is equivalently to

$$\max_{\theta_n} \ \mathrm{Tr}(\mathbf{C}\mathbf{V}) = \max_{\theta_n} \ 2\mathbb{R}\{e^{-j\theta_n}(\mathrm{Tr}(\mathbf{c}_2^H\mathbf{v}_1) + \mathrm{Tr}(\mathbf{v}_2\mathbf{c}_5))\}$$

whose optimal solution is straightforwardly given in closed-form by $\theta_n^\star = \angle\left(\mathrm{Tr}(\mathbf{c}_2^H\mathbf{v}_1) + \mathrm{Tr}(\mathbf{v}_2\mathbf{c}_5)\right)$. Substituting the values of $\mathbf{c}_2$ and $\mathbf{c}_5$ given in Eq. (37), and $\mathbf{v}_1$ and $\mathbf{v}_2$, the proof is complete.

REFERENCES

[1] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 4, pp. 2283–2314, 2020.

[2] 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, 2020.

[3] T. Marzetta, E. Larsson, H. Yang, and H. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[4] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[5] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, and M. Liyanage, "Survey on 6G Frontiers: Trends, Applications, Requirements, Technologies and Future Research," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 836–886, 2021.

[6] K. Zhi, C. Pan, H. Ren, and K. Wang, "Ergodic Rate Analysis of Reconfigurable Intelligent Surface-Aided Massive MIMO Systems With ZF Detectors," *IEEE Communications Letters*, vol. 26, no. 2, pp. 264–268, 2022.

[7] K. Zhi, C. Pan, G. Zhou, H. Ren, and K. Wang, "Analysis and Optimization of RIS-aided Massive MIMO Systems with Statistical CSI," in *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2021, pp. 153–158.

[8] E. Shtaiwi, H. Zhang, S. Vishwanath, M. Youssef, A. Abdelhadi, and Z. Han, "Channel Estimation Approach for RIS Assisted MIMO Systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 452–465, 2021.

[9] 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, 2019.

[10] M. Zeng, E. Bedeer, O. A. Dobre, P. Fortier, Q.-V. Pham, and W. Hao, "Energy-Efficient Resource Allocation for IRS-Assisted Multi-Antenna Uplink Systems," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1261–1265, 2021.

[11] 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, 2019.

[12] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy Efficiency and Spectral Efficiency Tradeoff in RIS-Aided Multiuser MIMO Uplink Transmission," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1407–1421, 2021.

[13] M. Forouzanmehr, S. Akhlaghi, A. Khalili, and Q. Wu, "Energy Efficiency Maximization for IRS-Assisted Uplink Systems: Joint Resource Allocation and Beamforming Design," *IEEE Communications Letters*, vol. 25, no. 12, pp. 3932–3936, 2021.

[14] H. Zhang, B. Di, Z. Han, H. V. Poor, and L. Song, "Reconfigurable Intelligent Surface Assisted Multi-User Communications: How Many Reflective Elements Do We Need?" *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1098–1102, 2021.

[15] A. Zappone, M. Di Renzo, X. Xi, and M. Debbah, "On the Optimal Number of Reflecting Elements for Reconfigurable Intelligent Surfaces," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 464–468, 2021.

[16] M. Forouzanmehr, S. Akhlaghi, A. Khalili, and Q. Wu, "Energy Efficiency Maximization for IRS-Assisted Uplink Systems: Joint Resource Allocation and Beamforming Design," *IEEE Communications Letters*, vol. 25, no. 12, pp. 3932–3936, 2021.

[17] Q. Ding, Y. Deng, X. Gao, and M. Liu, "Hybrid Precoding for mmWave Massive MIMO Systems with Different Antenna Arrays," *China Communications*, vol. 16, no. 10, pp. 45–55, 2019.

[18] E. Björnson and L. Sanguinetti, "Rayleigh Fading Modeling and Channel Hardening for Reconfigurable Intelligent Surfaces," *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 830–834, 2021.

[19] E. Björnson, O. T. Demir, and L. Sanguinetti, "A Primer on Near-Field Beamforming for Arrays and Reconfigurable Intelligent Surfaces," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 105–112.

[20] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large Intelligent Surface-Assisted Wireless Communication Exploiting Statistical CSI," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8238–8242, 2019.

[21] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[22] K. Shen and W. Yu, "Fractional Programming for Communication Systems—Part II: Uplink Scheduling via Matching," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2631–2644, 2018.

[23] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted Sum-Rate Maximization for Reconfigurable Intelligent Surface Aided Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3064–3076, 2020.

[24] Y. Jong, *An Efficient Global Optimization Algorithm for Nonlinear Sum-of-Ratios Problem*, 2012.

[25] K. Shen and W. Yu, "Fractional Programming for Communication Systems—Part I: Power Control and Beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.

[26] Z.-q. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[27] H. Li, J. Cheng, Z. Wang, and H. Wang, "Joint Antenna Selection and Power Allocation for an Energy-efficient Massive MIMO System," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 257–260, 2019.

[28] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[29] J. Tang, J. Luo, J. Ou, X. Zhang, N. Zhao, D. K. C. So, and K.-K. Wong, "Decoupling or Learning: Joint Power Splitting and Allocation in MC-NOMA With SWIPT," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5834–5848, 2020.

[30] N. I. Miridakis, T. A. Tsiftsis, and R. Yao, "Zero Forcing Uplink Detection Through Large-Scale RIS: System Performance and Phase Shift Design," *IEEE Transactions on Communications*, vol. 71, no. 1, pp. 569–579, 2023.

APPENDIX C – NOMA-aided Double RIS under Nakagami-m Fading: Channel and System Modelling

NOMA-aided Double RIS under Nakagami-m Fading: Channel and System Modelling. Conference paper. Accepted and presented in the Vehicular Technology Conference (VTC2023-Spring). June 2023.

NOMA-aided double RIS under Nakagami-m fading: Channel and System Modelling

Wilson de Souza Junior
Department of Electrical Engineering (DEEL)
State University of Londrina (UEL).
86057-970, Londrina, PR, Brazil.
wilsoonjr98@gmail.com

Taufik Abrão
Department of Electrical Engineering (DEEL)
State University of Londrina (UEL).
86057-970, Londrina, PR, Brazil.
taufik@uel.br

***Abstract*—We investigate the downlink outage performance of double-RIS-aided non-orthogonal multiple access (NOMA), where a near-BS and a near-users RISs setup are deployed. To extend the coverage to 360 degrees, we deploy a simultaneously transmitting and reflecting RIS (STAR-RIS) structure aiming to improve communication reliability for both indoor and outdoor users. New channel statistics for the end-to-end channel with Nakagami-m considering both the conventional-RIS and the STAR-RIS antenna elements features are derived using the moment-matching (MM) technique. The numerical results reveal that the double-RIS setup can outperform the single-RIS setups when the number of elements of STAR-RIS (R_S) and conventional RIS (R_C) is suitably adjusted. Moreover, when the link between the base station and the near-user RIS is in good condition, the double-RIS setup is outperformed by the single-RIS setup. Finally, the proposed analytical equations reveal to be very accurate under different channel and system configurations.**

***Index Terms*—Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surfaces (STAR-RIS); outage performance; moment-matching (MM)**

I. Introduction

The upcoming beyond fifth generation (B5G) wireless networks aim to provide high connectivity and different services for the most different kinds of users. Reconfigurable intelligent surface (RIS) has gained much space both in the research community and industries, and it is envisioned as a promising candidate to meet the stringent demands of future networks, mainly due to its ability to create a configurable propagation environment, enhancing metrics such as Spectral Efficiency (SE) [1], Energy Efficiency (EE) [2], as well as indicators as Quality of Service (QoS). However, depending on the place where the conventional RISs (C-RISs) are installed, users located in specific positions (blind spots) will not be served. Recently a new proposed architecture named simultaneously transmitting and reflecting RIS (STAR-RIS) can overcome this limitation by extending the coverage of communication by 180° to 360°. Besides, it has been shown that the RIS should be installed closer to the receiver to achieve the best performance gain[3]; thus, in order to improve the potential of this technology, double-RISs have been attracting much attention due to the high potential by combining the RIS deployment near to the transmitter and receiver.

Furthermore, with the exponential growth of connected devices in the network, schemes able to support massive connectivity are demanded, hence, non-orthogonal multiple access (NOMA) is considered a promising multiple access technique for B5G since it can provide access for a higher number of users, improving the system spectral efficiency. Recently, there is an effort to predict the behavior of the RIS-aided channels; hence, analytical expressions for *outage probability* (OP) and *ergodic capacity* (EC) considering different multiple access systems have been proposed in [4], [5], [6], [7]. These analytical approaches can be useful to assess the system's reliability and understand how the RIS-aided system scales according to some parameters. In [4], the authors derived analytical expressions for OP and EC for NOMA and OMA in downlink and uplink networks. In [5] the authors derived analytical equations for OP in a NOMA-aided single RIS scenario with random and optimal phase-shift design through the statistical moment-matching (MM) technique. In [6], the authors proposed a method to optimize the phase-shift of double-RIS-aided MIMO communication setup to maximize the system capacity. [7] investigated the performance of STAR-RIS assisted NOMA networks over Rician fading channels, and analytical equations for OP are derived.

To the best of the authors' knowledge, the analytical channel prediction behavior of a double-RIS (C-RIS + STAR-RIS)-aided NOMA has not yet been considered in the literature. Hence, motivated by the research gap considering this scenario we investigated the closed-form expressions for OP and EC; besides, we characterized conditions in which double-RIS configurations can be indeed profitable. The major contributions of this work are threefold: ***i***) we analyze a downlink double-RIS system, where a conventional RIS and a STAR-RIS are deployed to serve an indoor and an outdoor user. Through the MM technique, we statically characterize the end-to-end channel for this new scenario where all the communications links are considered as Nakagami-m; ***ii***) the OP and EC are derived in simple and accurate closed-form expressions and calculated for different parameters configurations of the double-RIS setup, as the number of elements in each RIS, as well as the m parameter of Nakagami distribution (LoS and NLoS impact). The developed analytical results are confirmed by Monte-Carlo simulations (MCs); ***iii***) we expose in which conditions and configurations the double-RIS setup can outperform the single-RIS setup.

Notations: Scalars, vectors, and matrices are denoted by the lower-case, bold-face lower-case, and bold-face upper-case let-

ters, respectively; $\mathbb{C}^{N\times M}$ denotes the space of $N\times M$ complex matrices; $|\cdot|$ denote the absolute operator; $\mathrm{diag}(\cdot)$ denotes the diagonal operator; $\Gamma(\cdot)$ is the gamma function; $\gamma(\cdot)$ is the lower incomplete gamma function; $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denotes the statistical expectation and variance operator; $\mathcal{G}(k,\theta)$ denotes the distribution of a Gamma random variable with shape and scale parameter k and θ respectively; $\mathcal{GG}(\alpha,\beta,\gamma,c)$ denotes an r.v. whose distribution is generalized-Gamma where the parameters are α,β,γ and c; $Nakagami(m,\Omega)$ denotes the distribution of a Nakagami-m random variable with shape and spread parameter m and Ω respectively;

II. System Model

As shown in Fig. 1, we consider a downlink double-RIS-aided two-user communication system, where the base station (BS), denoted by S, and the users are equipped with one single antenna. To extend the coverage to 360 degrees, we consider a STAR-RIS deployed in the building facade which can enhance the communication of indoor user (U_{I}) and outdoor user (U_{O}) simultaneously. We assume the U_{I} and U_{O} as the strongest and weakest users, respectively. The RISs are located near S (conventional RIS referred to as R_C) and near users (STAR-RIS referred to as R_S). Let us denote as N the total number of passive elements for the two RIS, where R_C and R_S consist of $N_C=|\mathcal{N}_C|$ and $N_S=|\mathcal{N}_S|$ elements respectively, with the set of RIS elements defined as $\mathcal{N}_C=\{1,\dots N_C\}$, $\mathcal{N}_S=\{1,\dots N_S\}$ and $N_C+N_S=N$. We define the *split factor* as $\eta\triangleq\frac{N_C}{N}$. Furthermore, we consider that U_{I} and U_{O} are paired to perform NOMA.

The R_C can be configured aiming to improve the communication through its phase-shift matrix given as $\mathbf{\Phi}_C=\mathrm{diag}([\nu_1 e^{j\theta_1},\dots,\nu_{N_C}e^{j\theta_{N_C}}])$, where $\theta_n\in[0,2\pi]$ and $\nu_n\in(0,1]$ are the phase-shift and the amplitude coefficient applied at the n-th element of R_C, respectively. Here, we adopt ideally $\nu_n=1$, $\forall n\in\{1,\dots,N_C\}$. Besides, concerning the STAR-RIS (R_S), its phase-shift matrix can be defined as $\mathbf{\Phi}_S^i=\mathrm{diag}([\xi_{i,1}e^{j\varphi_{i,1}},\dots,\xi_{i,N_S}e^{j\varphi_{i,N_S}}])$ with $i\in\{I,O\}$, where $\xi_{\mathrm{I},n}$ ($\varphi_{\mathrm{I},n}$) and $\xi_{\mathrm{O},n}$ ($\varphi_{\mathrm{O},n}$) denotes the amplitude coefficient (phase-shift) for refraction and reflection of n-th element respectively.

This paper focuses on the energy splitting (ES) protocol, where the STAR-RIS is assumed to operate in transmitting and reflecting mode simultaneously by splitting the total radiation energy into the refraction and reflection signals. By obeying the energy conservation law, the sum of the energies of the transmitted and reflected signals should be equal to the incident signal's energy, *i.e.*, $\xi_{\mathrm{I},n}+\xi_{\mathrm{O},n}=1\ \forall n\in\mathcal{N}_S$. Without loss of generality we adopt $\xi_{i,n}=\xi=0.5$, $\forall n\in\mathcal{N}_S$ and $\forall i\in\{I,O\}$.

A. Channel Model

In order to systematically characterize our proposed system model, let us denote $\mathbf{f}=[f_1,\dots,f_{N_C}]\in\mathbb{C}^{1\times N_C}$, $\mathbf{D}=[\mathbf{d}_1,\dots,\mathbf{d}_{N_S}]\in\mathbb{C}^{N_C\times N_S}$, $\mathbf{u}^i=[u_1^i,\dots,u_{N_S}^i]^T\in\mathbb{C}^{N_S\times 1}$ $\forall i\in\{\mathbf{I},\mathbf{O}\}$, $\mathbf{g}=[g_1,\dots,g_{N_C}]^T\in\mathbb{C}^{N_C\times 1}$ and $\mathbf{t}=[t_1,\dots,t_{N_S}]\in\mathbb{C}^{1\times N_S}$ as the channel between the $S\to R_C$, $R_C\to R_S$, $R_S\to U_i$, $R_C\to U_{\mathrm{O}}$[1] and $S\to R_S$ respectively. We consider that the absolute value of all elements of $\mathbf{f},\mathbf{D},\mathbf{u}^i,\mathbf{g}$ and $\mathbf{t}$ follow a Nakagami-m distribution with parameters (m_f,Ω_f), (m_D,Ω_D), (m_{U_i},Ω_{U_i}), (m_g,Ω_g) and (m_t,Ω_t), respectively. Since the channel estimation problem is out of the scope of this manuscript, we assume that perfect channel state information (CSI) is available at S.

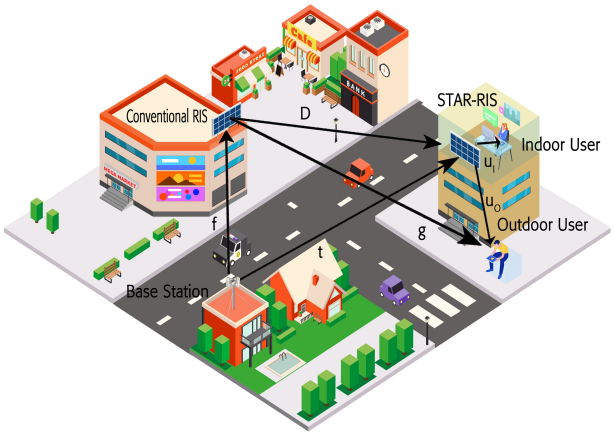


Fig. 1. Application scenario for double RIS-aided mobile communication.

III. New Statistics of the End-to-End Channel

In this section, we intend to derive new channel statistics for the proposed double-RIS-aided network, which will be useful to compute the OP and EC in the following subsections.

As illustrated in Fig. 1 we consider the double reflection link as well as the single reflection link from S to R_S towards the U_{I} and U_{O} and from the R_C to the U_{O}, therefore, one can write the cascaded channels of both user as follows:

$$h_I=\underbrace{\sqrt{\beta_t\beta_{U_{\mathrm{I}}}}\mathbf{t}\mathbf{\Phi}_S^I\mathbf{u}_I}_{\text{single-reflection}}+\underbrace{\sqrt{\beta_f\beta_D\beta_{U_{\mathrm{I}}}}\mathbf{f}\mathbf{\Phi}_C\mathbf{D}\mathbf{\Phi}_S^I\mathbf{u}_I}_{\text{double-reflection}}, \quad (1)$$

$$h_O=\underbrace{\sqrt{\beta_f\beta_g}\mathbf{f}\mathbf{\Phi}_C\mathbf{g}}_{\text{single-reflection}}+\underbrace{\sqrt{\beta_t\beta_{U_{\mathrm{O}}}}\mathbf{t}\mathbf{\Phi}_S^O\mathbf{u}_O}_{\text{single-reflection}}+\underbrace{\sqrt{\beta_f\beta_D\beta_{U_{\mathrm{O}}}}\mathbf{f}\mathbf{\Phi}_C\mathbf{D}\mathbf{\Phi}_S^O\mathbf{u}_O}_{\text{double-reflection}}, \quad (2)$$

where $\beta_v=\left(\frac{d_0}{d_v}\right)^{\alpha_v}$ denotes the *large-scale fading* gain (path-loss), d_0 is the reference distance, d_v is the distance of link v, and α_v is the path-loss coefficient, $\forall v\in\{f,g,t,D,u_{\mathrm{I}},u_{\mathrm{O}}\}$.

Notice that in order to maximize the single-reflection link for the U_{O} user, the optimal phase-shift in R_C can be applied setting $\theta_n=-(\angle f_n+\angle g_n)\ \forall n\in\mathcal{N}_C$; besides, aiming to explore the potential of the deployed R_S, we consider the case where it is configured to maximize the received power from the double-reflection link of both users, *i.e.*, the phase-shifts of the R_S are set as $\varphi_{i,n}=-(\angle([\mathbf{f}\mathbf{\Phi}_C\mathbf{D}]_n)+\angle u_{i,n})\ \forall n\in\mathcal{N}_S$, therefore, the signal of U_{I} and U_{O} will be coherently combined by the R_S. Thus, by setting the coherent phase shift design in both RISs, the following lemma is proposed.

[1]In this work, we do not consider the link between the $R_C\to U_{\mathrm{I}}$.

Lemma 1. *$|h_{\rm I}| \overset{approx}{\sim} \mathcal{G}\left(k_{h_{\rm I}}, \sqrt{\beta_{U_{\rm I}}}\theta_{h_{\rm I}}\right)$ whose scale and shape parameters are given by Eq.* (3)*. Moreover, the distribution $|h_O| \overset{approx}{\sim} \mathcal{G}\left(k_{h_O}, \theta_{h_O}\right)$ whose scale and shape parameters are given by Eq.* (4)*, where $\mu_v = \frac{\Gamma(m_v+1/2)\sqrt{\Omega_v}}{\Gamma(m_v)\sqrt{m_v}}$, $\forall v \in \{f, g, t, D, u_{\rm I}, u_{\rm O}\}$.*

$$\begin{cases} k_{h_{\rm I}} = \frac{N_S \mu_{U_{\rm I}}^2 \frac{\pi}{4}}{\left(\Omega_{U_{\rm I}} - \mu_{U_{\rm I}}^2 \frac{\pi}{4}\right)}, \\ \theta_{h_{\rm I}} = \frac{\xi\sqrt{\beta_{U_{\rm I}}}\sqrt{(\beta_t + \beta_f\beta_D N_C)}(\Omega_{U_{\rm I}} - \mu_{U_{\rm I}}^2 \frac{\pi}{4})}{\sqrt{\frac{\pi}{4}}\mu_{U_{\rm I}}}, \end{cases} \tag{3}$$

$$\begin{cases} k_{h_O} = \frac{\left(\sqrt{\beta_{U_{\rm O}}}\xi N_S \mu_{U_{\rm O}} \sqrt{\frac{\pi(\beta_t + \beta_f\beta_D N_C)}{4}} + \sqrt{\beta_f\beta_g} N_C \mu_f \mu_g\right)^2}{\beta_f\beta_g N_C\left(\Omega_f\Omega_g - \mu_f^2\mu_g^2\right) + \xi^2 \beta_{U_{\rm O}} N_S (\beta_t + \beta_f\beta_D N_C)\left(\frac{4\Omega_{U_{\rm O}} - \mu_{U_{\rm O}}^2 \pi}{4}\right)}, \\ \theta_{h_O} = \frac{\sqrt{\beta_{U_{\rm O}}}\xi N_S \mu_{U_{\rm O}} \sqrt{\frac{\pi(\beta_t + \beta_f\beta_D N_C)}{4}} + \sqrt{\beta_f\beta_g} N_C \mu_f \mu_g}{k_{h_O}}. \end{cases} \tag{4}$$

Proof. The proof can be found in Appendix A. ∎

Remark 1. If only the double-reflection link is considered, $\beta_t = \beta_g = 0$, besides, if only the single-reflection link is considered, we have that $\beta_D = 0$.

Besides, the probability density function (PDF) and cumulative distribution function (CDF) of $|h_i|$ can be given as [8]

$$f_{h_i}(x) = \frac{1}{\Gamma(k_{h_i})\theta_{h_i}^{k_{h_i}}} x^{k_{h_i}-1} e^{-\frac{x}{\theta_{h_i}}}, \tag{5}$$

$$F_{h_i}(x) = \frac{1}{\Gamma(k_{h_i})}\gamma\left(k_{h_i}, \frac{x}{\theta_{h_i}}\right). \tag{6}$$

IV. Outage Probability and Ergodic Capacity

We derive new OP closed-form expressions for double-RIS systems aided by STAR-RIS topology.

A. Signal Model of NOMA under Power Domain

For downlink communications, the S transmits the signal $x = \sqrt{P}\left(\sqrt{\lambda_I}s_I + \sqrt{\lambda_O}s_O\right)$, where P denotes the transmit power of the S, s_I and s_O denote the transmitted signals to $U_{\rm I}$ and $U_{\rm O}$, respectively, and λ_I and λ_O denote the power allocation coefficients for $U_{\rm I}$ and $U_{\rm O}$, respectively, with $\lambda_I + \lambda_O = 1$, and $\lambda_I < \lambda_O$. On the indoor and outdoor user side, the received signal is given by

$$y_i = h_i x + z_i, \qquad i \in \{I, O\}, \tag{7}$$

where the background noise $z_i \sim \mathcal{CN}(0, \sigma^2)$.

At $U_{\rm I}$, the signal of $U_{\rm O}$ is detected first, and the corresponding signal-to-interference-plus-noise ratio (SINR) is given by

$$\gamma_{{\rm O}\to{\rm I}} = \frac{\rho|h_I|^2\lambda_O}{\rho|h_I|^2\lambda_I + 1}, \tag{8}$$

where $\rho = \frac{P}{\sigma^2}$ is the transmit SNR under the AWGN channel. Besides, assuming perfect signal reconstruction and canceling in the successive interference cancellation (SIC) step, the SNR at the $U_{\rm I}$ is simply given by:

$$\gamma_{\rm I} = \rho|h_{\rm I}|^2\lambda_I. \tag{9}$$

At $U_{\rm O}$, its signal is decoded directly by regarding $U_{\rm I}$'s signal as interference, resulting in the following SINR

$$\gamma_{\rm O} = \frac{\rho|h_{\rm O}|^2\lambda_O}{\rho|h_{\rm O}|^2\lambda_I + 1}. \tag{10}$$

B. Outage Probability at Indoor User

According to the principle of NOMA, the outage behavior for the $U_{\rm I}$ occurs when it cannot effectively detect the user $U_{\rm O}$' message and its message consequently or when the $U_{\rm O}$' message is decoded successfully but an error occurs in the SIC process to decode its own message. Mathematically it can be formulated by the union of the following events:

$$\begin{aligned} P_{\rm out}^{U_{\rm I}} &= 1 - \Pr(\gamma_{{\rm I}\to{\rm O}} > \bar{\gamma}_{\rm th_O}, \gamma_{\rm I} > \bar{\gamma}_{\rm th_I}) \\ &= 1 - \Pr\left(|h_{\rm I}| > \sqrt{\frac{\bar{\gamma}_{\rm th_O}}{\rho(\lambda_O - \lambda_I\bar{\gamma}_{\rm th_O})}}, |h_{\rm I}| > \sqrt{\frac{\bar{\gamma}_{\rm th_I}}{\lambda_I\rho}}\right), \end{aligned}$$

where $\bar{\gamma}_{\rm th_O}$ and $\bar{\gamma}_{\rm th_I}$ are the SINR threshold for $U_{\rm O}$ and $U_{\rm I}$, respectively. Clearly, if $\lambda_O \le \lambda_I\bar{\gamma}_{\rm th_O}$, $P_{\rm out}^{U_{\rm I}} = 1$, otherwise by utilizing Eq. (6) the OP of $U_{\rm I}$ can be given as

$$P_{\rm out}^{U_{\rm I}} = \frac{1}{\Gamma(k_{h_{\rm I}})}\gamma\left(k_{h_{\rm I}}, \frac{\max\left\{\sqrt{\frac{\bar{\gamma}_{\rm th_O}}{\rho(\lambda_O - \lambda_I\bar{\gamma}_{\rm th_O})}}, \sqrt{\frac{\bar{\gamma}_{\rm th_I}}{\lambda_I\rho}}\right\}}{\theta_{h_{\rm I}}}\right), \tag{11}$$

where $k_{h_{\rm I}}$ and $\theta_{h_{\rm I}}$ are given respectively by Eq. (3).

C. Outage Probability at Outdoor User

In this case, the outage behavior for the $U_{\rm O}$ occurs since $U_{\rm O}$ cannot detect effectively its message. Hence, the OP of $U_{\rm O}$ is defined as

$$\begin{aligned} P_{\rm out}^{U_{\rm O}} &= \Pr(\gamma_{\rm O} < \bar{\gamma}_{\rm th_O}) = \Pr\left(|h_{\rm O}| < \sqrt{\frac{\bar{\gamma}_{\rm th_O}}{\rho(\lambda_O - \lambda_I\bar{\gamma}_{\rm th_O})}}\right) \\ &= \frac{1}{\Gamma(k_{h_{\rm O}})}\gamma\left(k_{h_{\rm O}}, \frac{\sqrt{\frac{\bar{\gamma}_{\rm th_O}}{\rho(\lambda_O - \lambda_I\bar{\gamma}_{\rm th_O})}}}{\theta_{h_{\rm O}}}\right), \end{aligned} \tag{12}$$

with $k_{h_{\rm O}}$ and $\theta_{h_{\rm O}}$ are given respectively by Eq. (4).

D. Ergodic Capacity

In this sub-section, the EC of the $U_{\rm I}$ and $U_{\rm O}$ for NOMA-Aided Conventional and STAR-RIS is derived based on Theorem 1.

Theorem 1. *In NOMA-Aided Conventional-RIS and STAR-RIS networks, the ergodic rate for the $U_{\rm I}$ and $U_{\rm O}$ can be predicted as*

$$\begin{cases} \bar{R}_{\mathrm{I}} \approx \log_2\left(1+\rho\lambda_{\mathrm{I}} N_S^2 \mu_{U_{\mathrm{I}}}^2 \pi \xi^2 \beta_{U_{\mathrm{I}}}(\beta_t+\beta_f\beta_D N_C)\right), \\ \bar{R}_{\mathrm{O}} \approx \\ \log_2\left(1+\dfrac{\rho\lambda_O\left(\frac{\xi N_S \mu_{U_{\mathrm{O}}}\sqrt{\beta_{U_{\mathrm{O}}}\pi(\beta_t+\beta_f\beta_D N_C)}+2\sqrt{\beta_f\beta_g}N_C\mu_f\mu_g}{2}\right)^2}{\rho\lambda_I\left(\frac{\xi N_S \mu_{U_{\mathrm{O}}}\sqrt{\beta_{U_{\mathrm{O}}}\pi(\beta_t+\beta_f\beta_D N_C)}+2\sqrt{\beta_f\beta_g}N_C\mu_f\mu_g}{2}\right)^2+1}\right), \end{cases} \tag{13}$$

Proof. The proof can be found in Appendix B. ■

V. Simulation Results

In this section, we provide numerical results to evaluate the accuracy of the proposed analytical equations. We consider three distinct scenarios: Sc.**A**: single- and double-reflection links are considered; Sc.**B**: only double-reflection link is considered; Sc.**C**: only single-reflection links are considered.

We show the impact of the reflecting elements on the performance of the NOMA-aided conventional and STAR-RIS system. The MCs results are averaged over 10^5 realizations. Unless stated otherwise, the adopted parameter values are presented in Table I.

TABLE I
Adopted Simulation Parameters.

Parameter	Value
NOMA-aided Conventional and STAR-RIS System	
Transmit SNR	$\rho \in [0; 50]$ [dB]
#Total Reflecting Meta-Surfaces	$N \in \{50, 100, 200\}$
Split Factor	$\eta = 0.35$
Power Allocation coefficients	λ_I = 0.15, λ_O = 0.75
Target Rate	$R_1 = R_2$ = 0.01 [BPCU]
Distance from S to R_C	25 [m]
Distance from R_S to U_{I} and U_{O}	5 and 20 [m]
Distance from R_C and R_S	15 [m]
Distance from R_C and U_{O}	35 [m]
Distance from S to R_S	35 [m]
Channel Parameters	
S-R_C link: **f**	$m_f = 8$, $\Omega_f = 1$
S-R_S link: **t**	$m_t = 1.5$, $\Omega_t = 1$
R_C-R_S link: **D**	$m_D = 8$, $\Omega_D = 1$
R_C-U_{O} link: **g**	$m_g = 1.8$, $\Omega_g = 1$
R_S-U_{I} link: $\mathbf{u}_I$	$m_u^I = 15$, $\Omega_u^I = 1$
R_S-U_{O} link: $\mathbf{u}_O$	$m_u^O = 7.5$, $\Omega_u^O = 1$
Path Loss Exponent	$\alpha_f = \alpha_D = 2.2$
	$\alpha_{U_{\mathrm{I}}} = \alpha_{U_{\mathrm{O}}} = 2$; $\alpha_g = 2.8$
	$\alpha_t \in \{2.8, 3.1, 3.4\}$
Reference Distance	$d_0 = 1$ [m]

Fig. 2 depicts the OP for both users *vs*. the split factor $\eta = \frac{N_C}{N}$ when $N = N_C + N_S = 200$. We can see that when the value N_C is low, U_{O} has a poor outage performance due to the marginal impact of the single reflection link $S \to R_C \to U_{\mathrm{O}}$ and double-reflection link $S \to R_C \to R_S \to U_{\mathrm{O}}$ once U_{O} is far from R_S and R_C. In contrast, since the U_{I} is near to the R_S, its performance is better when N_S assumes a high value.

Fig. 3 shows the ergodic sum rate *vs* the SNR (ρ) for three different scenarios, **A**, **B** and **C**. We can see that scenario **C** has the worse condition due to its low path-loss, besides, we can see the great potential of scenario **B** and how it can directly impact the double-RIS setup, outperforming even the single-RIS setup. Furthermore, we can see that the derived analytical equations (13) are very accurate when ρ varies.

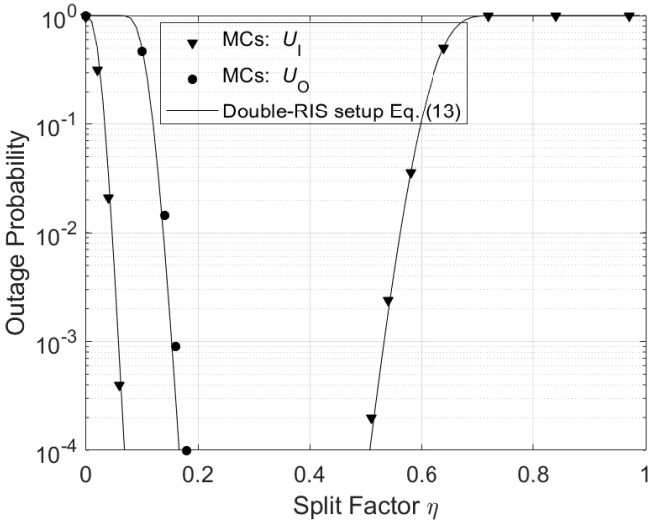


Fig. 2. Outage performance *vs* split factor (η) for $N = 200$ and $\alpha_t = 3.4$ on scenario A.

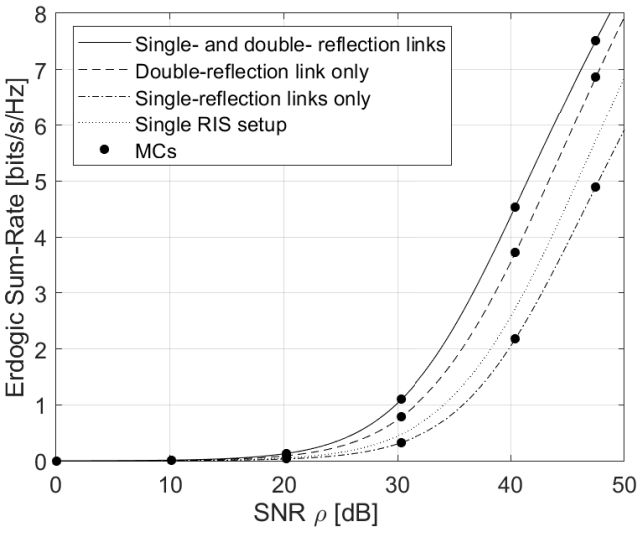


Fig. 3. Ergodic sum-rate ($\bar{R}_{\mathrm{I}} + \bar{R}_{\mathrm{O}}$) *vs* SNR ($\rho$) $N = 200$, $\eta = 0.35$ and $\alpha_t = 3.4$.

Fig. 4 depicts the ergodic sum rate for both users versus the split factor when $N = 200$ for three different path-loss coefficients, $\alpha_t = \{2.8, 3.1, 3.4\}$, where the path-loss $\beta_t = \left(\frac{d_0}{d_t}\right)^{\alpha_t}$. We first notice that for all scenarios considered the analytical equations as summarized in Eq. (13) are very accurate for any value of the split factor η. Furthermore, we can see that, by improving the conditions of the direct link between the S and the R_S, in terms of path-loss, the performance of the system can be improved in a general

way, however, within these conditions, the double-RIS setup can be outperformed by the single-RIS setup. On the other hand, when $\alpha_t = \{3.1, 3.4\}$, the double-RIS setup can obtain higher data rates than the single-RIS setup, revealing its great potential in scenarios where the direct link between the S and R_S is poor. Finally, we can see the scenarios where the double-RIS setup performs well, the number of meta-surface elements N_S and N_C can be strategically optimized aiming to obtain as better performance as possible.

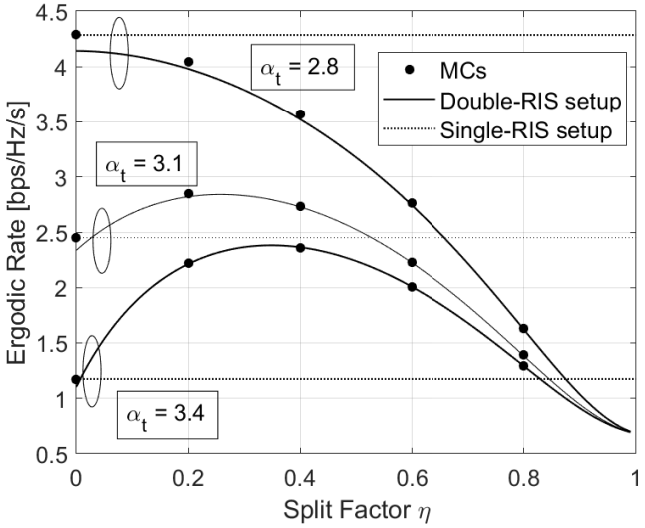


Fig. 4. EC *vs* split factor (η) over an NOMA-aided conventional and STAR-RIS system *vs* single-RIS setup with $N = 200$ and $\rho = 35$ dB.

Fig. 5 depicts the OP for both $U_{\rm I}$ and $U_{\rm O}$ also for the three different values of N. Firstly, one can notice that the derived analytical expressions (11) and (12) are very accurate for the considered parameters. Furthermore, we can notice that better performance for both $U_{\rm I}$ and $U_{\rm O}$ users occurs for high values of N. Additionally, we can observe how important is to know the CSI aiming to optimize the phase-shift matrix of R_S and R_C since it can provide very high-performance gains over the random phase-shift matrix strategy.

VI. Conclusions

In this paper, we have characterized the system performance of a NOMA-aided conventional and STAR-RIS under Nakagami-m channels. We derived analytical equations for OP and EC using the Central Limit Theorem (CLT) and MM technique. The numerical (analytical and MC simulation) results reveal that the double-RIS setup can outperform the single-RIS setups when the values of R_S and R_C are suitably adjusted. Furthermore, in scenarios where the direct link between the $S \to R_S$ are in poor conditions w.r.t. the path-loss (large α_t), the double-RIS can outperform the single-RIS setup. Finally, the simulation results demonstrated that the analytical proposed equations are very accurate under the parameters of the system.

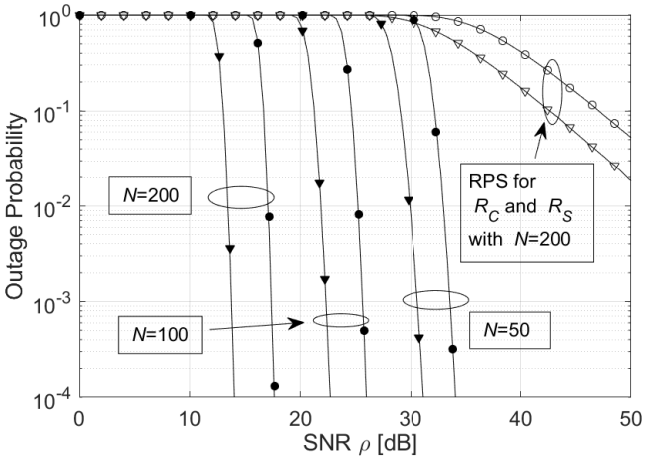


Fig. 5. OP *vs* SNR (ρ) when $N = \{50, 100, 200\}$, $\alpha_t = 3.4$ and $\bar{\gamma}_{\rm I} = \bar{\gamma}_{\rm O} = 0.5$.

Acknowledgment

This work was supported by the National Council for Scientific and Technological Development (CNPq) of Brazil under Grant 310681/2019-7, the CAPES (Financial Code 001).

Appendix A
Proof of Lemma 1

According to section II-A all links follow Nakagami-m distribution whose mean and variance can be given as

$$\mu_v = \frac{\Gamma(m_v + 1/2)\sqrt{\Omega_v}}{\Gamma(m_v)\sqrt{m_v}}, \tag{14}$$

$$\sigma_v^2 = \Omega_v - \mu_v^2, \qquad \forall v \in \{f, g, t, D, u_{\rm I}, u_{\rm O}\}, \tag{15}$$

The channel of $U_{\rm O}$ in (2) can be written as

$$h_O = \sqrt{\beta_{U_{\rm O}}} \underbrace{\left(\underbrace{\sqrt{\beta_t}\mathbf{t} + \sqrt{\beta_f \beta_D}\, \underbrace{\mathbf{f}\mathbf{\Phi}_C\mathbf{D}}_{\mathbf{w}}}_{\mathbf{x}} \right) \mathbf{\Phi}_S^O \mathbf{u}_{\rm O}}_{Z_{\rm O}} + \sqrt{\beta_f \beta_g}\, \underbrace{\mathbf{f}\mathbf{\Phi}_C\mathbf{g}}_{Y},$$

$$= \sqrt{\beta_{U_{\rm O}}} Z_{\rm O} + \sqrt{\beta_f \beta_g} Y, \tag{16}$$

by setting $\theta_n = -(\angle f_n + \angle g_n) \ \forall n \in \mathcal{N}_C$, thus, Y can be written as

$$Y = \sum_{n=1}^{N_C} |f_n||g_n|, \tag{17}$$

according to Lemma 2 of [5], the distribution of Y can be approximated as $Y \overset{\text{approx}}{\sim} \mathcal{G}\left(\frac{\mu_Y^2}{\sigma_Y^2}, \frac{\sigma_Y^2}{\mu_Y}\right)$, where $\mu_Y = N_C \mu_f \mu_g$ and $\sigma_Y^2 = N_C(\Omega_f \Omega_g - \mu_f^2 \mu_g^2)$. Since $\mathbf{\Phi}_C$ is set to maximize the power of the single reflection link $S \to R_C \to U_{\rm O}$, its combination in $\mathbf{w}$ will appear random, hence, by applying Lemma 4 of [5] we can approximate $\mathbf{w}$ as

$$\mathbf{w} \overset{\text{approx}}{\sim} \mathcal{CN}\left(\mathbf{0}, N_C \mathbf{I}_{N_S}\right), \tag{18}$$

one can notice that [9] shows that the Central Limit Theorem (CLT) provides a good approximation for $\mathbf{w}$, however, the correlation clearly impacts the approximation of $|h_{\mathrm{O}}|$ and it will be characterized in the numerical results in sub-section. Besides, the distribution of $\mathbf{x}$ can be given as follows

$$\mathbf{x} \overset{\text{approx}}{\sim} \mathcal{CN}\left(\mathbf{0}, (\beta_t + \beta_f \beta_D N_C)\mathbf{I}_{N_S}\right), \tag{19}$$

since the distribution of x_n follows complex Gaussian, the distribution of $|x_n|$, follows Rayleigh $\forall n \in \mathcal{N}_S$ whose mean is given as $\mu_x \triangleq \mathbb{E}[|x_n|] = \sqrt{\frac{\pi(\beta_t+\beta_f\beta_D N_C)}{4}}$ and variance can be given as $\sigma_x^2 \triangleq \mathbb{V}[|x_n|] = \frac{(4-\pi)(\beta_t+\beta_f\beta_D N_C)}{4}$.

Furthermore we set $\varphi_n^{\mathrm{O}} = -\left(\angle x_n + \angle u_{\mathrm{O},n}\right)\ \forall n \in \mathcal{N}_S$, in order to maximize the received power of U_{O} thus Z_{O} can be written as follows

$$Z_{\mathrm{O}} = \xi \sum_{n=1}^{N_S} |x_n||u_{\mathrm{O},n}|, \tag{20}$$

we should notice that differently of (17), eq. (20) represents the sum of the product of a Rayleigh and Nakagami-m r.v.; however, since Rayleigh can be viewed as a particular case of a Nakagami-m r.v., we also applied Lemma 2 of [5] to approximate the PDF of Z_{O} as Gamma r.v., hence, $Z_{\mathrm{O}} \overset{\text{approx}}{\sim} \mathcal{G}\left(\frac{\mu_{Z_{\mathrm{O}}}^2}{\sigma_{Z_{\mathrm{O}}}^2}, \xi\frac{\sigma_{Z_{\mathrm{O}}}^2}{\mu_{Z_{\mathrm{O}}}}\right)$, where $\mu_{Z_{\mathrm{O}}} = \xi N_S \mu_x \mu_{U_{\mathrm{O}}}$ and $\sigma_{Z_{\mathrm{O}}}^2 \approx \xi^2 N_S(\Omega_{U_{\mathrm{O}}}(\beta_t + \beta_f\beta_D N_C) - \mu_{U_{\mathrm{O}}}^2\mu_x^2)$, where the *approx.* is due to the correlation in $\mathbf{x}$.

Hence, the distribution of $|h_O|$ in (16), is the sum of two scaled gamma random variables with different shapes and scales parameters, since the exact closed-form expression for its PDF can be toilsome to obtain, we approximate the distribution of $|h_O|$ according to the well-known *Welch-Satterthwaite* [10] as follows

$$|h_O| \sim \mathcal{G}\left(\frac{\left(\sqrt{\beta_{U_{\mathrm{O}}}}\xi N_S \mu_x \mu_{U_{\mathrm{O}}} + \sqrt{\beta_f\beta_g}N_C\mu_f\mu_g\right)^2}{\xi^2\beta_{U_{\mathrm{O}}}\sigma_{Z_O}^2 + \beta_f\beta_g\sigma_Y^2}, \frac{\xi^2\beta_{U_{\mathrm{O}}}\sigma_{Z_O}^2 + \beta_f\beta_g\sigma_Y^2}{\xi\sqrt{\beta_{U_{\mathrm{O}}}}N_S\mu_x\mu_{U_{\mathrm{O}}} + \sqrt{\beta_f\beta_g}N_C\mu_f\mu_g}\right). \tag{21}$$

By following the same step in (16) and applying optimal phase-shift in φ_n^{I}, the channel of U_{I} in (1) can be written as

$$|h_I| = |\sqrt{\beta_{U_{\mathrm{I}}}}\underbrace{\mathbf{x}\mathbf{\Phi}_S^I\mathbf{u}_{\mathrm{I}}}_{Z_{\mathrm{I}}}| = \sqrt{\beta_{U_{\mathrm{I}}}}Z_I, \tag{22}$$

following the same steps done above, we obtain that $Z_{\mathrm{I}} \overset{\text{approx}}{\sim} \mathcal{G}\left(\frac{\mu_{Z_{\mathrm{I}}}^2}{\sigma_{Z_{\mathrm{I}}}^2}, \frac{\sigma_{Z_{\mathrm{I}}}^2}{\mu_{Z_{\mathrm{I}}}}\right)$, where the mean and variance are given as $\mu_{Z_{\mathrm{I}}} = \xi N_S \mu_x \mu_{U_{\mathrm{I}}}$ and $\sigma_{Z_{\mathrm{I}}}^2 \approx \xi^2 N_S(\Omega_{U_{\mathrm{I}}}(\beta_t + \beta_f\beta_D N_C) - \mu_{U_{\mathrm{I}}}^2\mu_x^2)$ respectively, therefore

$$|h_{\mathrm{I}}| \sim \mathcal{G}\left(\frac{\mu_{Z_{\mathrm{I}}}^2}{\sigma_{Z_{\mathrm{I}}}^2}, \sqrt{\beta_{U_{\mathrm{I}}}}\frac{\sigma_{Z_{\mathrm{I}}}^2}{\mu_{Z_{\mathrm{I}}}}\right). \tag{23}$$

Appendix B
Proof of Theorem 1

The EC of U_{I} and U_{O} can be written as

$$\bar{R}_i = \mathbb{E}\left[\log_2\left(1+\gamma_i\right)\right], \qquad \forall i \in \{\mathrm{I}, \mathrm{O}\}, \tag{24}$$

where γ_{I} and γ_{O} are given respectively by (9) and (10). By utilizing Jensen's inequality we can obtain an upper bound based on $\log_2(1+x)$, which is concave w.r.t. x, therefore

$$\bar{R}_{\mathrm{I}} \leq \log_2\left(1 + \rho\lambda_I\mathbb{E}[|h_{\mathrm{I}}|^2]\right), \tag{25}$$

$$\bar{R}_{\mathrm{O}} \leq \log_2\left(1 + \frac{\rho\lambda_O\mathbb{E}[|h_{\mathrm{O}}|^2]}{\rho\lambda_I\mathbb{E}[|h_{\mathrm{O}}|^2]+1}\right). \tag{26}$$

Since $|h_i| \sim \mathcal{G}(k_{h_i}, \theta_{h_i})$, we have that

$$f_{|h_i|^2}(x) = f_{|h_i|}(x)\frac{1}{2\sqrt{x}} = \frac{1}{2\theta_i^{k_i}\Gamma(k_i)}x^{\frac{k_i}{2}-1}e^{-\frac{x^{1/2}}{\theta_i}}, \tag{27}$$

therefore $|h_i|^2 \sim \mathcal{GG}(k_i, \theta_i^2, 0, 1/2)$ [11, Eq. 17.116], thus, the mean of generalized gamma r.v. is given by [11, Eq. 17.118]

$$\mathbb{E}[|h_i|^2] = \theta_i^2\frac{\Gamma(k_i+2)}{\Gamma(k_i)}, \tag{28}$$

One can notice that if $k_i \to +\infty$, $\Gamma\left(k_i+2\right) \approx \Gamma\left(k_i\right)k_i^2$, thus (28) can be approximate as

$$\mathbb{E}[|h_i|^2] \approx \theta_i^2 k_i^2, \tag{29}$$

thus substituting (29) into (25) and (26) respectively, (13) are obtained. This completes the proof.

References

[1] J. Zuo, Y. Liu, Z. Ding, L. Song, and H. V. Poor, "Joint Design for Simultaneously Transmitting and Reflecting (STAR) RIS Assisted NOMA Systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 611–626, 2023.

[2] Q. Zhang, Y. Wang, H. Li, S. Hou, and Z. Song, "Resource Allocation for Energy Efficient STAR-RIS Aided MEC Systems," *IEEE Wireless Communications Letters*, pp. 1–1, 2023.

[3] X. Gan, C. Zhong, C. Huang, and Z. Zhang, "RIS-Assisted Multi-User MISO Communications Exploiting Statistical CSI," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 6781–6792, 2021.

[4] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, and H. Vincent Poor, "Downlink and Uplink Intelligent Reflecting Surface Aided Networks: NOMA and OMA," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3988–4000, 2021.

[5] B. Tahir, S. Schwarz, and M. Rupp, "Analysis of Uplink IRS-Assisted NOMA Under Nakagami-m Fading via Moments Matching," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 624–628, 2021.

[6] Y. Han, S. Zhang, L. Duan, and R. Zhang, "Double-IRS Aided MIMO Communication Under LoS Channels: Capacity Maximization and Scaling," *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2820–2837, 2022.

[7] X. Yue, J. Xie, Y. Liu, Z. Han, R. Liu, and Z. Ding, "Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface Assisted NOMA Networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 189–204, 2023.

[8] A. Papoulis, *Probability and statistics*. Prentice-Hall, Inc., 1990.

[9] Z. Ding, R. Schober, and H. V. Poor, "On the Impact of Phase Shifting Designs on IRS-NOMA," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1596–1600, 2020.

[10] A. T. Abusabah, L. Irio, R. Oliveira, and D. B. da Costa, "Approximate distributions of the residual self-interference power in multi-tap full-duplex systems," *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 755–759, 2021.

[11] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. Wiley, 1994, vol. 1.