UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

HEBERT DOUGLAS PEREIRA

NOMA-RA: PROTOCOLO DE ACESSO ALEATÓRIO BASEADO EM MÚLTIPLO ACESSO NÃO ORTOGONAL PARA MIMO MASSIVO

TESE

CURITIBA

2021

HEBERT DOUGLAS PEREIRA

NOMA-RA: PROTOCOLO DE ACESSO ALEATÓRIO BASEADO EM MÚLTIPLO ACESSO NÃO ORTOGONAL PARA MIMO MASSIVO

NOMA-RA: Non-Orthogonal Multiple Access-Based Random Access Protocol for Massive MIMO

Tese apresentado(a) como requisito para obtenção do título(grau) de Doutor em Engenharia Elétrica e Informática Industrial, do Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Prof. Dr. Glauber Gomes de Oliveira Brante
Coorientador: Dr. Jamil de Araújo Farhat

CURITIBA

2021

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Curitiba

HEBERT DOUGLAS PEREIRA

NOMA-RA: PROTOCOLO DE ACESSO ALEATÓRIO BASEADO EM MÚLTIPLO ACESSO NÃO ORTOGONAL PARA MIMO MASSIVO

Trabalho de pesquisa de doutorado apresentado como requisito para obtenção do título de Doutor Em Ciências da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Telecomunicações E Redes.

Data de aprovação: 26 de Novembro de 2021

Prof Glauber Gomes De Oliveira Brante, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Joao Luiz Rebelatto, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Jose Carlos Marinello Filho, - Universidade Tecnológica Federal do Paraná

Prof Richard Demo Souza, Doutorado - Universidade Federal de Santa Catarina (Ufsc)

Prof Samuel Montejo Sanchez, Doutorado - Universidad Tecnológica Metropolitana Del Estado Del Chile - Utem

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 26/11/2021.

To GOD.

To my great love, my wife Kaendra and my children Elliot and Jhosef.

To my parents Aparício and Ozirema and to my brothers Daython, Erick and Ludyana.

ACKNOWLEDGEMENTS

First of all, I want to express my gratitude to GOD, for the breath of life, for his grace and for giving me intelligence and capacity to conclude this work. "For of Him, and through Him, and to Him, are all things: to whom be glory for ever. Amem. Rm 11.36"

My sincere thanks to Professor Glauber Brante for his support, guidance and willingness to help. Your support and encouragement was essential to reach the end of this work. Also many thanks to my co-advisor Jamil Farhat for his friendship and dedication. Also my gratitude to Professor Richard Souza for contribute and good advice.

I want to thank my dear wife Kaendra, who has always been by my side, for her dedication, support, encouragement and care of our family. I couldn't have done it without her support, understanding and dedication, thank you very much. Thanks to my children Elliot and Jhosef who brought new meaning to our lives with their births.

I want also to express my thanks to my parents Aparício and Ozirema, who even being physically distant, always were present, helping and encouraging. I also want to thank my brothers Daython, Erick and Ludyana along with their families who also supported and encouraged me on this journey.

I want also to express my thanks to all who directly or indirectly contributed to this work, to all the members of the LabSC, professors and all the students for the great moments together, which unfortunately the pandemic caused great damage.

Also, I would like to thanks to the UTFPR and CPGEI for the opportunity to participate in this program; to CAPES, CNPq and Fundação Araucária for the financial support.

Finally, to all my great friends that I do not name them because they are several, that even though some are physically distant, they are present with advice, support and encouragement. Thanks for their partnership, friendship and company.

This study was financed in part by the CAPES Brazil, finance Code 001.

”If you can’t fly then run, if you can’t run then walk, if you can’t walk then crawl, but whatever you do you have to keep moving forward.”

(Martin Luther King Jr.)

RESUMO

PEREIRA, Hebert Douglas. **NOMA-RA: Protocolo de acesso aleatório baseado em múltiplo acesso não ortogonal para MIMO massivo**. 2021. 60 f. Tese (Doutorado em Engenharia Elétrica e Informática Industrial) – Universidade Tecnológica Federal do Paraná. Curitiba, 2021.

Nesta tese, investigamos o problema do acesso aleatório (RA do inglês: *Random Access*) em um cenário MIMO massivo. Em redes 5G, os dispositivos/usuários que desejam se conectar com a estação base, (BS do inglês: *Base Station*) precisam escolher uma sequência piloto dentre um conjunto de sequências ortogonais disponíveis. Porém considerando o cenário de internet das coisas (IoT do inglês: *Internet of Things*), o número de usuários/dispositivos que tentam acessar a BS é muito maior do que o número de sequências pilotos disponíveis, com isso colisões podem ocorrer com frequência. Considerando esse cenário, o protocolo de resolução de colisão de usuário mais forte (SUCRe do inglês: *Strongest-User Collision Resolution*) é modificado para incluir um acesso múltiplo não ortogonal (NOMA do inglês: *Non-Orthogonal Multiple Access*). O protocolo proposto, NOMA-RA, permite resolver colisões entre usuários que tentam acessar o meio utilizando a mesma sequência piloto. Os resultados numéricos mostram que a estratégia proposta NOMA-RA atinge um melhor desempenho em termos *sum-rate* com menor atraso médio em comparação com o esquema SUCRe tradicional, mesmo considerando a estimativa de canal imperfeita e cancelamento sucessivo de interferência (SIC do inglês: *Sucessive Interference Cancellation*). Por exemplo, em um cenário em que 10 sequências piloto estão disponíveis e 20 dispositivos competem para acessar o canal, em comparação com o protocolo SUCRe, o atraso médio é reduzido em 30%, enquanto a *sum-rate* é melhorada em 33% e 22% considerando, respectivamente, decodificadores SIC perfeitos e imperfeitos, quando a imperfeição do SIC foi fixada em 10%. Além disso, a porcentagem de usuários que não conseguem acessar a BS nesta situação cai de $66{,}5\%$ com SUCRe para $38{,}5\%$ com o esquema proposto NOMA-RA.

Palavras-chave: MIMO massivo. Resolução de colisão do usuário mais forte. Não ortogonal. Acesso Multiplo.

ABSTRACT

PEREIRA, Hebert Douglas. **NOMA-RA: Non-Orthogonal Multiple Access-Based Random Access Protocol for Massive MIMO**. 2021. 60 p. Thesis (PhD in Electrical and Computer Engineering) – Federal University of Technology - Paraná. Curitiba, 2021.

In this thesis, we investigate the random access (RA) problem in a massive MIMO scenario. In 5G networks, devices/users willing to connect to the base station (BS) need to choose a pilot sequence from a set of available orthogonal sequences. However, considering the internet of things (IoT) scenario, the number of users/devices trying to access the BS is much higher than the number of available pilot sequences, so that collisions may occur frequently. Considering this scenario, the strongest-user collision resolution (SUCRe) protocol is modified to include a non-orthogonal multiple access (NOMA) approach. The proposed NOMA-RA protocol allows to resolve collisions between users who try to access the medium using the same pilot signal. Numerical results show that the proposed NOMA-RA strategy achieves a better performance in terms of sum-rate with lower average delay in comparison with the traditional SUCRe scheme, even considering imperfect channel estimation and successive interference cancellation (SIC). For instance, in a scenario when 10 pilot sequences are available and 20 devices compete to access the channel, comparing with the SUCRe protocol, the average delay is reduced by 30% while the sum-rate is improved by 33% and 22% considering, respectively, perfect and imperfect SIC decoders, when the imperfection of SIC was fixed in 10%. In addition, the percentage of users that fail to access the BS in this situation drops from 66.5% with SUCRe to 38.5% with the proposed NOMA-RA scheme.

Keywords: Massive MIMO. Strongest-User Collision Resolution. Non-Orthogonal. Multiple-Access.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

1G	First-generation of cellular network
2G	Second-generation of cellular network
3G	Third-generation of cellular network
3GPP	Third Generation Partnership Project
4G	Fourth-generation of cellular network
ACB	Access Class Barring
ACBPC	Access Class Barring Power Control
BS	Base Station
CDMA	Code Division Multiple Access
CD-NOMA	Code Domain NOMA
CS	Compressed Sensing
CSI	Channel State Information
DL	Downlink
eMBB	Enhanced mobile broadband
FDMA	Frequency Division Multiple Access
HTC	Human-Type Communications
IoT	Internet-of-Things
IRS	Intelligent Reflecting Surface
ITU	International Telecommunication Union
LoRa	Long Range Radio
LPWAN	Low Power Wide Area Networks
LTE-M	Long Term Evolution-machine
MA	Multiple-Access
MMSE	Minimum Mean Square Error
mMTC	Massive Machine-Type Communication
MTC	Machine-Type Communications
NB-IoT	Narrow-band IoT
NOMA	Non-Orthogonal Multiple Access
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PD-NOMA	Power Domain NOMA
QoS	Quality of Service
RA	Random Access
RB	Resource Block
RM	Reed-Muller
SDMA	Space Division Multiple Access
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SUCRe	Strongest-User Collision Resolution
SUCR-GBPA	Strongest-User Collision Resolution - Graph-Based Pilot Access

SUCR-IPA	Strongest-User Collision Resolution - Idle Pilot Access
TDD	Time-Division Duplex
TDMA	Time Division Multiple Access
UL	Uplink
URLLC	Ultra-Reliable Low-Latency Communications
UWB	Ultra-Wideband
XL-MIMO	Extra-Large MIMO

LIST OF SYMBOLS

R	Cell radius
R_{Min}	Minimum radius with respect to the BS
M	Number of antennas at the BS
T	Number of channel uses into a coherence block
$\mathbf{h}_k$	Channel vector between user k and its BS
$\underline{\mathbf{h}}_k$	Small-scale fading
$\mathcal{CN}(\cdot,\cdot)$	Circularly-symmetric complex Gaussian distribution
$\mathbf{I}_M$	Identity matrix $M \times M$
β_k	Channel gain between user k and its BS
d_k	Distance in meters between the BS and the k-th user
ϱ	Lognormal standard deviation of the shadowing
$\mathcal{N}(\cdot,\cdot)$	Gaussian distribution
$\mathcal{U}_0$	Set of users in the cell
$\mathcal{A}_0$	Subset of active users in this cell at a given time
P_{a}	Access probability of users becoming active
$\mathcal{K}_0$	Subset of *inactive users* in the cell
K_0	Set of users that share the same pilot
τ_{p}	Number of available pilot sequences mutually orthogonal
$\boldsymbol{\psi}_{\tau_{\mathrm{p}}}$	Uplink pilots Sequences $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{\tau_{\mathrm{p}}} \in \mathbb{C}^{\tau_{\mathrm{p}}}$
ρ_k	Power transmission of the k-th user
$\mathcal{S}_t$	Set of users that select the same pilot t
$\lvert\mathcal{S}_t\rvert$	Number of users that transmit the t-th sequence $\boldsymbol{\psi}_t$
$\mathbf{Y}$	Signal receive at the BS before $\boldsymbol{\psi}_t$-correlated
$\mathbf{N}$	Receiver noise before $\boldsymbol{\psi}_t$-correlated
$\mathbf{y}_t$	Received pre-processed $\boldsymbol{\psi}_t$-correlated signal at the BS
$\mathbf{n}_t$	The effective receiver noise
σ^2	Noise variance
α_t	Sum of the signal and interference gains received at the BS during the transmission of pilot sequence $\boldsymbol{\psi}_t$
$\boldsymbol{\phi}_t$	Downlink pilot sequences
$\mathbf{V}$	Precoded downlink signal
q	BS transmit power per pilot
$\mathbf{z}_k^T$	Signal received at user $k \in \mathcal{S}_t$
$\mathbf{z}_k$	Signal received at user $k \in \mathcal{S}_t$ after $\boldsymbol{\phi}_t$-correlated
$\boldsymbol{\eta}_k^T$	Received noise before $\boldsymbol{\phi}_t$-correlated
ζ_k	Noise at receiver after $\boldsymbol{\phi}_t$-correlated
$\Re(z_k)$	Real part of z_k
$\hat{\alpha}_{t,k}$	α_t estimated for each user k
$\mathcal{R}_k$	Decision rule repeat
$\mathcal{I}_k$	Decision rule inactive
ϵ_k	Bias term

$\mathcal{Q}(\cdot)$	$\mathcal{Q}$-function
$\gamma(\cdot,\cdot)$	Incomplete gamma function
P_{an}	New access probability after first attempt failed
$\text{SNR}_k^{\text{SUCRe}}$	SNR at the BS with respect to the k-th user for the SUCRe protocol
R_k^{SUCRe}	Available data rate of the k-th user for the SUCRe protocol
$R_{\text{sum}}^{\text{SUCRe}}$	Sum-rate for the SUCRe protocol
A_k	Number of attempts from the user k
$\mathcal{A}^{\text{SUCRe}}$	Average delay
θ	Fraction of $\rho_1\beta_1\tau_{\text{p}}$ could not be perfectly cancelled in the first iteration of the SIC decoder
$\mathbf{R}_k$	Covariance matrix of the channel
$\mathbf{Q}$	Normalized covariance matrix of the observation after correlating with the pilot sequence
$\hat{\mathbf{h}}_k$	Channel estimation by MMSE
$\tilde{\mathbf{h}}_k$	Channel estimation error
$\mathbf{\Phi}_k$	Covariance matrix of the channel estimation error
$R_1^{\text{NOMA}-\text{RA}}$	Achievable data rates of the strongest user
$R_2^{\text{NOMA}-\text{RA}}$	Achievable data rates of the weakest user
$R_{\text{sum}}^{\text{NOMA}-\text{RA}}$	Sum-rate of NOMA-RA protocol
δ	Parameter that can be tuned to improve a desired system metric
$\delta^{\star}_{(\text{sch})}$	Optimal parameter δ that maximizes the sum-rate of the scheme (sch)

CONTENTS

1 INTRODUCTION

The number of wirelessly connected devices is growing rapidly. It was reported there were 8.8 billion mobile devices and connections in 2018 worldwide, with a forecast of reaching 13.1 billion by 2023, what correspond to 71% of population with mobile conectivity, in which 1.4 billion of those will be 5G capable (CISCO, 2021). It is estimated that 5G subscriptions will reach 1 billion devices two years earlier than 4G (ERICSSON, 2021). This growth is strongly driven by video streaming, social networking and new communication use cases, such as machine-type communications (MTC), also known as internet-of-things (IoT), where data transmission between various MTC devices, and with the underlying network, takes place with little or no human intervention (TALEB; KUNZ, 2012; TULLBERG *et al.*, 2016).

The big amount of available MTC applications are contributing in a major way to the growth of devices and connections, such as smart meters, video surveillance, healthcare monitoring, transportation, package or asset tracking, smart buildings, logistics tracking, and smart agriculture. By 2023, according to recent Cisco annual Internet report (2018-2023), there will be around 29.3 billion networked devices, where the number of MTC devices is predicted to be 14.7 billion, which will represent 50% of the total devices and connections (CISCO, 2021). For instance, 26.66 billion MTC devices were reported active by the end of 2020, an average of 127 new connected devices every second, and with a forecast of 75 billion MTC devices in the world by 2025 (SAFEATLAST, 2021). This tremendous growth rate will further increase over the next decade and can reach hundreds of billions, with a connection density of 10 million devices per km^2 by 2030 (CHEN *et al.*, 2021).

Table 1 – Comparison of existing wireless system supporting IoT.

	Zigbee	**Bluetooth**	**WiFi**	**LoRa**	**Cellular**
Spectrum	Unlicensed	Unlicensed	Unlicensed	Unlicensed	Licensed
Connectivity	Moderate	Small	Large	Massive	Massive
Throughput	Moderate	Low	High	High	High
Range	Short	Short	Moderate	Long	Long
Security	Moderate	Low	Moderate	High	High
Power	Low	Low	High	Low	Low
Mobility	No	No	No	Yes	Yes
Latency	Low	Low	Low	Low	Low

Source: (CHEN *et al.*, 2021)

To explore the full potential of the massive IoT, reliable communication is mandatory, which can be provided by an enormous amount of available wireless network with low-cost

commercial technologies, such as Zigbee (VARGHESE *et al.*, 2019), Bluetooth (CHANG, 2014) and WiFi (POKHREL; WILLIAMSON, 2018). However, these networks can support only short-range wireless access to a moderate number of devices, achieving a few hundred devices in an indoor environment or in a small area. This current infrastructure is not enough, front the perspective of newly emerging IoT services, that require an interconnected and ubiquitous wireless infrastructure to connect, interact, and exchange data anywhere and anytime (AL-FUQAHA *et al.*, 2015). On the other hand, long range radio (LoRa) and cellular IoT are two main access technologies for low power wide area networks (LPWAN) which can provide a wide coverage (RAZA *et al.*, 2017; SULYMAN *et al.*, 2017; SUNDARAM *et al.*, 2020). LoRa technology has a disadvantage compared to cellular IoT since it requires new network infrastructure, whereas cellular IoT can reuse the existing cellular infrastructure. Table 1 provides a comparison of existing wireless systems supporting IoT. For 5G and beyond, the wireless communications systems must support massive connectivity, with high throughput and long range, while maintaining a good mobility and low latency, which is the focus of this thesis.

Given the huge variety of available IoT devices, with different features and functionality, the international telecommunication union (ITU) and third generation partnership project (3GPP) perceived the need to classify these devices according to network usage scenarios. Considering the different services characteristics and their diverse quality of service (QoS) requirements, three scenarios were defined: enhanced mobile broadband (eMBB), massive MTC (mMTC) and ultra-reliable low-latency communications (URLLC) (ITU, 2015). Among these network usage scenarios, mMTC was selected by 3GPP as one of three main use cases of 5G wireless networks and provided dedicated specifications for cellular IoT, with narrow-band IoT (NB-IoT) providing support for fixed and low-rate scenarios, while long term evolution-machine LTE-machine (LTE-M) covers mobile and high-rate scenarios (3GPP, 2015). Long-Term Evolution (LTE) is the usual name for the fourth generation (4G) cellular system.

To enable multiple access with limited system resources is an inherent issue in cellular networks, specially when the demand grows fast. However, to deal with the mMTC in cellular network, appropriate multiple access techniques are required. Thus, to cope with this exponential increase in mobile communications demand, massive MIMO technology (LARSSON *et al.*, 2014; BJÖRNSON *et al.*, 2017b) is one of the key-enablers to achieve better spectral efficiency, enhanced system capacity, as well as to improve random access (RA) performance (CARVALHO *et al.*, 2017; SORENSEN *et al.*, 2018; SANGUINETTI *et al.*, 2018; MUKHERJEE *et al.*, 2019).

Considering such scenario, several medium access protocols operating under crowded networks have been proposed in the literature. These protocols usually fall into two categories: *i.)* grant-based (BJÖRNSON *et al.*, 2017a; HAN *et al.*, 2017a; HAN *et al.*, 2017b; MARINELLO *et al.*, 2020; MARINELLO; ABRÃO, 2019), in which nodes compete to receive permission to access the network , and *ii.)* grant-free (HAN *et al.*, 2020a; DING *et al.*, 2019; DING; CHOI, 2020b; YU *et al.*, 2020; SHAHAB *et al.*, 2020), which offers a reduced signaling overhead for both base station (BS) and users at the cost of potential collisions. Consequently, a disadvantage related to grant-free schemes is the need of robust decoding strategies for improved performance, requiring relatively complex signal processing techniques at the BS (HAN *et al.*, 2020a; DING *et al.*, 2019; DING; CHOI, 2020b; YU *et al.*, 2020; SHAHAB *et al.*, 2020).

Recently, the authors in (DING *et al.*, 2019) have proposed a semi grant-free protocol, in which one grant-free device transmits simultaneously with another grant-based user, using non-orthogonal multiple access (NOMA). Such concept draw considerably attention in the recent literature, *e.g.*, in (ZHANG *et al.*, 2020a; YANG *et al.*, 2020; GERASIN *et al.*, 2020; JAYANTH *et al.*, 2020; ZHANG *et al.*, 2020b). The idea is to enhance network performance by enabling grant-based users – usually eMBB users – and grant-free devices – typically with URLLC demands – to share the same spectrum resources (ZHANG *et al.*, 2020a; YANG *et al.*, 2020). In addition, scheduling and power allocation algorithms for eMBB users can also be employed to deal with the constraints imposed by grant-free URLLC transmissions (GERASIN *et al.*, 2020). Nevertheless, the coordination among grant-free and grant-based transmissions is done in a centralized way by the BS in (ZHANG *et al.*, 2020a; YANG *et al.*, 2020; GERASIN *et al.*, 2020; JAYANTH *et al.*, 2020; ZHANG *et al.*, 2020b). Interestingly, the authors in (FACENDA; SILVA, 2020) recently compare state-of-the-art grant-based and grant-free schemes, showing that, if a small feedback is allowed, the grant-based solutions perform closely to the state-of-art of grant-free access while using simpler coding schemes, suggesting that novel grant-based schemes should not be dismissed as a potential solution to the massive RA problem.

Considering the favorable propagation and channel hardening characteristics of the massive MIMO channel, the authors in (BJÖRNSON *et al.*, 2017a) propose a grant-based protocol to deal with crowded scenarios, in which the number of available pilots is smaller than the number of users willing to communicate with the BS. Such protocol, named strongest-user collision resolution (SUCRe), is employed to resolve collisions between users who try to access the medium eventually choosing the same pilot signal. Nevertheless, it is interesting to note

that the impact of NOMA is still incipient in the context of RA grant-based massive MIMO distributed protocols. For instance, a distributed grant-based RA protocol is proposed in (ZHANG *et al.*, 2020a), where users try to access the BS, which identifies which pilots collided. Then, the BS broadcasts the information of the pilots free of collisions, such that only these users transmit. The rest of the users, with pilot collisions, remain silent. On the other hand, by exploring power domain NOMA technique allows multiple users to communicate with the BS using the same time, frequency and spreading code resources, which increases the system throughput when compared to orthogonal multiple access (OMA) (OTAO *et al.*, 2012). Additionally, an interesting way to optimize the system performance, by mitigating the interference among users, relies in the application of NOMA with successive interference cancellation (SIC) (SAITO *et al.*, 2013; XU *et al.*, 2017). With SIC, the users are decoded in decreasing order of their statistical channel gains, while the other signals are treated as interference. Then, the first decoded signal is subtracted from the original received signal and the process continues until all users are decoded. As well known, the channel estimation accuracy plays a central role in exploring power domain NOMA-SIC strategy.

1.1 GOALS

1.1.1 Main Goal

To develop a random access protocol, based on knowledge of the SUCRe protocol proposed in (BJÖRNSON *et al.*, 2017a), by employing a NOMA-based collision resolution scheme. Considering the fact that most of the collisions in SUCRe occurs between two users only (BJÖRNSON *et al.*, 2017a), the NOMA technique is employed at the uplink in order to allow both users to transmit in a non-orthogonal power domain fashion using the same pilot in a decentralized and distributed way. This proposed protocol will be named as NOMA-RA.

1.1.2 Specific Goals

- To employ NOMA in a massive MIMO scenario in order to improve the system sum-rate;
- To compare the performance of the proposed NOMA-RA protocol in terms of average delay, fail access probability and sum-rate with the SUCRe protocol;

- To identify the effect of different bias term parameter to maximize the sum-rate;
- To investigate the system performance damages while considering imperfect channel state information (CSI) and imperfect SIC.

1.2 ORGANIZATION

The remainder of this work is organized as follows. Chapter 2 presents the basic concepts of random access in mobile communication, the main protocols and characteristics. It is also presented the difference between grant-free and grant-based random access protocols. The SUCRe protocol and its main characteristics are presented in Chapter 3.

In the sequel, Chapter 4 presents the proposed NOMA-RA protocol and its main characteristics. An important issue in wireless networks, that required especial attention, especially in NOMA systems, is the CSI estimation. So, it is considered perfect and imperfect CSI in the analysis. It is also investigated the effect of the bias term parameter, that helps the users to resolve a contention and improve the achievable sum-rate.

Chapter 5 presents some numerical examples and some considerations comparing NOMA-RA and SUCRe protocols in terms of collision probability, sum-rate and average delay. Finally, Chapter 6 presents the final considerations and proposals for future works.

The main contributions of this thesis were addressed in:

- H. D. Pereira, G. Brante, J. Farhat, R. D. Souza, J. C. M. Filho and T. Abrão, "*On the Sum-Rate of Contention Resolution in Massive MIMO With NOMA,*" in **IEEE Access**, vol. 9, pp. 24965-24974, 2021, doi: 10.1109/ACCESS.2021.3055579.

2 BASIC CONCEPTS

In this chapter we present the basic concepts of random access techniques to wireless networks, the main existing protocols and their characteristics.

2.1 MULTIPLE ACCESS TECHNIQUES

Multiple-access (MA) techniques have been regarded as one of the key-enablers in wireless communications with a significant evolution across network generations, which recently has focus on the support of the future massive IoT cellular networks. A variety of effective MA technique have been employed in cellular networks, such as frequency division multiple access (FDMA) in the first-generation (1G) of cellular network, time division multiple access (TDMA) in second-generation (2G), code division multiple access (CDMA) in third-generation (3G), and orthogonal frequency division multiple access (OFDMA) in 4G and 5G (CAI *et al.*, 2018).

The existing MA techniques may be categorized into OMA and NOMA strategies. The first allow only a single device to be served within the same time/frequency resource block (RB), whereas NOMA allows multiple devices sharing the same RB. Due to the limitation of the available radio resources, OMA technique may not satisfy the stringent QoS requirements of massive access, and thus new massive orthogonal access techniques that exploit the extra degrees of freedom in 5G and beyond-5G networks are required. One example comes from 5G massive MIMO deployments, in which BSs are equipped with a large antenna arrays.

Massive MIMO technology can achieve better spectral efficiency, enhanced system capacity, as well as improve RA. In some setups being already deployed, the BS can be equipped with 64, or more, antenna elements (MARZETTA, 2010; LARSSON *et al.*, 2014; BJÖRNSON *et al.*, 2017b; NGO *et al.*, 2013). Therefore, massive MIMO can provide access for a large number of devices simultaneously, separating the devices in the spatial domain, *i.e.*, using space division multiple access (SDMA) (CARVALHO *et al.*, 2017; SORENSEN *et al.*, 2018; SANGUINETTI *et al.*, 2018; MUKHERJEE *et al.*, 2019; HAN *et al.*, 2020b).

The performance of massive access is determined by the acquisition of accurate CSI at the BS, which can be a bottleneck due to high signaling and overhead. Another factor that can reduce the accuracy of CSI is the co-channel interference, due to pilot contamination, since there is not enough pilot sequences available for all devices and it is necessary to share

them (KHANSEFID; MINN, 2015; ELIJAH *et al.*, 2016; BJÖRNSON *et al.*, 2018).

Due to such limitation in OMA, NOMA is identified as a promising technology to provide massive connectivity for 5G and beyond (SAITO *et al.*, 2013; ZHANG *et al.*, 2016; DING *et al.*, 2016; CHEN *et al.*, 2021; WANG *et al.*, 2016; DING; CHOI, 2020a; CHEN *et al.*, 2021; LIU; YU, 2017; SHAHAB *et al.*, 2020). In NOMA the RB are shared among the users, so it has the potential to improve spectral efficiency, *i.e.*, NOMA can admit significantly more devices within same RB compared to OMA. Hence, NOMA is able to support massive access, even when radio spectrum is limited.

Despite these benefits, NOMA suffers with severe interference at the receiver side, since the received signal is a superposition of signals from different devices. An alternative to overcome such difficulties relies in SIC, which is a promising interference cancellation technique (DING *et al.*, 2014; LIU *et al.*, 2016; CHEN *et al.*, 2019). In the literature NOMA is classified into two categories, namely power domain NOMA (PD-NOMA) and code-domain NOMA (CD-NOMA). Usually, PD-NOMA is employed at the uplink of cellular communications, *i.e.*, when multiple devices transmit to the BS at the same time/frequency, applying SIC to decrease the co-channel interference in order to correctly decode the superposition of received signals. With SIC, the receiver first decodes the signal with the strongest transmit power and removes it from the received signal. Then, it decodes the second strongest signal, with the second highest transmit power, until all signals from different devices are recovered (DING *et al.*, 2014; ZHANG *et al.*, 2016; ISLAM *et al.*, 2017). On the other hand, in CD-NOMA (DAI *et al.*, 2018) the access devices are separated within the code domain, in which each device assigns different codes for multiplexing. Compared to conventional CDMA, the assigned codes are sparse, which can still offer spreading gains for suppressing undesired co-channel interference.

In this thesis, we adopt the PD-NOMA technique exploring the power difference at the received signals, with the advantage of a lower complexity in the processing at the BS, in comparison with CD-NOMA. In addition, PD-NOMA can improve spectral efficiency and capacity, user fairness and also can achieve massive connectivity, since NOMA has the ability to increase the number of simultaneous connections compared to OMA (LIAQAT *et al.*, 2018).

2.2 RANDOM ACCESS PROTOCOLS

Access protocols are used to coordinate the access requests of the inactive devices that want to become active in the network. Nevertheless, both the access efficiency and reliability can

be heavily affected when devices are subjected to fading, interference and noise. In addition, the devices QoS requirements, such as rate and latency, can affect the selection of suitable access protocols and techniques. Thus, the design of multiple access is a fundamental task to guarantee network requirements depending on a given application.

Its worth noting the difference between IoT devices and eMBB use cases, in which the latter typically refers to the human-type communications (HTC). IoT devices usually have random activity, mostly in uplink, small transmit-data size per device, partially/fully autonomous communication and, most importantly, sporadic transmission. On the other hand, in HTC the number of devices is smaller, communication is majorly in downlink due to multimedia activity, and the data size per device is large (SHAHAB *et al.*, 2020). Thus, to achieve good performance, an efficient access protocol is essential. The random access protocols fall into two categories, grant-based and grant-free random access protocols (LIU *et al.*, 2018b). These two random access protocols will be discussed in the following.

2.2.1 Grant-Free Random Access

In grant-free random access, each active device directly transmits without any grant (ABEBE; KANG, 2017). Moreover grant-free transmission using OMA is usually not feasible due to huge number of collisions, since the available orthogonal pilot sequence are limited. Hence, the grant-free random access is usually associated with NOMA, due to the massive number of devices (LIU *et al.*, 2018b; CHEN *et al.*, 2021). The key idea of grant-free random access is to detect the active devices based on the received pilot sequences, which occurs as follows. The devices that have data to transmit, send their pilot sequences and data together to the BS. The BS first detects which pilot sequence was used, detecting the active device. Next, if the pilot sequence is correctly detected, the BS estimates the device's channel based on this received pilot sequence and then decode the data (LIU; YU, 2017).

However, these protocols face important challenges, with a notable issue being the active device detection. The active device detection demands huge processing at the BS, since it is not possible to assign an orthogonal pilot sequence to all devices (DING; CHOI, 2020b; DING *et al.*, 2019; YU *et al.*, 2020; LIU; YU, 2017; HAN *et al.*, 2020a). Since the number of IoT devices and antennas at the BS are very large in the 5G systems and beyond, it can result in a high dimensional device state matrix, implying in huge computational complexity (CHEN *et al.*, 2021).

In general, three strategies are employed to detect the device's activity: compressed sensing (CS), covariance-based approach and unsourced random access, in which CS and covariance-based approaches are more common. These three strategies will be discussed as follows.

2.2.1.1 Compressed Sensing

Due to the sporadic traffic generated by IoT devices, where only the active devices send their pilot sequence, the received pilot signal at the BS is typically sparse. Hence, compressed sensing exploits the sparsity of the signal to recover it using far fewer samples than required by the Nyquist criteria, using efficient signal processing techniques (SENEL; LARSSON, 2018; CHEN *et al.*, 2018).

The CS problem is generally non-convex and to obtain the global optimal is not a simple task, thus different approaches for the design of CS-based massive device detection algorithms have been studied, such as: sensing matrix, CS algorithms, joint device detection and channel estimation, and joint device and data detection. The design of the sensing matrix is an important and not trivial task, since the pilot sequences are non-orthogonal and the performance of grant-free random access is determined by it. For instance, Zadoff-Chu sequences (DING; CHOI, 2020a) have good auto- and cross-correlation properties, Gaussian sequences (CHEN *et al.*, 2018; LIU; YU, 2018) can be easily generated and are convenient for performance analysis, the Reed-Muller (RM) sequences (WANG *et al.*, 2019; YU, 2021) have reduced storage space requirements and a low-complexity activity detection algorithm, while a deep auto-encoded sequences exploit the properties of sparsity patterns, showing low computational complexity and are especially useful without analytical models (LI *et al.*, 2019; ZHANG *et al.*, 2020b; CUI *et al.*, 2021).

2.2.1.2 Covariance approach

In some cases, when only detecting the device's activity at the BS is required, without CSI estimation, the received signal covariance matrix at the BS can be used to formulate the device detection as a maximum likelihood estimation problem, as long as the BS be equipped with a large number of antennas (HAGHIGHATSHOAR *et al.*, 2018; CHEN *et al.*, 2019; CHENG *et al.*, 2020; FENGLER *et al.*, 2021). This approach is able to detect a very large number of

active devices, which can scale quadratically with the length of the pilot sequences (CHEN *et al.*, 2019; CHENG *et al.*, 2020).

2.2.1.3 Unsourced Random Access

Recently, a new paradigm in massive random access emerged, which is called unsourced random access (POLYANSKIY, 2017). This paradigm is a natural consequence of a very large number of potential devices of which only a finite number is active in each time slot. Compared to grant-based and original grant-free random access protocols where each device assigns a unique preamble sequence, in unsourced random access one codebook is shared for all devices (FENGLER *et al.*, 2019; SHAO *et al.*, 2020; TRUHACHEV *et al.*, 2021). The recent results have shown a significantly decrease in the minimum energy per bit required in a reliable communication, as well as some progress on codebook design for massive access (SHYIANOV *et al.*, 2021). However, unsourced massive random access is still an open and very challenging problem, specially in terms of the efficient codebook design and activity detection algorithms.

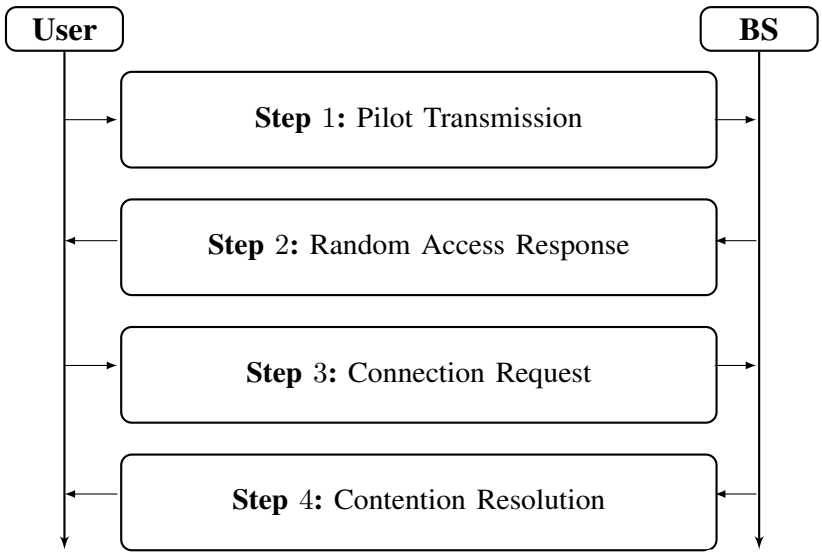
2.2.2 Grant-Based Random Access

Grant-based is the random access protocol adopted in the current 5G NB-IoT network communications (LIU *et al.*, 2018b). As suggested by the protocol's name, the device needs to negotiate with the BS to access the network. As the pilot sequence assigned to the devices are orthogonal, this protocol typically falls into the category of OMA. The grant procedure, as shown in Fig. 1, works similarly as in ALOHA and includes four steps between the device and the BS (HASAN *et al.*, 2013), in which the steps are summarized below.

1. *Pilot transmission*: Each active device picks a random pilot sequence, from a predefined set of orthogonal pilots, and send it to the BS to inform that it has data to transmit. Pilot sequence is the name of the orthogonal sequences adopted in the 5G nomenclature, what is also known as the preamble in the LTE nomenclature.

2. *Random Access Response*: The BS responds to each active device with the radio resource allocated to the device, authorizing it to send a connection request in the next step. In case a device does not receive the answer from the BS within a predefined waiting time, it postpones the access attempt to the next random access channel opportunity.

Figure 1 – Grant-Based random access protocol.



Source: The Author

3. *Connection Request*: Each device that has received a response to its pilot transmission sends a connection request to the BS, demanding resources for subsequent data transmission.

4. *Contention Resolution*: In case a pilot sequence has been selected by a single device, the connection request of the device is granted by the BS, which in turn sends a contention-resolution message to inform the device about the allocated resources. Otherwise, the access request is not granted. However if two or more devices have selected the same pilot sequence a collision occur. When the BS detects a collision, it does not reply with a contention-resolution message, so that the affected devices restart the random access procedure after an answer time expires.

The simple processing at the BS is the main advantage of the grant-based random access protocol. However, it has some shortcomings in the context of massive access. As the number of available pilot sequences is limited, they are shared among all devices. If a pilot sequence is selected by more than one device, a collision will occur, thus in massive access the devices can suffer from a high probability of access failure due to collisions, considerably increasing the average access latency, sometimes beyond tolerable. Another shortcoming is the signaling overhead, since the four steps depicted in Fig 1 are a major performance bottleneck and source of excessive delay (SHIRVANIMOGHADDAM *et al.*, 2015; CHEN *et al.*, 2018).

As this access mechanism is derived from the classic ALOHA, the number of active devices that can obtain the grant to access is limited. Recently, extensive efforts have been

devoted to different variations of the random access schemes with advanced contention-resolution strategies (BURSALIOGLU *et al.*, 2016; BJÖRNSON *et al.*, 2017a; HAN *et al.*, 2017b; HAN *et al.*, 2017a; MARINELLO; ABRÃO, 2019; MARINELLO *et al.*, 2020; NISHIMURA *et al.*, 2020). For instance, in (BURSALIOGLU *et al.*, 2016) a coded pilot approach was proposed, which can substantially avoid pilot contamination and increase multiplexing gains. To resolve the contention resolution problem, the authors in (BJÖRNSON *et al.*, 2017a) proposed a strongest-user collision resolution protocol, where the contention resolution is solved at the user side. Several authors extended the work presented in (BJÖRNSON *et al.*, 2017a). For instance, the authors in (HAN *et al.*, 2017a) propose an idle pilot access (SUCR-IPA) protocol, which allows a fraction of the users that failed to access the BS to re-select their pilots by means of an access class barring (ACB) factor, alleviating pilot collision. Another extension can be found in (HAN *et al.*, 2017b), in which a graph-based pilot access (SUCR-GBPA) protocol is proposed, allowing all users that lost the contention resolution to randomly select a new pilot. More recently, the authors in (MARINELLO *et al.*, 2020) improve the SUCRe protocol combining ACB with power control in the ACBPC protocol, which allows each user to estimate, without additional overhead, the number of users that collide with the same pilot. Additionally, in (MARINELLO; ABRÃO, 2019) a modification of the decision rule of the SUCRe protocol is proposed, in which each user decides to retransmit or not its pilot sequence according to the probability of being the strongest contender. Furthermore, the authors in (NISHIMURA *et al.*, 2020) propose a grant-based RA protocol to operate in extra-large MIMO (XL-MIMO) systems, where the large array size and the proximity with the users give rise to spatial non-stationarities across the array. Hence, the paper proposes modifications in the SUCRe protocol to exploit the non-overlapping visibility regions in XL-MIMO.

Even with the improvements in grant-based access procedure achieved until now, they might not be enough to support the mMTC traffic and access. The main reason is the increasing number of devices and their transmission pattern, since the RA approaches were basically designed for HTC. Therefore, in the next subsection we provide some discussions on the subject of massive access, which motivates the proposal under this thesis.

2.3 DISCUSSIONS

First of all, it is important to remember the main characteristics of the three 5G use cases: mMTC, eMBB and URLLC. The mMTC applications require reliable and energy-efficient

data transmission, supporting connection for a massive number of devices. The eMBB services are requested for applications that demand extremely high data rates and moderate requirements on latency and reliability. Finally, URLLC are requested for applications that demand very low latency (lower than 1ms) and very high reliability (99.999%) with moderate amounts of traffic. These three usage cases have different QoS requirements, which lead to significant channel resource consumption, even when low amounts of traffic is generated, as in the URLLC case.

As a consequence, different QoS requirements request different random access protocols. Grant-based RA protocols allow high throughput, with moderate latency and reliability. Such strategies are usually adopted for eMBB applications, in which high signaling and overhead are tolerated. Grant-free RA protocols allow low latency, which is usually the urge for URLLC applications, such as self-driving vehicles, industry automation, etc. For mMTC applications, the use of grant-free and grant-based schemes still have important performance trade-offs, since number of collisions may increase beyond tolerable in grant-free RA, while a limited number of devices will be served by grant-based approaches.

Recently, semi grant-free protocols have been proposed and draw considerably attention in the recent literature (DING *et al.*, 2019; YANG *et al.*, 2020; ZHANG *et al.*, 2020a; ZHANG *et al.*, 2020b; JAYANTH *et al.*, 2020; GERASIN *et al.*, 2020; DING *et al.*, 2021). Its aim is to take advantage of the benefits offered by both the grant-free and the grant-based protocols, since creating a new network structure for MTC devices is very expensive and current networks do not support their high demand. Semi grant-free protocols can be viewed as a compromise between grant-free transmission and conventional grant-based schemes, where one grant-based user eMBB shares the same resource with other grant-free mMTC or URLLC device. The main idea is to enhance network performance, decreasing the latency, signaling and overhead at the cost of higher processing at the BS, which must coordinate grant-free and grant-based transmissions.

A summary of grant-based, grant-free and semi-grant-free proposals in the recent literature is presented in Table 2. In addition, we also include the work in (PEREIRA *et al.*, 2021), which summarizes the main achievements of this thesis. It can be observed from Table 2 that our contribution is to apply NOMA in grant-based schemes. Our main goal is improve connectivity in a mMTC scenario, by employing NOMA, so that two devices may communicate with the BS using the same pilot sequence, at the same time and frequency. Our approach is based on the SUCRe protocol proposed in (BJÖRNSON *et al.*, 2017a), whose steps will be summarized

in Chapter 3. Then, the proposed NOMA-RA protocols will be detailed in Chapter 4.

Table 2 – Sumary of random access protocols

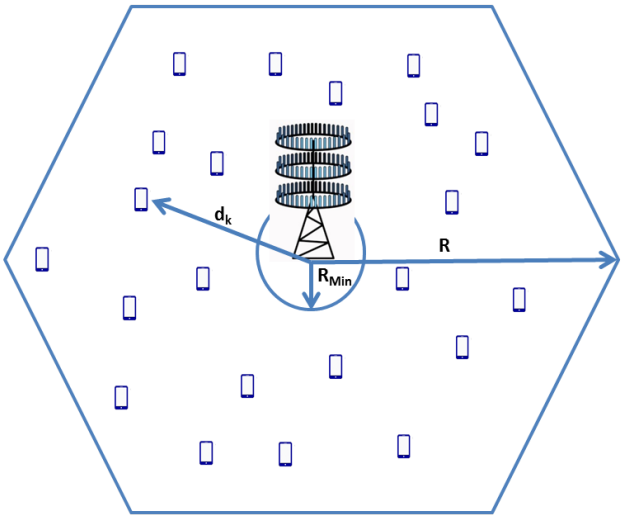
Reference	Protocol	Scenario	Access
(BJÖRNSON *et al.*, 2017a)	Grant-based	eMBB	OMA
(BURSALIOGLU *et al.*, 2016)	Grant-based	eMBB	OMA
(HAN *et al.*, 2017b)	Grant-based	MTC	OMA
(HAN *et al.*, 2017a)	Grant-based	eMBB	OMA
(MARINELLO; ABRÃO, 2019)	Grant-based	eMBB	OMA
(MARINELLO *et al.*, 2020)	Grant-based	eMBB/MTC	OMA
(NISHIMURA *et al.*, 2020)	Grant-based	eMBB	OMA
(ZHANG *et al.*, 2020a)	Grant-based	MTC	OMA
(PEREIRA *et al.*, 2021)	Grant-based	MTC	NOMA
(LIU; YU, 2018)	Grant-free	mMTC	NOMA
(LIU; YU, 2017)	Grant-free	mMTC	NOMA
(DING *et al.*, 2019)	Grant-free	MTC	OMA
(DING; CHOI, 2020b)	Grant-free	MTC	OMA
(YU *et al.*, 2020)	Grant-free	MTC	NOMA
(HAN *et al.*, 2020a)	Grant-free	MTC	NOMA
(SHAHAB *et al.*, 2020)	Grant-free	mMTC	NOMA
(SENEL; LARSSON, 2018)	Grant-free	mMTC	NOMA
(CHEN *et al.*, 2018)	Grant-free	mMTC	NOMA
(DING *et al.*, 2019)	Semi-grant-free	eMBB/mMTC	NOMA
(YANG *et al.*, 2020)	Semi-grant-free	eMBB/MTC	NOMA
(ZHANG *et al.*, 2020a)	Semi-grant-free	eMBB/MTC	NOMA
(ZHANG *et al.*, 2020b)	Semi-grant-free	eMBB/URLLC	NOMA
(JAYANTH *et al.*, 2020)	Semi-grant-free	eMBB/MTC	NOMA
(GERASIN *et al.*, 2020)	Semi-grant-free	eMBB/URLLC	OMA/NOMA
(DING *et al.*, 2021)	Semi-grant-free	eMBB/mMTC	OMA

Source: The Author

3 SUCRE PROTOCOL

We revisit the SUCRe protocol proposed in (BJÖRNSON *et al.*, 2017a), which is a distributed method to resolve pilot collisions at the user side. Let us consider a hexagonal cellular network, with radius R, as illustrated in Fig. 2. We assume that the users are uniformly distributed in the area, but also that no user is less than R_{Min} away from the BS. The BS is equipped with M antennas and the users are equipped with a single antenna. The system operates in time-division duplex (TDD) mode and we consider that time-frequency resources are divided into coherence blocks of T channel uses, with constant and frequency flat channel responses between the BS and its users within a block, while varying between blocks. Also, following (BJÖRNSON *et al.*, 2017a), only a few quantity of blocks are dedicated for radio access, with most part of the blocks used to payload data transmission for active users.

Figure 2 – System model of the hexagonal network with the users uniformly distributed within the cell of radius R.



Source: The Author

We consider Rayleigh fading channels, thus the channel $\mathbf{h}_k \in \mathbb{C}^M$ between the k-th user and the BS is

$$\mathbf{h}_k = \sqrt{\beta_k}\,\underline{\mathbf{h}}_k, \tag{1}$$

where $\underline{\mathbf{h}}_k \sim \mathcal{CN}(0,\mathbf{I}_M)$ is the small-scale fading, $\mathcal{CN}(\cdot,\cdot)$ denotes the circularly-symmetric complex Gaussian distribution, $\mathbf{I}_M$ denotes an $M \times M$ identity matrix, $\beta_k > 0$ represents the

channel gain and describes the path loss, which is modeled based on the urban micro scenario defined by 3GPP, given in dB by (3GPP, 2018, Table 5.1)

$$\beta_k[\text{dB}] = -34.53 - 38\log_{10}(d_k) + 10\varrho, \tag{2}$$

where d_k is the distance in meters between the BS and the k-th user, and 10 dB is the lognormal standard deviation of the shadowing, while $\varrho \sim \mathcal{N}(0,1)$, where $\mathcal{N}(\cdot,\cdot)$ denotes the Gaussian distribution.

Considering $\mathcal{U}_0$ the set of users in the cell, and $\mathcal{A}_0 \subset \mathcal{U}_0$ the subset of active users in this cell at a given time, in crowded MTC scenarios we have that $|\mathcal{U}_0| \gg T$ channel uses. However, the active users satisfy $|\mathcal{A}_0| < T$, therefore the BS temporarily assign orthogonal pilot sequences to these users and reclaim them when they finished their transmissions. We also define the probability of the users that desire becoming active as *access probability*, denoted by P_a, with $P_\text{a} \leq 1$. In addition, assuming $\mathcal{K}_0 = \mathcal{U}_0 \setminus \mathcal{A}_0$ as the subset of *inactive users* in the cell, we consider that $K_0 = |\mathcal{K}_0|$ users share $\tau_\text{p} \ll K_0$ mutually orthogonal uplink pilot sequences $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{\tau_\text{p}} \in \mathbb{C}^{\tau_\text{p}}$.

Taking this into account, each inactive user that want to become active, selects one of the τ_p pilots $\boldsymbol{\psi}_t$ at random in each RA block, with $t \in \{1,2,\ldots,\tau_\text{p}\}$, and transmits such pilot with power $\rho_k > 0$. We also define the set of users that select the same pilot t as $\mathcal{S}_t$, so that the number of users that transmit the t-th sequence $\boldsymbol{\psi}_t$ is denoted by $|\mathcal{S}_t|$, which follows a binomial distribution

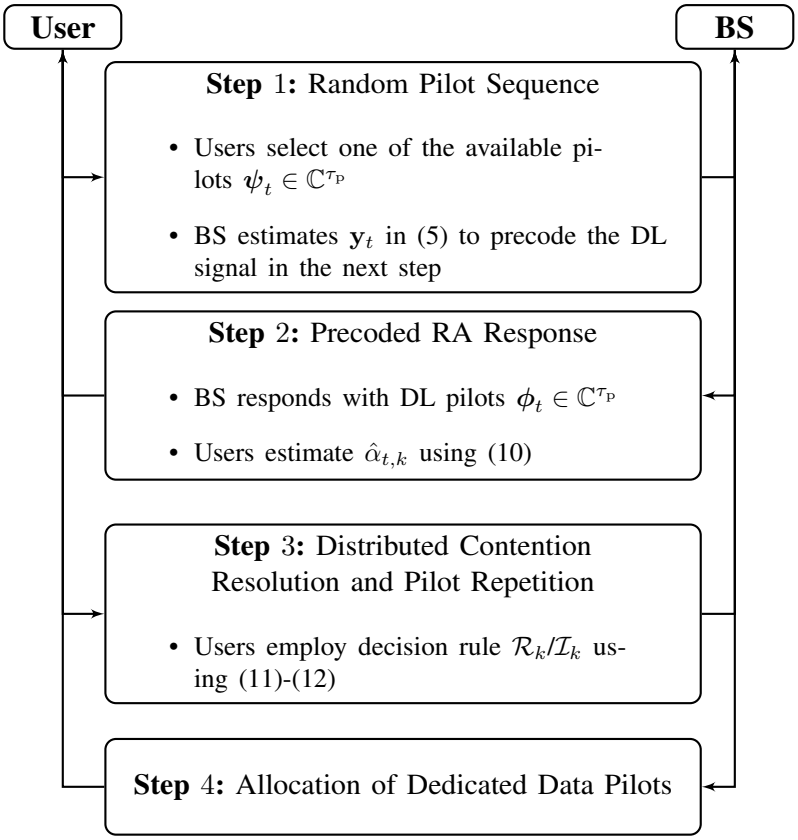
$$|\mathcal{S}_t| \sim \mathcal{B}\left(K_0, \frac{P_\text{a}}{\tau_\text{p}}\right). \tag{3}$$

The SUCRe protocol consists of four steps illustrated in Fig. 3. Notice that a previous step, called Step 0, is performed by the BS broadcasting a control signal in order to let the users estimate their own average CSI, defined by β_k in (2), and to synchronize itself with the BS.

In Step 1 some inactive user picks a pilot sequence from the pool of RA pilots and sends it to the BS. After receiving the RA pilots for all active users, the BS estimates the CSI between each user and the BS by correlating the received uplink (UL) signals with the t-th orthogonal sequence pilot $\boldsymbol{\psi}_t$. Let us remark that, if more than one user chooses the same pilot at this step, the estimated CSI will be the combination of the CSI of these multiple users. Hence, the BS receives the signal $\mathbf{Y} \in \mathbb{C}^{M\times\tau_p}$ from the pilot transmission,

$$\mathbf{Y} = \sum_{k\in\mathcal{K}_0} \sqrt{\rho_k}\,\mathbf{h}_k\,\boldsymbol{\psi}_t^T + \mathbf{N}, \tag{4}$$

Figure 3 – Steps of the SUCRe protocol.



Source: The Author

where $\mathbf{N} \in \mathbb{C}^{M\times\tau_p}$ is the receiver noise with each element distributed as $\mathcal{CN}(0,\sigma^2)$. Then, the received pre-processed $\boldsymbol{\psi}_t$-correlated signal at the BS is (BJÖRNSON *et al.*, 2017a)

$$\begin{aligned}\mathbf{y}_t =& \mathbf{Y}\frac{\boldsymbol{\psi}_t^*}{||\boldsymbol{\psi}_t||} = \sum_{i\in\mathcal{S}_t}\sqrt{\rho_i}||\,\boldsymbol{\psi}_t||\,\mathbf{h}_i + \mathbf{n}_t \\ =& \sum_{i\in\mathcal{S}_t}\sqrt{\rho_i\,\tau_\mathrm{p}}\,\mathbf{h}_i + \mathbf{n}_t,\end{aligned} \tag{5}$$

where $\mathbf{y}_t \in \mathbb{C}^{M\times 1}$ and $\mathbf{n}_t \sim \mathcal{CN}(0,\sigma^2\mathbf{I}_M)$ is the effective receiver noise, with σ^2 as the noise variance.

Then, the signal gain received at the BS during the transmission of pilot sequence $\boldsymbol{\psi}_t$, denoted by α_t, is given by (BJÖRNSON *et al.*, 2017a)

$$\alpha_t = \sum_{i\in\mathcal{S}_t}\rho_i\,\beta_i\,\tau_\mathrm{p}. \tag{6}$$

Next, in Step 2, the BS responds to the users from Step 1 by sending downlink (DL) pilots $\boldsymbol{\phi}_t \in \mathbb{C}^{\tau_\mathrm{p}}$, where the sequences $\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_{\tau_\mathrm{p}} \in \mathbb{C}^{\tau_\mathrm{p}}$ are mutually orthogonal and satisfy $\|\boldsymbol{\phi}_t\| = \sqrt{\tau_\mathrm{p}}$. In addition, the BS signal will be precoded, generating the downlink pilot

signal

$$\mathbf{V} = \sqrt{q} \sum_{t=1}^{\tau_{\mathrm{p}}} \frac{\mathbf{y}_t^*}{\|\mathbf{y}_t\|} \boldsymbol{\phi}_t^T, \tag{7}$$

where $\mathbf{V} \in \mathbb{C}^{M \times \tau_{\mathrm{p}}}$ is the precoded dowlink signal and q is the BS transmit power per pilot.

Thus, the signal received at user $k \in \mathcal{S}_t$ is $\mathbf{z}_k^T$ denoted by

$$\mathbf{z}_k^T = \mathbf{h}_k^T \mathbf{V} + \boldsymbol{\eta}_k^T, \tag{8}$$

with $\mathbf{z}_k \in \mathbb{C}^{\tau_{\mathrm{p}}}$ and where $\boldsymbol{\eta}_k^T \sim \mathcal{CN}\left(0, \sigma^2 \mathbf{I}_{\tau_{\mathrm{p}}}\right)$ is the noise. After correlating $\mathbf{z}_k^T$ with $\boldsymbol{\phi}_t$ the user calculates

$$z_k = \mathbf{z}_k^T \frac{\boldsymbol{\phi}_t^*}{\|\boldsymbol{\phi}_t\|} = \sqrt{q\,\tau_{\mathrm{p}}}\, \mathbf{h}_k^T \frac{\mathbf{y}_t^*}{\|\mathbf{y}_t\|} + \zeta_k, \tag{9}$$

where $\zeta_k = \boldsymbol{\eta}_k \frac{\boldsymbol{\phi}_t^*}{\|\boldsymbol{\phi}_t\|} \sim \mathcal{CN}\left(0, \sigma^2\right)$.

In order to estimate α_t for each user k, we follow (BJÖRNSON *et al.*, 2017a) to solve $\Re(z_k) = \mathbb{E}\{z_k\}$ with respect to α_t, so that

$$\hat{\alpha}_{t,k} = \max\left\{ \left(\frac{\Gamma(M + \frac{1}{2})}{\Gamma(M)} \right)^2 \frac{q \rho_k \beta_k^2 \tau_{\mathrm{p}}^2}{(\Re(z_k))^2} - \sigma^2,\, \rho_k \beta_k \tau_{\mathrm{p}} \right\}. \tag{10}$$

Next, at Step 3 the pilot collisions are solved in a distributed manner, in which only one user repeats the pilot transmission when collision resolution is successful. Since each user knows its own average CSI, obtained in Step 0, and estimates $\hat{\alpha}_{t,k}$ during Step 2, then, each user detects if a collision has occurred when these estimates considerably diverge. In order to solve the collision, the user k defines how strong is its own signal relative to the sum of all the contenders, by employing the following decision rule

$$\mathcal{R}_k: \quad \rho_k\, \beta_k\, \tau_{\mathrm{p}} > \frac{\hat{\alpha}_{t,k}}{2} + \epsilon_k \quad \text{(repeat)} \tag{11}$$

$$\mathcal{I}_k: \quad \rho_k\, \beta_k\, \tau_{\mathrm{p}} \leq \frac{\hat{\alpha}_{t,k}}{2} + \epsilon_k \quad \text{(inactive)} \tag{12}$$

If $\mathcal{R}_k$ is true, the k-th user concludes that it has the strongest signal gain, and it repeats the pilot $\boldsymbol{\psi}_t$ in Step 3. Otherwise, if $\mathcal{I}_k$ is true, then the k-th user remains inactive by pulling out the RA attempt and trying to communicate later. Let us notice that the bias term, $\epsilon_k \in \mathbb{R}$, is a variable that can be appropriate tuned in order to reach some performance criteria, *e.g.*, to maximize the average number of resolved collisions or to minimize the probability of false negatives. The bias term ϵ_k is defined as (BJÖRNSON *et al.*, 2017a)

$$\epsilon_k = \frac{\delta \beta_k \tau_{\mathrm{p}}}{\sqrt{M}}. \tag{13}$$

By employing the distributed decision rule, the probability that the k-th user decides to repeat its pilot transmission is given by (BJÖRNSON *et al.*, 2017a)

$$\Pr\{\mathcal{R}_k\} = 1 - \Pr\left\{\Re(z_k) \leq \sqrt{\vartheta_k}\right\} + \Pr\left\{\Re(z_k) \leq -\sqrt{\vartheta_k}\right\} \tag{14}$$

where

$$\vartheta_k = \left(\frac{\Gamma(M+\frac{1}{2})}{\Gamma(M)}\right)^2 \frac{q\,\rho_k\,\beta_k^2\,\tau_{\mathrm{p}}^2}{\sigma^2 + 2(\rho_k\beta_k\tau_{\mathrm{p}} - \epsilon_k)} \tag{15}$$

and

$$\begin{aligned}\Pr\{\Re(z_k) \leq b\} = \mathcal{Q}\left(-\frac{b\sqrt{2}}{\sqrt{\lambda_2}}\right) - \sum_{k=0}^{M-1} \frac{e^{-\frac{b^2}{\lambda_2}\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)}}{\Gamma(k+1)\lambda_1^k\sqrt{\pi\lambda_2}} \\ \times \sum_{n=0}^{2k}\binom{2k}{n}\frac{\Gamma(\frac{n+1}{2}) + c_n(b)\,\gamma(\frac{n+1}{2}, \frac{b^2}{\lambda_2}\frac{\lambda_1}{\lambda_1+\lambda_2})}{2\left(\frac{b}{\lambda_2}\right)^{n-2k}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right)^{2k+\frac{1}{2}-\frac{n}{2}}},\end{aligned} \tag{16}$$

with $\lambda_1 = \frac{\rho_k q \beta_k^2 \tau_{\mathrm{p}}^2}{\alpha_t + \sigma^2}$ and $\lambda_2 = \sigma^2 + q\beta_k\tau_{\mathrm{p}} - \lambda_1$. Moreover, $\mathcal{Q}(\cdot)$ denotes the $\mathcal{Q}$-function (ABRAMOWITZ; STEGUN, 1974, §26.2.3), $\gamma(\cdot,\cdot)$ is the incomplete gamma function (ABRAMOWITZ; STEGUN, 1974, §6.5.2), and

$$c_n(x) = \begin{cases} (-1)^n & x \geq 0 \\ -1 & x < 0 \end{cases}. \tag{17}$$

Then, the probability of contention resolution for the set $|\mathcal{S}_t|$ is given by

$$\begin{aligned}P_{\mathrm{CR}}^{\mathrm{SUCRe}} = &\Pr\{\mathcal{R}_1,\mathcal{I}_2,\ldots,\mathcal{I}_{|\mathcal{S}_t|}\} + \\ &\Pr\{\mathcal{I}_1,\mathcal{R}_2,\mathcal{I}_3,\ldots,\mathcal{I}_{|\mathcal{S}_t|}\} + \cdots + \\ &\Pr\{\mathcal{I}_1,\ldots,\mathcal{I}_{|\mathcal{S}_t|-1},\mathcal{R}_{|\mathcal{S}_t|}\}.\end{aligned} \tag{18}$$

In the case of successful collision resolution, the BS receives the repeated RA pilot transmissions from the user with the strongest signal gain, followed by the UL messages (user identity and a request for payload transmission).

Finally, at Step 4, the BS uses the pilot signal to estimate the channel to the user, or multiple users, that sent the pilot $\boldsymbol{\psi}_t$ in Step 3, and tries to decode the corresponding message. If the decoding is successful, the BS has identified one of the users in $\mathcal{S}_t$ and admits it to the payload coherence blocks by allocating a dedicated data pilot sequence. If the decoding fails, the SUCRe protocol has failed to resolve the collision.

The users that were not admitted in Step 4 can still try to transmit new RA pilots in a next block, after a random waiting time with a new access probability P_{an}. If the user has not succeeded to access the BS after a maximal of 10 attempts, then it stops sending pilot sequences and declares failed access attempt.

3.1 SYSTEM SUM-RATE

In order to define the sum-rate metric for the SUCRe protocol, we consider the transmissions from the BS to users and from the users to BS reciprocal, occurring with a fixed constant power $\rho_k = q, \forall k$. Then, the signal-to-noise ratio (SNR) at the BS with respect to the k-th user after Step 4 ($\mathrm{SNR}_k^{\mathrm{SUCRe}}$) is

$$\mathrm{SNR}_k^{\mathrm{SUCRe}} = \frac{\rho_k \beta_k}{\sigma^2}, \tag{19}$$

so that the available data rate of the k-th user (R_k^{SUCRe}) is

$$R_k^{\mathrm{SUCRe}} = \log_2\left(1 + \frac{\rho_k \beta_k}{\sigma^2}\right), \tag{20}$$

yielding the following sum-rate ($R_{\mathrm{sum}}^{\mathrm{SUCRe}}$) for the SUCRe protocol

$$R_{\mathrm{sum}}^{\mathrm{SUCRe}} = \sum_{t=1}^{\tau_{\mathrm{p}}} \sum_{k \in \mathcal{S}_t} \frac{R_k^{\mathrm{SUCRe}}}{A_k}, \tag{21}$$

in which we compute the individual data rate of each user, R_k^{SUCRe}, normalized by the delay A_k, given in terms of the number of attempts of that user to access the BS. For example, if the k-th user that transmits the t-th pilot secures access to the BS only in the third attempt, $A_k = 3$, and the data rate is computed as $\frac{R_k^{\mathrm{SUCRe}}}{3}$. Furthermore, if after 10 attempts the user has failed to access the channel, the individual data rate of this user is not computed. In addition, the sum-rate takes each pilot sequence $t \in \{1, 2, \ldots, \tau_{\mathrm{p}}\}$ into account, so that, for each t, a different set $\mathcal{S}_t$ of contending users is considered.

3.2 AVERAGE DELAY

Another important system metric is given by the average number of attempts to access the BS, which we denote by the average delay $\mathcal{A}^{\mathrm{SUCRe}} = \mathbb{E}\{A_k\}$, so that

$$\mathcal{A}^{\mathrm{SUCRe}} = \frac{\displaystyle\sum_{t=1}^{\tau_{\mathrm{p}}} \sum_{k \in \mathcal{S}_t} A_k}{\displaystyle\sum_{t=1}^{\tau_{\mathrm{p}}} |\mathcal{S}_t|}. \tag{22}$$

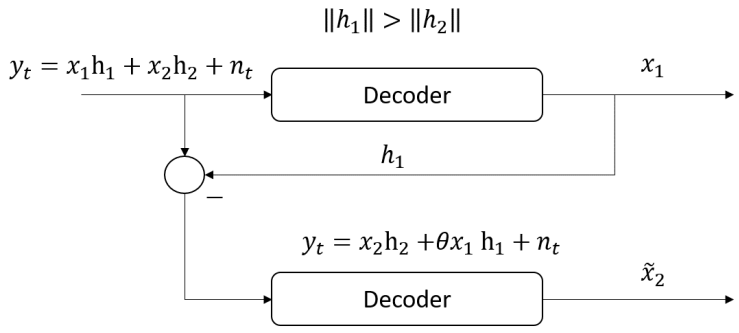
In the case that an user exceeds 10 attempts, the access is considered unsuccessful and its individual delay is accounted as $A_k = 10$ in (22).

4 NOMA-RA PROTOCOL

In order to optimize the performance of the SUCRe protocol, we propose a novel scheme, denoted by NOMA-RA, which exploits the power domain for simultaneous access of multiple users served at the same time, frequency, and pilot. To that end, a SIC strategy is employed at the BS in order to decode the superposition of the transmission from users with the same pilot sequence.

The SIC process is illustrated in Fig. 4, where the signal with highest channel gain is decoded first and subtracted from the overall received signal. In the sequel, the signal with the second highest channel gain is decoded and also subtracted from the remaining of the overall signal, and so on. Fig. 4 illustrates the process for two superimposed symbols, x_1 and x_2, subjected to the channel coefficient vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, respectively, with $||\mathbf{h}_1|| > ||\mathbf{h}_2||$, x_1 and x_2 can be considered as pilot sequences at the random access phase or data at the payload data transmission phase. Also, we assume that SIC may not be perfect, so a fraction $0 \leq \theta \leq 1$ of the first signal may still appear as interference at the second decoding step, with $\theta = 0$ representing the perfect SIC scenario.

Figure 4 – SIC decoder considering two superimposed signals, where the fraction $\theta x_1 \mathbf{h}_1$, with $0 \leq \theta \leq 1$, is the residue due the imperfect cancellation.



Source: The Author

Our aim is to exploit NOMA after Step 2 of the SUCRe protocol, which occurs with the BS receiving the superposition of signals from the users who declared themselves winners at the contention resolution phase. The steps of the proposed NOMA protocol are similar to those of SUCRe, except for Step 3, which we detail in the sequel.

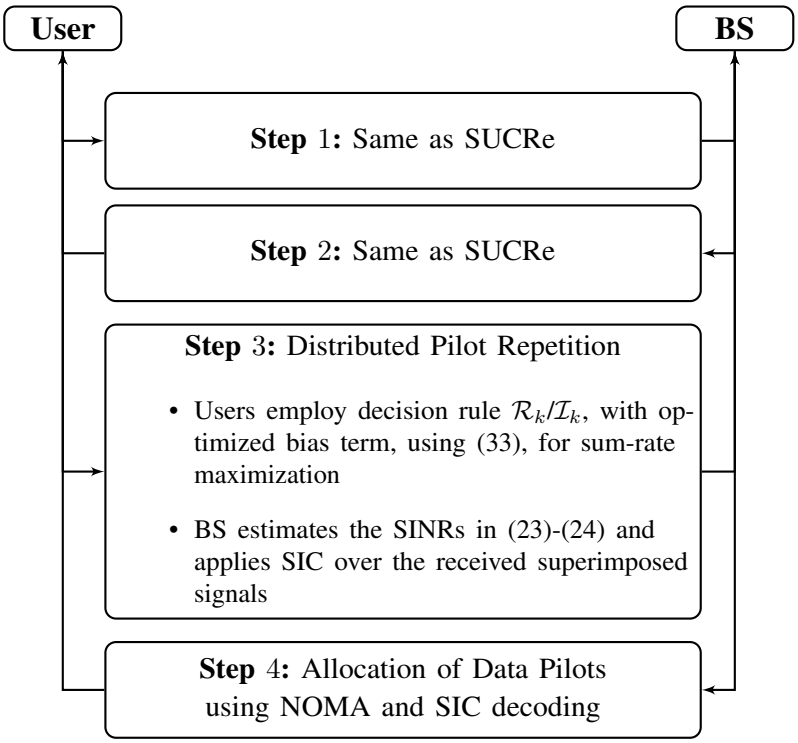
Considering that most of the collisions occur between two users choosing the same pilot, and since interference after SIC is not always perfectly canceled, which limits performance under multiple simultaneous transmissions, we consider that only the cases where at most two users decide to retransmit in Step 3 can be solved through NOMA. Then, the signal-to-interference-plus-noise ratio (SINR) related to these two users are used to indicate the SIC decoding order, which can be written as

$$\mathrm{SINR}_1^{\mathrm{NOMA-RA}} = \frac{\rho_1 \beta_1 \tau_\mathrm{p}}{\rho_2 \beta_2 \tau_\mathrm{p} + \sigma^2}, \tag{23}$$

$$\mathrm{SINR}_2^{\mathrm{NOMA-RA}} = \frac{\rho_2 \beta_2 \tau_\mathrm{p}}{\theta \rho_1 \beta_1 \tau_\mathrm{p} + \sigma^2}, \tag{24}$$

where we assume $\rho_1 \beta_1 \tau_\mathrm{p} > \rho_2 \beta_2 \tau_\mathrm{p}$, and that a fraction θ of $\rho_1 \beta_1 \tau_\mathrm{p}$ could not be perfectly cancelled in the first iteration of the SIC decoder, appearing at the denominator of (24).

Figure 5 – Steps of the NOMA-RA protocol highlighting the differences with respect to SUCRe.



Source: The Author

Then, keeping the nomenclature consistent with that used in the SUCRe protocol, we define *contention resolution* in the NOMA-RA protocol as the probability that one or two users

simultaneously declare themselves as winners in Step 3, which implies in

$$\begin{aligned} P_{\mathrm{CR}}^{\mathrm{NOMA-RA}} = & \Pr\{\mathcal{R}_1,\mathcal{I}_2,\ldots,\mathcal{I}_{|\mathcal{S}_t|}\}+ \\ & \Pr\{\mathcal{I}_1,\mathcal{R}_2,\mathcal{I}_3,\ldots,\mathcal{I}_{|\mathcal{S}_t|}\}+\cdots+ \\ & \Pr\{\mathcal{I}_1,\ldots,\mathcal{I}_{|\mathcal{S}_t|-1},\mathcal{R}_{|\mathcal{S}_t|}\}+ \\ & \Pr\{\mathcal{R}_1,\mathcal{R}_2,\mathcal{I}_3,\ldots,\mathcal{I}_{|\mathcal{S}_t|}\}+ \\ & \Pr\{\mathcal{R}_1,\mathcal{I}_2,\mathcal{R}_3,\ldots,\mathcal{I}_{|\mathcal{S}_t|}\}+\cdots+ \\ & \Pr\{\mathcal{I}_1,\cdots,\mathcal{R}_{|\mathcal{S}_t|-1},\mathcal{R}_{|\mathcal{S}_t|}\}. \end{aligned} \tag{25}$$

The steps of the NOMA-RA protocol are illustrated in Fig. 5, highlighting only the differences with respect to SUCRe. In addition, let us remark that the same Step 0 as in SUCRe is also needed for initial average CSI β_k estimation and synchronization.

4.1 CSI ESTIMATION

Estimating CSI is challenging in a NOMA setup when using the SIC algorithm. To that end, we follow (BJÖRNSON *et al.*, 2018), so that the channel estimation in the case that two users choose the same pilot sequence is given by the minimum mean square error (MMSE) estimator

$$\hat{\mathbf{h}}_k = \frac{1}{\sqrt{\rho}}\mathbf{R}_k\,\mathbf{Q}^{-1}\,\mathbf{Y}\,\boldsymbol{\psi}_t^*, \tag{26}$$

where $\mathbf{Y}$ is defined in (4), $\mathbf{R}_k$ is the covariance matrix of the channel, denoted by

$$\mathbf{R}_k = \beta_k \mathbf{I}_M \tag{27}$$

and $\mathbf{Q}$ is the normalized covariance matrix of the observation after correlating with the pilot sequence, which is denoted by

$$\mathbf{Q} = \frac{1}{\rho}\mathbb{E}\{\mathbf{Y}\boldsymbol{\psi}_t^*(\mathbf{Y}\boldsymbol{\psi}_t^*)^H\} = \mathbf{R}_1 + \mathbf{R}_2 + \frac{1}{\rho}\mathbf{I}_M. \tag{28}$$

In addition, the estimate $\hat{\mathbf{h}}_k$ and the estimation error $\tilde{\mathbf{h}}_k = \mathbf{h}_k - \hat{\mathbf{h}}_k$ are independent random vectors distributed as $\hat{\mathbf{h}}_k \sim \mathcal{CN}(0, \boldsymbol{\Phi}_k)$ and $\tilde{\mathbf{h}}_k \sim \mathcal{CN}(0, \mathbf{R}_k - \boldsymbol{\Phi}_k)$ with $\boldsymbol{\Phi}_k = \mathbf{R}_k\,\mathbf{Q}^{-1}\,\mathbf{R}_k$.

Furthermore, according to (BJÖRNSON *et al.*, 2018, Remark 2) the MMSE estimator in (26) implies in knowing the channel statistics, since the BS can only compute $\hat{\mathbf{h}}_k$ if it knows the covariance matrices $\mathbf{R}_k$. One option to estimate $\mathbf{R}_k$ is by means of a regularized sample covariance matrix, given realizations of $\mathbf{h}_k$ observed in only noise (YIN *et al.*, 2013), or where

some observations are regular pilot transmissions containing the desired channel plus noise and some contain only the noise (BJÖRNSON *et al.*, 2016). Moreover, (BJÖRNSON *et al.*, 2016) has shown that the covariance information is relatively easy to acquire.

In addition, another option is that the user provides side information about their estimates of β_k, obtained in Step 0, during the pilot repetition in Step 3. In this case, the SIC decoding can be performed in two steps. We can first approximate the covariance matrix of the channel of the strongest user by the normalized covariance matrix $\mathbf{Q}$, *i.e.*,

$$\mathbf{R}_1 \approx \frac{1}{\rho}\mathbb{E}\{\mathbf{Y}\boldsymbol{\psi}_t^*(\mathbf{Y}\boldsymbol{\psi}_t^*)^H\}. \tag{29}$$

Although this estimate may be coarse, as the transmission in Step 3 of the RA process is designed to be robust, it is very likely to be sufficient to allow the BS to obtain the information of β_1, transmitted by the strongest user in the payload concatenated to the pilot transmission. With β_1 acquired, then MMSE estimation for the first iteration of the SIC decoder can be performed using (26), while $\mathbf{R}_2$ can be obtained after subtracting $\mathbf{R}_1$ and $\frac{1}{\rho}\mathbf{I}_M$ from $\mathbf{Q}$. Nevertheless, a deeper discussion on the particular method for channel estimation needs to be explored as a future investigation.

4.2 SYSTEM SUM-RATE AND AVERAGE DELAY IN NOMA-RA

In order to write the sum-rate for the NOMA-RA protocol, we assume that both users accessing the BS successfully with NOMA-RA will be granted the same pilots for data transmission, after Step 4. Thus, we first define their achievable data rates at this point $R_1^{\mathrm{NOMA-RA}}$ and $R_2^{\mathrm{NOMA-RA}}$, which can be written as

$$R_1^{\mathrm{NOMA-RA}} = \log_2\left(1 + \frac{\rho_1\beta_1}{\rho_2\beta_2 + \sigma^2}\right), \tag{30}$$

$$R_2^{\mathrm{NOMA-RA}} = \log_2\left(1 + \frac{\rho_2\beta_2}{\theta\rho_1\beta_1 + \sigma^2}\right), \tag{31}$$

while in the case of the retransmission of a single user, the achievable data rate is given by (20).

Then, the sum-rate of NOMA-RA ($R_{\mathrm{sum}}^{\mathrm{NOMA-RA}}$) is defined as

$$R_{\mathrm{sum}}^{\mathrm{NOMA-RA}} = \sum_{t=1}^{\tau_{\mathrm{P}}}\sum_{k=1}^{|\mathcal{S}_t|}\frac{R_k^{\mathrm{NOMA-RA}}}{A_k}, \tag{32}$$

in which the number of attempts to access the BS is also limited to $A_k = 10$. As can be observed, the SUCRe achievable rate and sum-rate, from equations (20) and (21), present a lower

performance compared to (30) and (32), since NOMA considers the sum of the achievable rates of two users, whereas only one user per pilot is accounted for the achievable rate of SUCRe.

In addition, the average delay of NOMA-RA, denoted by $\mathcal{A}^{\mathrm{NOMA-RA}}$, can be obtained similarly as in (22). Nevertheless, a smaller average delay is expected for the NOMA-RA scheme compared to SUCRe, since two users may simultaneously access the BS with the same pilot.

4.3 BIAS TERM

The bias term defined in (13) adds $\delta \cdot \tau_{\mathrm{p}}$ standard deviations around the users path-loss, β_k, in which δ is a parameter that can be tuned to improve some system metric. For instance, δ has been numerically optimized in (BJÖRNSON *et al.*, 2017a) in order to maximize the probability of collision resolution.

Nevertheless, two users choosing the same pilot can be very beneficial in terms of sum-rate for the NOMA-RA protocol, while disadvantageous regarding SUCRe. As a consequence, in the numerical comparisons we consider the optimal $\delta^{\star}_{(\mathrm{sch})}$ that maximizes the sum-rate of the scheme $(\mathrm{sch}) \in \{\mathrm{SUCRe}, \mathrm{NOMA-RA}\}$, *i.e.*,

$$\delta^{\star}_{(\mathrm{sch})} = \operatorname{argmax} R^{(\mathrm{sch})}_{\mathrm{sum}}. \tag{33}$$

Due to the complexity of the involved expressions, $\delta^{\star}_{(\mathrm{sch})}$ is obtained numerically, using an exhaustive search approach. Moreover, we assume that the same optimized $\delta^{\star}_{(\mathrm{sch})}$ is employed by all users at the same time. Therefore, in a practical setup, the BS is responsible for optimizing it, using for instance some automatic learning approaches, as well as broadcasting this configuration to all users.

5 NUMERICAL RESULTS

In order to provide some numerical examples comparing the SUCRe and NOMA-RA protocols, we consider the simulation parameters listed in Table 3.

Table 3 – Simulation Parameters

Parameter	Value
Radius of the cell	$R = 250$ m
Minimum radius with respect to the BS	$R_{Min} = 25$ m
Number of BS antennas	$M = 100$
BS/Users transmit power	$q = \rho_k = 27$ dBm
Noise power	$\sigma^2 = -98.65$ dBm
Number of inactive users	$K_0 \in [0; 20000]$
Bias term	$\delta = [-16 : -2 : 0]$
Access probability (1st attempt)	$P_{\mathrm{a}} = 0.001$
Access probability (new attempts)	$P_{\mathrm{an}} = 0.5$
Number of available pilot sequences	$\tau_{\mathrm{p}} = 10$
Channel estimation error	$\sigma_{\varepsilon}^2 \in \{0, 0.5\}$
SIC imperfection factor	$\theta \in \{0, 0.1\}$

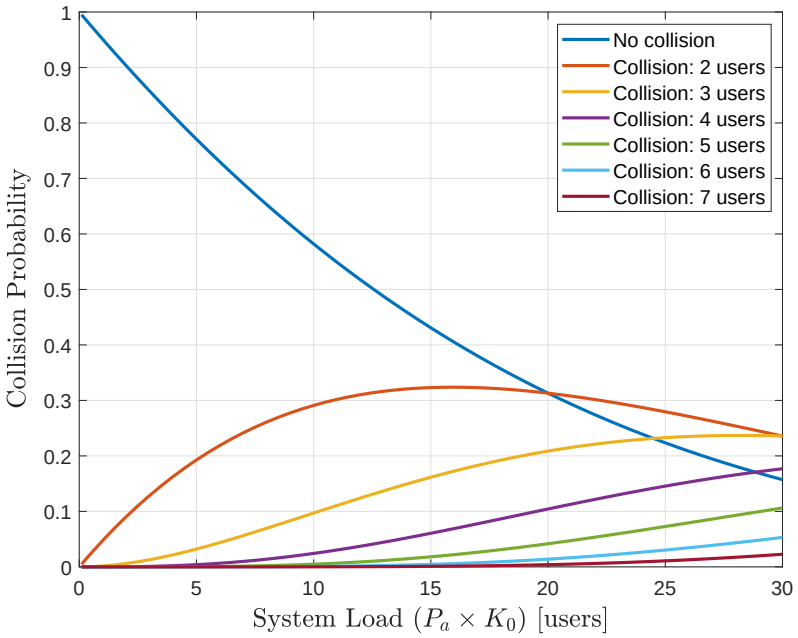
Source: The Author

First, Fig. 6 plots the probability considering 1 (no collision) to 7 users simultaneously choosing the same pilot, as a function of the system load ($P_{\mathrm{a}} \times K_0$) using the SUCRe protocol. As we can observe, the probability of no collision, when $|\mathcal{S}_t| = 1$, combined with the probability that only 2 users choose the same pilot sequence represent the majority of the cases, which justifies the proposed approach considering NOMA for solving collisions with two users transmitting simultaneously the same pilot signal.

Next, Fig. 7 plots the probabilities of collision resolution and false positives for SUCRe and NOMA-RA protocols, as a function of δ. A false positive happens when two users choose the same pilot in the case of SUCRe, or when three or more choose the same pilot in NOMA-RA. We also consider the NOMA-RA protocol with both perfect CSI and using the MMSE estimator using (26)[1]. As we notice, the NOMA-RA protocol shows better performance in terms of resolved collisions, taking advantage of the cases when two users choose the same pilot, which is treated as a false positive in the SUCRe protocol. Thus, we also observe that the probability of false positives in the NOMA-RA protocol decreases considerably compared to SUCRe. Furthermore, we also observe that the use of the MMSE estimator has a small impact in the NOMA-RA performance in terms of collision resolution. Additionally, we observe that an appropriate choice

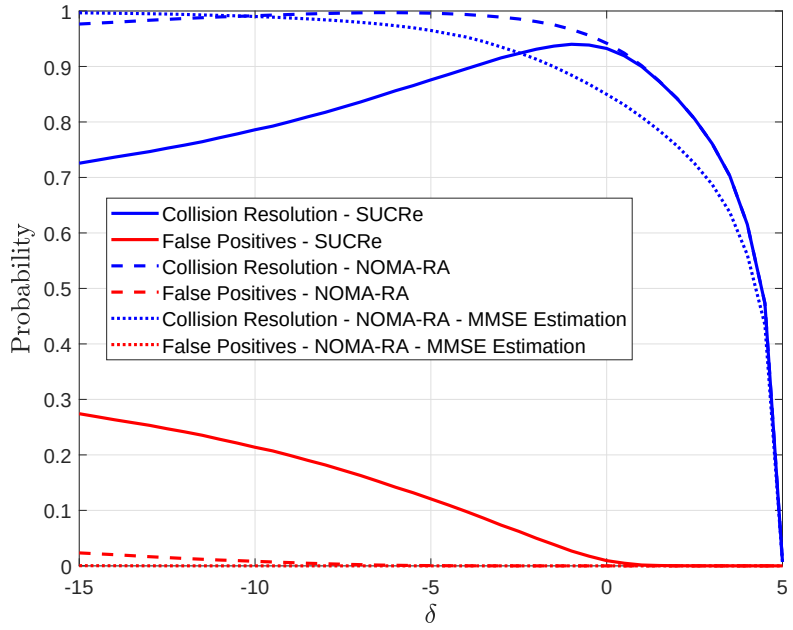
[1] Let us remark that the imperfect SIC decoding impacts only the SNR of the NOMA-RA protocol, and thus θ has no effect on the collision resolution and false positive probabilities in Fig 7.

Figure 6 – Collision probability from 1 (no collision) to 7 users simultaneously choosing the same pilot, as a function of the system load $P_{\mathrm{a}} \times K_0$.



Source: The Author

Figure 7 – Probability of collision resolution and false positives as a function of δ for the SUCRe and NOMA-RA protocols, considering $K_0 = 20000$, perfect CSI and MMSE estimator.

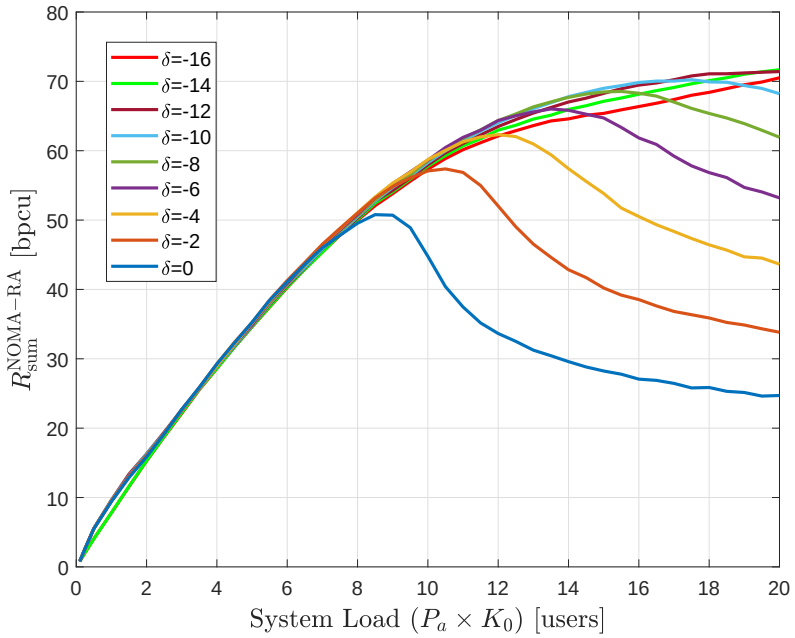


Source: The Author

of δ is important in order to optimize the desired probability metric.

Fig. 8 plots the sum-rate of the NOMA-RA protocol as a function of $P_{\mathrm{a}} \times K_0$ for $\delta \in [-16, 0]$ and $\theta = 0$. We can observe that the sum-rate is maximized with different values of

Figure 8 – Sum-rate of the NOMA-RA protocol as a function of the system load ($P_{\mathrm{a}} \times K_0$) for $\delta \in [-16, 0]$ and $\theta = 0$.



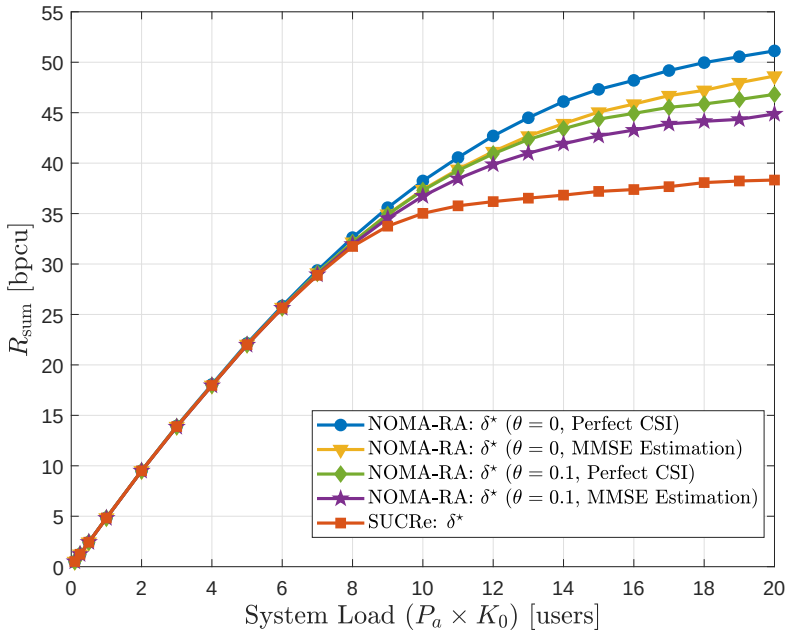
Source: The Author

δ depending on K_0. When $P_{\mathrm{a}} \times K_0 < 8$, the performance is approximately the same regardless of δ. Nevertheless, when the system load increases, the optimization of δ becomes crucial since the number of pilot collisions also increases with $P_{\mathrm{a}} \times K_0$.

Considering the optimization of the bias term, Fig. 9 plots the sum-rate for SUCRe and NOMA-RA protocols as a function of the system load by finding the optimal $\delta^{\star}_{(\mathrm{sch})}$. In addition, for the NOMA-RA protocol we also consider perfect and imperfect SIC decoders, with $\theta \in \{0, 0.1\}$, as well as perfect CSI and the MMSE estimator. As we observe, the NOMA-RA protocol outperforms SUCRe in terms of sum-rate, especially when the system load is higher. For instance, when $P_{\mathrm{a}} \times K_0 = 20$ (*i.e.*, $K_0 = 20000$) the proposed NOMA-RA protocol with perfect CSI yields a sum-rate 33% higher than SUCRe with $\theta = 0$ and 22% higher with $\theta = 0.1$. When the MMSE estimator is considered, a small degradation of around 5% is observed comparing it to the scenario with perfect CSI, still leading to a considerable gain with respect to SUCRe.

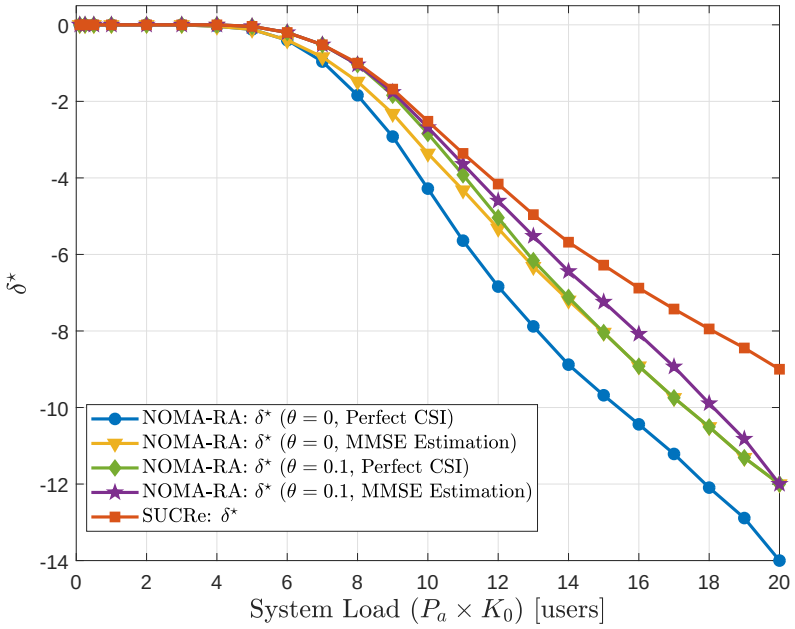
Fig. 10 complements the analysis by showing the optimal $\delta^{\star}_{(\mathrm{sch})}$ that maximizes the sum-rate for both NOMA-RA and SUCRe protocols, as a function of the system load. As we observe, the optimal $\delta^{\star}_{(\mathrm{sch})}$ decreases when the system load increases for all schemes. Nevertheless, $\delta^{\star}_{\mathrm{NOMA-RA}}$ is usually lower than $\delta^{\star}_{\mathrm{SUCRe}}$, once the NOMA-RA scheme leverages from pilot collisions between two users, which is especially important when the system load increases. In

Figure 9 – Sum-rate with optimized $\delta^{\star}_{(\text{sch})}$ for SUCRe and NOMA-RA protocols as a function of the system load. Perfect and imperfect SIC decoders are considered, with $\theta \in \{0, 0.1\}$, as well as perfect CSI and the MMSE estimator.



Source: The Author

Figure 10 – $\delta^{\star}_{(\text{sch})}$ that maximizes the sum-rate of NOMA-RA and SUCRe protocols, as a function of the system load, for $\theta \in \{0, 0.1\}$, perfect CSI and the MMSE estimator.



Source: The Author

addition, we also observe that $\delta^{\star}_{\text{NOMA}-\text{RA}}$ slightly increases when either imperfect SIC or MMSE estimation are considered.

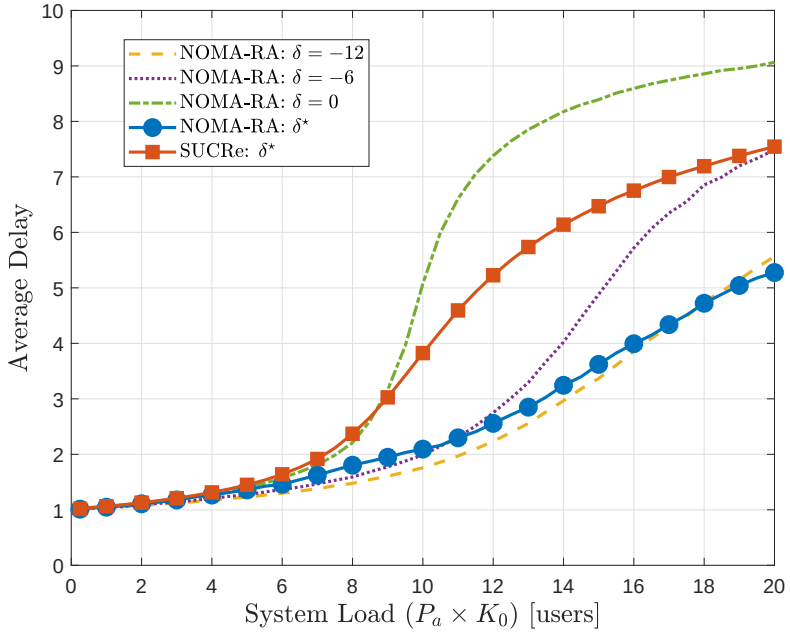
Table 4 – Minimum percentage of the sum-rate achieved by the NOMA-RA schemes with $\delta_{\mathrm{NOMA-RA}} = -12$ compared to that with $\delta^{\star}_{\mathrm{NOMA-RA}}$.

SIC and CSI configuration	Percentage of the sum-rate
$\theta = 0$ and perfect CSI	97.49%
$\theta = 0$ and MMSE estimation	95.32%
$\theta = 0.1$ and perfect CSI	95.04%
$\theta = 0.1$ and MMSE estimation	91.99%

Source: The Author

In addition, in order to illustrate how the optimization of δ impacts the overall performance of the NOMA-RA scheme, we compare the sum-rate achieved by NOMA-RA using a fixed δ with the optimal one. As indicated by Fig. 8, $\delta_{\mathrm{NOMA-RA}} = -12$ may be an interesting choice, since it exhibits a near-optimal performance in the range of the evaluated system load. Therefore, Table 4 shows the minimum percentage of the sum-rate of the NOMA-RA protocol that is achieved when $\delta_{\mathrm{NOMA-RA}} = -12$ is used instead of the optimal $\delta^{\star}_{\mathrm{NOMA-RA}}$. As we observe, $\delta_{\mathrm{NOMA-RA}} = -12$ is a suitable trade-off choice in the case of perfect SIC and perfect CSI. Nevertheless, the sum-rate decreases considerably when θ increases or the MMSE estimator is employed, which is more representative of a real deployment.

Figure 11 – Average delay of SUCRe and NOMA-RA schemes as a function of the system load, with $\theta = 0$, perfect CSI, $\delta_{\mathrm{SUCRe}} = \delta^{\star}$ and $\delta_{\mathrm{NOMA-RA}} \in \{-12, -6, 0, \delta^{\star}\}$.

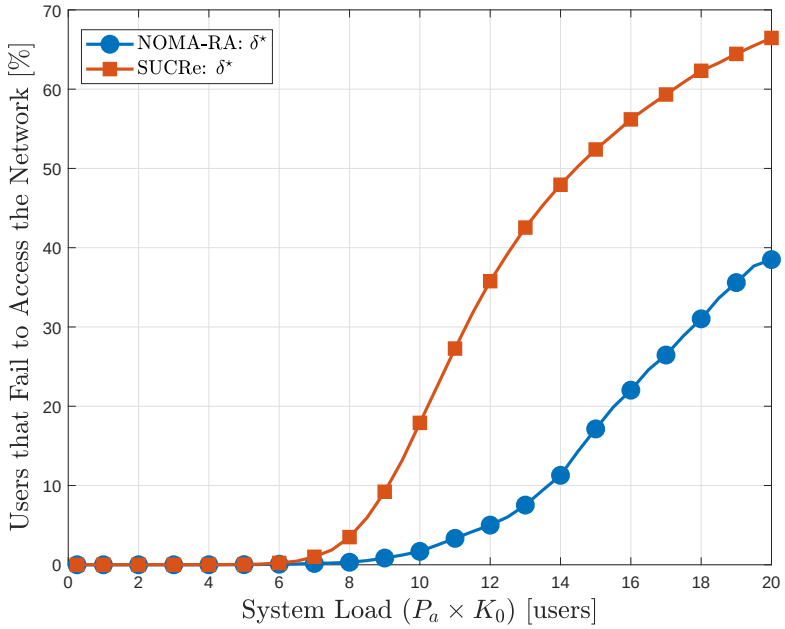


Source: The Author

Fig. 11 plots the average transmission delay of SUCRe and NOMA-RA schemes, respectively $\mathcal{A}^{\mathrm{SUCRe}}$ and $\mathcal{A}^{\mathrm{NOMA-RA}}$, as a function of the system load. Moreover, Fig. 11 only

considers the case of $\theta = 0$ and perfect CSI, once these parameters do not affect directly the average delay, while $\delta_{\text{SUCRe}} = \delta^{\star}$ and $\delta_{\text{NOMA-RA}} \in \{-12, -6, 0, \delta^{\star}\}$. First, we notice that the average delay increases with the system load, with the NOMA-RA scheme achieving smaller delays compared to SUCRe when $\delta_{\text{NOMA-RA}} < 0$ for all system loading considered. For example, when $P_{\text{a}} \times K_0 = 20$ the average delay with NOMA-RA is 30% smaller than that of SUCRe. Furthermore, the optimization of δ impacts the average delay, even in the case of perfect CSI and perfect SIC. As we can observe, $\delta^{\star}_{\text{NOMA-RA}}$ has been optimized to maximize the sum-rate, and not to minimize the average delay. Therefore, $\delta^{\star}_{\text{NOMA-RA}}$ slightly increases the delay compared to, *e.g.*, $\delta_{\text{NOMA-RA}} = -12$, depending on the system load. Fig. 11 also reinforces the importance of the optimization of δ in the collision resolution. For instance, for $\delta = 0$, (when the δ parameter does not cause any influence in the decision rule) and $P_{\text{a}} \times K_0 > 8$, the average delay presents a high increase.

Figure 12 – Average number of users that fail to access the network, for the NOMA-RA and SUCRe protocols using $\delta^{\star}_{\text{NOMA-RA}}$ and $\delta^{\star}_{\text{SUCRe}}$, as a function of the system load.



Source: The Author

Finally, to better understand the issue of users that exceed the maximum connection attempts in the system performance, Fig. 12 plots the average number of users that fail to access the network as a function of the system load, considering $\delta^{\star}_{\text{NOMA-RA}}$ and $\delta^{\star}_{\text{SUCRe}}$. As we can observe, the NOMA-RA protocol presents a reduced average number of failed access attempts, mainly in crowded and over-crowded scenarios. For instance, when $P_{\text{a}} \times K_0 = 20$, 66.5% of

users fail to access the network with SUCRe, while 38.5% fail with the proposed NOMA-RA scheme.

6 CONCLUSIONS AND FINAL CONSIDERATIONS

In addition to the massive growth of MTC devices, the number of HTC-type connections has also grown considerably, consequently increasing exponentially the demand for network access. Considering this context, this work proposes a new form of grant-based random access based on the SUCRe protocol. In addition, this work investigated the system performance in terms of the sum-rate and average delay.

The proposed random access protocol is named NOMA-RA and employs NOMA with SIC at the BS in order to allow two users to communicate using the same pilot sequence, time slot and frequency. Furthermore, the bias term associated with the distributed collision resolution is the mechanism that encourage a user to retransmit when a collision occurs, so it is optimized numerically in order to maximize the sum-rate. In order to consider the analysis more realistic, the proposed NOMA-RA protocol was also investigated considering imperfect SIC and imperfect CSI estimation. In relation to imperfect SIC, it was considered that the interference from the strongest user is not totally canceled, while imperfect CSI considers an approximation for the MMSE estimation.

Our results show that the proposed NOMA-RA protocol for massive MIMO systems can decrease the average delay in comparison with the SUCRe protocol, while gains in terms of sum-rate are obtained. For instance, in a scenario when 10 pilot sequences are available and 20 devices compete to access the channel, comparing with the SUCRe protocol, the average delay is reduced by 30% while the sum-rate is improved by 33% and 22% considering, respectively, perfect and imperfect SIC decoders. In addition, the percentage of users that fail to access the BS in this situation drops from 66.5% with SUCRe to 38.5% with the proposed NOMA-RA scheme.

6.1 FUTURE WORKS

Despite the significant research efforts dedicated to facilitating massive access in 5G and beyond-5G wireless networks, there are still many challenging issues that worth of future investigation. In the following, we discuss some future research directions of this work.

6.1.1 Further Improvements on NOMA-RA

Considering the proposed work, the number of contenders, $|\mathcal{S}_t|$, is unknown, so that a user can only compare its own signal gain with the summation of the gains of its contenders $\hat{\alpha}_{t,k}$ in (10). Then, we encourage the two strongest users to retransmit in the case of a collision, by numerically optimizing the bias term. However, having the two strongest users retransmitting may not be optimal in the point of view of NOMA, since a certain difference between both SNRs may improve performance. Thus, in order to avoid using the bias term in the analysis, another technique that can help to improve the sum-rate, and still remain in a decentralized way, is to employ different thresholds so that the strongest user may be paired with another user with lower instantaneous SNR. In some recent analysis, we have observed that changing the decision rule in (11)-(12) to $0.45\ \hat{\alpha}_{t,k}$ for the strongest user and defining a window between $0.22\ \hat{\alpha}_{t,k}$ and $0.45\ \hat{\alpha}_{t,k}$ for the weaker user may worth of future investigation. However, it is still necessary a good mathematical analysis to verify it and propose further enhancements.

Another proposal in this sense is to exploit some processing at the BS, in order to pair which users should perform NOMA. For instance, by extending the proposed MMSE estimation for the CSI acquisition, the BS could possibly form the best combination of users that should retransmit, in case multiple devices choose the same pilot. In principle, this strategy could even extend the possibility to perform NOMA with more than two users, if the result is interesting in terms of sum-rate maximization, for example.

6.1.2 Power Allocation and Beamforming

The use of power allocation techniques has been shown as an efficient way to improve the overall network performance. The authors in (ISLAM *et al.*, 2017) focus on potentials and challenges of NOMA in 5G systems and showed that to obtain the maximum benefits offered by NOMA, a perfect superposition coding at the transmitter and error-free SIC at the receiver, optimum power allocation, QoS-oriented user fairness, appropriate user pairing, and good link adaptation are also required. In addition the authors presented a discussion of several important issues, such as dynamic user pairing, distortion analysis, interference analysis, resource allocation, heterogeneous networks, carrier aggregation, and transmit antenna selection, are expected to facilitate and provide a basis for further research on NOMA in 5G.

Another technique that improves the network performance with NOMA is called

beamforming. Several works have addressed either of these techniques, *e.g.*, in (ALI *et al.*, 2017; CUI *et al.*, 2018; SUN *et al.*, 2018) (CHEN *et al.*, 2017; NGUYEN *et al.*, 2021; MU *et al.*, 2020). These works have tried to maximize, for instance, the overall cell capacity (ALI *et al.*, 2017), the system utility (CUI *et al.*, 2018), or the sum-rate (SUN *et al.*, 2018) while guaranteeing a minimum required target rate for the users.

Considering the works cited above, power allocation and beamforming can be applied jointly to improve overall network performance. For a correct decoding of the signals received simultaneously at the BS, these signals need to have different power levels. Therefore, applying an optimal power allocation at the users will improve the decoding capability at the BS in Steps 1 and 3, as depicted in Fig. 5.

6.1.3 Intelligent Reflecting Surfaces

An intelligent reflecting surface (IRS) is a panel with a large number of low-cost passive reflecting elements, which can be smartly tuned to direct the radio signal propagation to a desired location or device. The IRS has the capacity of dynamically adjusting wireless channels to enhance the communication performance, and can be integrated to multiple access networks for boosting the spectrum/energy efficiency and enlarging network coverage/connections, constituting a good cost-effective solution.

Due to the resources and improvements that IRS can offer, in 5G and beyond-5G networks, it has been largely studied in recent literature (ZHENG *et al.*, 2020; ZUO *et al.*, 2020; MU *et al.*, 2020; MU *et al.*, 2021). In (ZHENG *et al.*, 2020) a theoretical performance comparison between NOMA and OMA is pursued in the IRS-assisted downlink communication. It is analyzed and numerically compared the minimum transmit powers required by different multiple access schemes. In (MU *et al.*, 2021) the authors considered three multiple access schemes (NOMA, FDMA and TDMA) and formulated a weighted sum rate maximization problem for joint optimization of the deployment location and the reflection coefficients of the IRS as well as the power allocation at the AP.

In order to maximize the system throughput, the authors in (ZUO *et al.*, 2020) formulated a joint optimization problem over the channel assignment, decoding order of NOMA users, power allocation, and reflection coefficients. The results showed that NOMA-IRS outperform the conventional NOMA and OMA-IRS. It is also shown that the choice of the IRS location is fundamental to enhance the overall system performance. In (MU *et al.*, 2020) IRS and

beamforming was addressed jointly in order to maximize the sum-rate of all users. The results showed the system sum-rate can be significantly improved with the proposed scheme and also achieve higher system sum rate than the IRS-aided OMA system.

In the works cited above, IRS has proven to be an excellent alternative to improve network performance, especially when combined with NOMA. Therefore, to merge the concepts of IRS with the proposed NOMA-RA scheme can improve the network performance, particularly when the IRS location is also optimized, as show in (MU *et al.*, 2021; ZUO *et al.*, 2020). Moreover, beamforming can be addressed jointly with IRS in order to maximize the system sum-rate, as in (MU *et al.*, 2020).

6.1.4 Cooperative NOMA

Over the past two decades, cooperative communications have been investigated and implemented for conventional OMA techniques, which are not enough to meet all the requirements of future generation networks (LIAQAT *et al.*, 2018). Cooperative communications and NOMA were first addressed in (DING *et al.*, 2015), where a SIC receiver is used to detect the multi-user signal. Similar to conventional cooperation, cooperative NOMA also relies on two phases, broadcast and cooperative. In the broadcast phase, the BS sends downlink messages to all users, whereas in the cooperative phase, the cooperating users transmit their signals via their short range communication channels, such as Bluetooth or ultra-wideband (UWB).

Recently, cooperative NOMA have gained considerable attention in the literature (KIM; LEE, 2015; WEI *et al.*, 2017; WAN *et al.*, 2018; LV *et al.*, 2018; KARA; KAYA, 2019b; KARA; KAYA, 2019a; LIU, 2019; JU *et al.*, 2019; LIU *et al.*, 2018a; XIE *et al.*, 2020). The relaying was considered from different ways at the analysis, such as with a fixed relay or by employing relay selection. Considering the works mentioned above, a possible future direction of this work is to adopt the concept of cooperative communications, in which the strongest user can act as a relay or not. Since the strongest user has better channel conditions and it is nearer from the BS than weakest user, it can be used as a relay to help the weakest user to delivery the information to the BS. Therefore, achieving a diversity gain for the weakest user. This technique can be explored to improve the user's achievable rate, sum-rate and also others important variables, such as latency and access failure probability.

REFERENCES

3GPP. *Technical specification group GSM/EDGE radio access network; Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) (3GPP TR 45.820 version 13.1.0 Release 13). [S.l.]*, 2015.

3GPP. *Universal Mobile Telecommunications System (UMTS); Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (3GPP TR 25.996 version 15.0.0 Release 15). [S.l.]*, 2018. Accessed in: 2020/04/18.

ABEBE, Ameha T.; KANG, Chung G. Comprehensive grant-free random access for massive low latency communication. *In*: **2017 IEEE International Conference on Communications (ICC)**. *[S.l.: s.n.]*, 2017. p. 1–6.

ABRAMOWITZ, M.; STEGUN, I.A. **Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables**. New York, NY, USA: Dover Publications, Inc., 1974. ISBN 0486612724.

AL-FUQAHA, Ala; GUIZANI, Mohsen; MOHAMMADI, Mehdi; ALEDHARI, Mohammed; AYYASH, Moussa. Internet of things: A survey on enabling technologies, protocols, and applications. **IEEE Communications Surveys Tutorials**, v. 17, n. 4, p. 2347–2376, 2015.

ALI, Shipon; HOSSAIN, Ekram; KIM, Dong In. Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation. **IEEE Access**, v. 5, p. 565–577, 2017.

BJÖRNSON, E.; CARVALHO, E. de; SORENSEN, J. H.; LARSSON, E. G.; POPOVSKI, P. A random access protocol for pilot allocation in crowded massive MIMO systems. **IEEE Transactions on Wireless Communications**, v. 16, n. 4, p. 2220–2234, April 2017. ISSN 1536-1276.

BJÖRNSON, Emil; HOYDIS, Jakob; SANGUINETTI, Luca. 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, E.; HOYDIS, J.; SANGUINETTI, L. Massive MIMO has unlimited capacity. **IEEE Transactions on Wireless Communications**, v. 17, n. 1, p. 574–590, 2018.

BJÖRNSON, E.; SANGUINETTI, L.; DEBBAH, M. Massive MIMO with imperfect channel covariance information. *In*: **50th Asilomar Conference on Signals, Systems and Computers**. *[S.l.: s.n.]*, 2016. p. 974–978.

BURSALIOGLU, Ozgun Y.; WANG, Chenwei; PAPADOPOULOS, Haralabos; CAIRE, Giuseppe. RRH based massive MIMO with "on the fly" pilot contamination control. *In*: **2016 IEEE International Conference on Communications (ICC)**. *[S.l.: s.n.]*, 2016. p. 1–7.

CAI, Yunlong; QIN, Zhijin; CUI, Fangyu; LI, Geoffrey Ye; MCCANN, Julie A. Modulation and multiple access for 5G networks. **IEEE Communications Surveys Tutorials**, v. 20, n. 1, p. 629–646, 2018.

CARVALHO, E. De; BJORNSON, E.; SORENSEN, J. H.; POPOVSKI, P.; LARSSON, E. G. Random access protocols for massive MIMO. **IEEE Communications Magazine**, v. 55, n. 5, p. 216–222, 2017.

CHANG, Kuor-Hsin. Bluetooth: a viable solution for IoT? [industry perspectives]. **IEEE Wireless Communications**, v. 21, n. 6, p. 6–7, 2014.

CHEN, Chen; CAI, Wenbo; CHENG, Xiang; YANG, Liuqing; JIN, Ye. Low complexity beamforming and user selection schemes for 5G MIMO-NOMA systems. **IEEE Journal on Selected Areas in Communications**, v. 35, n. 12, p. 2708–2722, 2017.

CHEN, He; ABBAS, Rana; CHENG, Peng; SHIRVANIMOGHADDAM, Mahyar; HARDJAWANA, Wibowo; BAO, Wei; LI, Yonghui; VUCETIC, Branka. Ultra-reliable low latency cellular networks: Use cases, challenges and approaches. **IEEE Communications Magazine**, v. 56, n. 12, p. 119–125, 2018.

CHEN, Xiaoming; JIA, Rundong; NG, Derrick Wing Kwan. On the design of massive non-orthogonal multiple access with imperfect successive interference cancellation. **IEEE Transactions on Communications**, v. 67, n. 3, p. 2539–2551, 2019.

CHEN, Xiaoming; NG, Derrick Wing Kwan; YU, Wei; LARSSON, Erik G.; AL-DHAHIR, Naofal; SCHOBER, Robert. Massive access for 5G and beyond. **IEEE Journal on Selected Areas in Communications**, v. 39, n. 3, p. 615–637, 2021.

CHEN, Zhilin; SOHRABI, Foad; LIU, Ya-Feng; YU, Wei. Covariance based joint activity and data detection for massive random access with massive MIMO. *In*: **ICC 2019 - 2019 IEEE International Conference on Communications (ICC)**. *[S.l.: s.n.]*, 2019. p. 1–6.

CHEN, Zhilin; SOHRABI, Foad; YU, Wei. Sparse activity detection for massive connectivity. **IEEE Transactions on Signal Processing**, v. 66, n. 7, p. 1890–1904, 2018.

CHENG, Lei; LIU, Liang; CUI, Shuguang. A covariance-based user activity detection and channel estimation approach with novel pilot design. *In*: **2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)**. *[S.l.: s.n.]*, 2020. p. 1–5.

CISCO. Cisco annual internet report (2018-2023) white paper. Jun 2021. Accessed in 2021/06/19. Available at: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

CUI, Jingjing; DING, Zhiguo; FAN, Pingzhi. Outage probability constrained MIMO-NOMA designs under imperfect CSI. **IEEE Transactions on Wireless Communications**, v. 17, n. 12, p. 8239–8255, 2018.

CUI, Ying; LI, Shuaichao; ZHANG, Wanqing. Jointly sparse signal recovery and support recovery via deep learning with applications in MIMO-based grant-free random access. **IEEE Journal on Selected Areas in Communications**, v. 39, n. 3, p. 788–803, 2021.

DAI, Linglong; WANG, Bichai; DING, Zhiguo; WANG, Zhaocheng; CHEN, Sheng; HANZO, Lajos. A survey of non-orthogonal multiple access for 5G. **IEEE Communications Surveys Tutorials**, v. 20, n. 3, p. 2294–2323, 2018.

DING, Jie; CHOI, Jinho. Comparison of preamble structures for grant-free random access in massive MIMO systems. **IEEE Wireless Communications Letters**, v. 9, n. 2, p. 166–170, 2020.

DING, J.; CHOI, J. Preamble-data superposition random access in massive MIMO systems. **IEEE Wireless Communications Letters**, v. 9, n. 6, p. 906–910, 2020.

DING, Jie; FENG, Mingjie; NEMATI, Mahyar; CHOI, Jinho. Performance analysis of massive mimo assisted semi-grant-free random access. *In*: **2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC)**. *[S.l.: s.n.]*, 2021. p. 1–7.

DING, J.; QU, D.; JIANG, H.; JIANG, T. Success probability of grant-free random access with massive MIMO. **IEEE Internet of Things Journal**, v. 6, n. 1, p. 506–516, 2019.

DING, Zhiguo; PENG, Mugen; POOR, H. Vincent. Cooperative non-orthogonal multiple access in 5G systems. **IEEE Communications Letters**, v. 19, n. 8, p. 1462–1465, 2015.

DING, Z.; SCHOBER, R.; FAN, P.; POOR, H. V. Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access. **IEEE Transactions on Communications**, v. 67, n. 6, p. 4464–4478, 2019.

DING, Zhiguo; SCHOBER, Robert; POOR, H. Vincent. A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment. **IEEE Transactions on Wireless Communications**, v. 15, n. 6, p. 4438–4454, 2016.

DING, Zhiguo; YANG, Zheng; FAN, Pingzhi; POOR, H. Vincent. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. **IEEE Signal Processing Letters**, v. 21, n. 12, p. 1501–1505, 2014.

ELIJAH, Olakunle; LEOW, Chee Yen; RAHMAN, Tharek Abdul; NUNOO, Solomon; ILIYA, Solomon Zakwoi. A comprehensive survey of pilot contamination in massive MIMO-5G system. **IEEE Communications Surveys Tutorials**, v. 18, n. 2, p. 905–923, 2016.

ERICSSON. Ericsson mobility report. Jun 2021. Accessed in 2021/06/17.

FACENDA, G. Kasper; SILVA, D. Efficient scheduling for the massive random access gaussian channel. **IEEE Transactions on Wireless Communications**, v. 19, n. 11, p. 7598–7609, 2020.

FENGLER, Alexander; HAGHIGHATSHOAR, Saeid; JUNG, Peter; CAIRE, Giuseppe. Non-bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver. **IEEE Transactions on Information Theory**, v. 67, n. 5, p. 2925–2951, 2021.

FENGLER, Alexander; JUNG, Peter; CAIRE, Giuseppe. SPARCs and AMP for unsourced random access. *In*: **2019 IEEE International Symposium on Information Theory (ISIT)**. *[S.l.: s.n.]*, 2019. p. 2843–2847.

GERASIN, I.; KRASILOV, A.; KHOROV, E. Flexible multiplexing of grant-free URLLC and eMBB in uplink. *In*: **IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications**. *[S.l.: s.n.]*, 2020. p. 1–6.

HAGHIGHATSHOAR, Saeid; JUNG, Peter; CAIRE, Giuseppe. Improved scaling law for activity detection in massive MIMO systems. *In*: **2018 IEEE International Symposium on Information Theory (ISIT)**. *[S.l.: s.n.]*, 2018. p. 381–385.

HAN, H.; GUO, X.; LI, Y. A high throughput pilot allocation for M2M communication in crowded massive MIMO systems. **IEEE Transactions on Vehicular Technology**, v. 66, n. 10, p. 9572–9576, 2017.

HAN, H.; LI, Y.; GUO, X. A graph-based random access protocol for crowded massive MIMO systems. **IEEE Transactions on Wireless Communications**, v. 16, n. 11, p. 7348–7361, 2017.

HAN, H.; LI, Y.; ZHAI, W.; QIAN, L. A grant-free random access scheme for M2M communication in massive MIMO systems. **IEEE Internet of Things Journal**, v. 7, n. 4, p. 3602–3613, 2020.

HAN, Yonghee; RAO, Bhaskar D.; LEE, Jungwoo. Massive uncoordinated access with massive MIMO: A dictionary learning approach. **IEEE Transactions on Wireless Communications**, v. 19, n. 2, p. 1320–1332, 2020.

HASAN, Monowar; HOSSAIN, Ekram; NIYATO, Dusit. Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches. **IEEE Communications Magazine**, v. 51, n. 6, p. 86–93, 2013.

ISLAM, S. M. Riazul; AVAZOV, Nurilla; DOBRE, Octavia A.; KWAK, Kyung-sup. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. **IEEE Communications Surveys Tutorials**, v. 19, n. 2, p. 721–742, 2017.

ITU. IMT vision - framework and overall objectives of the future development of IMT for 2020 and beyond. **International Telecommunication Union**, Geneva, Switzerland, Recommendation ITU-R M.2083-0, Sep 2015.

JAYANTH, N.; CHAKRABORTY, P.; GUPTA, M.; PRAKRIYA, S. Performance of semi-grant free uplink with non-orthogonal multiple access. *In*: **IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications**. *[S.l.: s.n.]*, 2020. p. 1–6.

JU, Jinjuan; DUAN, Wei; SUN, Qiang; GAO, Shangce; ZHANG, Guoan. Performance analysis for cooperative NOMA with opportunistic relay selection. **IEEE Access**, v. 7, p. 131488–131500, 2019.

KARA, Ferdi; KAYA, Hakan. On the error performance of cooperative-NOMA with statistical CSIT. **IEEE Communications Letters**, v. 23, n. 1, p. 128–131, 2019.

KARA, Ferdi; KAYA, Hakan. Threshold-based selective cooperative-NOMA. **IEEE Communications Letters**, v. 23, n. 7, p. 1263–1266, 2019.

KHANSEFID, Amin; MINN, Hlaing. On channel estimation for massive MIMO with pilot contamination. **IEEE Communications Letters**, v. 19, n. 9, p. 1660–1663, 2015.

KIM, Jung-Bin; LEE, In-Ho. Non-orthogonal multiple access in coordinated direct and relay transmission. **IEEE Communications Letters**, v. 19, n. 11, p. 2037–2040, 2015.

LARSSON, E. G.; EDFORS, O.; TUFVESSON, F.; MARZETTA, T. L. Massive MIMO for next generation wireless systems. **IEEE Communications Magazine**, v. 52, n. 2, p. 186–195, 2014.

LI, Shuaichao; ZHANG, Wanqing; CUI, Ying; CHENG, Hei Victor; YU, Wei. Joint design of measurement matrix and sparse support recovery method via deep auto-encoder. **IEEE Signal Processing Letters**, v. 26, n. 12, p. 1778–1782, 2019.

LIAQAT, M.; NOORDIN, K. A.; LATEF, T. A.; DIMYATI, K. Power-domain non orthogonal multiple access (PD-NOMA) in cooperative networks: An overview. **Wireless Networks**, v. 26, p. 1–23, Jul 2018.

LIU, Gang; CHEN, Xianhao; DING, Zhiguo; MA, Zheng; YU, F. Richard. Hybrid half-duplex/full-duplex cooperative non-orthogonal multiple access with transmit power adaptation. **IEEE Transactions on Wireless Communications**, v. 17, n. 1, p. 506–519, 2018.

LIU, Liang; LARSSON, Erik G.; YU, Wei; POPOVSKI, Petar; STEFANOVIC, Cedomir; CARVALHO, Elisabeth de. Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things. **IEEE Signal Processing Magazine**, v. 35, n. 5, p. 88–99, 2018.

LIU, Liang; YU, Wei. Massive device connectivity with massive MIMO. *In*: **2017 IEEE International Symposium on Information Theory (ISIT)**. *[S.l.: s.n.]*, 2017. p. 1072–1076.

LIU, Liang; YU, Wei. Massive connectivity with massive MIMO-part I: Device activity detection and channel estimation. **IEEE Transactions on Signal Processing**, v. 66, n. 11, p. 2933–2946, 2018.

LIU, Yuan. Exploiting NOMA for cooperative edge computing. **IEEE Wireless Communications**, v. 26, n. 5, p. 99–103, 2019.

LIU, Yuanwei; ELKASHLAN, Maged; DING, Zhiguo; KARAGIANNIDIS, George K. Fairness of user clustering in MIMO non-orthogonal multiple access systems. **IEEE Communications Letters**, v. 20, n. 7, p. 1465–1468, 2016.

LV, Lu; CHEN, Jian; NI, Qiang; DING, Zhiguo; JIANG, Hai. Cognitive non-orthogonal multiple access with cooperative relaying: A new wireless frontier for 5G spectrum sharing. **IEEE Communications Magazine**, v. 56, n. 4, p. 188–195, 2018.

MARINELLO, J. C.; ABRÃO, T. Collision resolution protocol via soft decision retransmission criterion. **IEEE Transactions on Vehicular Technology**, v. 68, n. 4, p. 4094–4097, 2019.

MARINELLO, J. C.; ABRÃO, T.; SOUZA, R. D.; CARVALHO, E. de; POPOVSKI, P. Achieving fair random access performance in massive MIMO crowded machine-type networks. **IEEE Wireless Communications Letters**, v. 9, n. 4, p. 503–507, 2020.

MARZETTA, Thomas L. Noncooperative cellular wireless with unlimited numbers of base station antennas. **IEEE Transactions on Wireless Communications**, v. 9, n. 11, p. 3590–3600, 2010.

MU, Xidong; LIU, Yuanwei; GUO, Li; LIN, Jiaru; AL-DHAHIR, Naofal. Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization. **IEEE Transactions on Wireless Communications**, v. 19, n. 10, p. 6884–6898, 2020.

MU, Xidong; LIU, Yuanwei; GUO, Li; LIN, Jiaru; SCHOBER, Robert. Joint deployment and multiple access design for intelligent reflecting surface assisted networks. **IEEE Transactions on Wireless Communications**, p. 1–1, 2021.

MUKHERJEE, Sudarshan; SINHA, Alok Kumar; MOHAMMED, Saif Khan. Timing advance estimation and beamforming of random access response in crowded TDD massive MIMO systems. **IEEE Transactions on Communications**, v. 67, n. 6, p. 4004–4019, 2019.

NGO, Hien Quoc; LARSSON, Erik G.; MARZETTA, Thomas L. Energy and spectral efficiency of very large multiuser MIMO systems. **IEEE Transactions on Communications**, v. 61, n. 4, p. 1436–1449, 2013.

NGUYEN, Kha-Hung; NGUYEN, Hieu V.; BUI, Van-Phuc; SHIN, Oh-Soon. Dynamic user pairing for non-orthogonal multiple access in downlink networks. *In*: **2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)**. *[S.l.: s.n.]*, 2021. p. 111–114.

NISHIMURA, O. S.; MARINELLO, J. C.; ABRÃO, T. A grant-based random access protocol in extra-large massive MIMO system. **IEEE Communications Letters**, v. 24, n. 11, p. 2478–2482, 2020.

OTAO, N.; KISHIYAMA, Y.; HIGUCHI, K. Performance of non-orthogonal access with SIC in cellular downlink using proportional fair-based resource allocation. *In*: **International Symposium on Wireless Communication Systems (ISWCS)**. *[S.l.: s.n.]*, 2012. p. 476–480. ISSN 2154-0225.

PEREIRA, Hebert Douglas; BRANTE, Glauber; FARHAT, Jamil; SOUZA, Richard Demo; MARINELLO, José Carlos; Abrão, T. On the sum-rate of contention resolution in massive MIMO with NOMA. **IEEE Access**, v. 9, p. 24965–24974, 2021.

POKHREL, Shiva Raj; WILLIAMSON, Carey. Modeling compound TCP over WiFi for IoT. **IEEE/ACM Transactions on Networking**, v. 26, n. 2, p. 864–878, 2018.

POLYANSKIY, Yury. A perspective on massive random-access. *In*: **2017 IEEE International Symposium on Information Theory (ISIT)**. *[S.l.: s.n.]*, 2017. p. 2523–2527.

RAZA, Usman; KULKARNI, Parag; SOORIYABANDARA, Mahesh. Low power wide area networks: An overview. **IEEE Communications Surveys Tutorials**, v. 19, n. 2, p. 855–873, 2017.

SAFEATLAST. 80 insightful internet of things statistics (infographic). Jun 2021. Accessed in 2021/06/17. Available at: https://safeatlast.co/blog/iot-statistics/#gref.

SAITO, Y.; KISHIYAMA, Y.; BENJEBBOUR, A.; NAKAMURA, T.; LI, A.; HIGUCHI, K. Non-orthogonal multiple access (NOMA) for cellular future radio access. *In*: **IEEE 77th Vehicular Technology Conference (VTC Spring)**. *[S.l.: s.n.]*, 2013. p. 1–5. ISSN 1550-2252.

SANGUINETTI, Luca; D'AMICO, Antonio A.; MORELLI, Michele; DEBBAH, Mérouane. Random access in massive MIMO by exploiting timing offsets and excess antennas. **IEEE Transactions on Communications**, v. 66, n. 12, p. 6081–6095, 2018.

SENEL, Kamil; LARSSON, Erik G. Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach. **IEEE Transactions on Communications**, v. 66, n. 12, p. 6164–6175, 2018.

SHAHAB, M. B.; ABBAS, R.; SHIRVANIMOGHADDAM, M.; JOHNSON, S. J. Grant-free non-orthogonal multiple access for IoT: A survey. **IEEE Communications Surveys Tutorials**, v. 22, n. 3, p. 1805–1838, 2020.

SHAO, Xiaodan; CHEN, Xiaoming; NG, Derrick Wing Kwan; ZHONG, Caijun; ZHANG, Zhaoyang. Cooperative activity detection: Sourced and unsourced massive random access paradigms. **IEEE Transactions on Signal Processing**, v. 68, p. 6578–6593, 2020.

SHIRVANIMOGHADDAM, Mahyar; LI, Yonghui; DOHLER, Mischa; VUCETIC, Branka; FENG, Shulan. Probabilistic rateless multiple access for machine-to-machine communication. **IEEE Transactions on Wireless Communications**, v. 14, n. 12, p. 6815–6826, 2015.

SHYIANOV, Volodymyr; BELLILI, Faouzi; MEZGHANI, Amine; HOSSAIN, Ekram. Massive unsourced random access based on uncoupled compressive sensing: Another blessing of massive MIMO. **IEEE Journal on Selected Areas in Communications**, v. 39, n. 3, p. 820–834, 2021.

SORENSEN, Jesper H.; CARVALHO, Elisabeth de; STEFANOVIC, Cedomir; POPOVSKI, Petar. Coded pilot random access for massive MIMO systems. **IEEE Transactions on Wireless Communications**, v. 17, n. 12, p. 8035–8046, 2018.

SULYMAN, Ahmed Iyanda; OTEAFY, Sharief M. A.; HASSANEIN, Hossam S. Expanding the cellular-IoT umbrella: An architectural approach. **IEEE Wireless Communications**, v. 24, n. 3, p. 66–71, 2017.

SUN, Xiaofang; YANG, Nan; YAN, Shihao; DING, Zhiguo; NG, Derrick Wing Kwan; SHEN, Chao; ZHONG, Zhangdui. Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks. **IEEE Transactions on Wireless Communications**, v. 17, n. 8, p. 5367–5381, 2018.

SUNDARAM, Jothi Prasanna Shanmuga; DU, Wan; ZHAO, Zhiwei. A survey on LoRa networking: Research problems, current solutions, and open issues. **IEEE Communications Surveys Tutorials**, v. 22, n. 1, p. 371–388, 2020.

TALEB, T.; KUNZ, A. Machine type communications in 3GPP networks: potential, challenges, and solutions. **IEEE Communications Magazine**, v. 50, n. 3, p. 178–184, 2012.

TRUHACHEV, Dmitri; BASHIR, Murwan; KARAMI, Alireza; NASSAJI, Ehsan. Low-complexity coding and spreading for unsourced random access. **IEEE Communications Letters**, v. 25, n. 3, p. 774–778, 2021.

TULLBERG, H.; POPOVSKI, P.; LI, Z.; UUSITALO, M. A.; HOGLUND, A.; BULAKCI, O.; FALLGREN, M.; MONSERRAT, J. F. The METIS 5G system concept: Meeting the 5G requirements. **IEEE Communications Magazine**, v. 54, n. 12, p. 132–139, 2016.

VARGHESE, Susan G.; KURIAN, Ciji Pearl; JOHN, V. I. Anupriya; NAYAK, Varsha; UPADHYAY, Anil. Comparative study of ZigBee topologies for IoT-based lighting automation. **IET Wireless Sensor Syst.**, v. 9, n. 4, p. 201–207, 2019.

WAN, Dehuan; WEN, Miaowen; JI, Fei; LIU, Yun; HUANG, Yu. Cooperative NOMA systems with partial channel state information over Nakagami-m fading channels. **IEEE Transactions on Communications**, v. 66, n. 3, p. 947–958, 2018.

WANG, Bichai; DAI, Linglong; MIR, Talha; WANG, Zhaocheng. Joint user activity and data detection based on structured compressive sensing for NOMA. **IEEE Communications Letters**, v. 20, n. 7, p. 1473–1476, 2016.

WANG, Jue; ZHANG, Zhaoyang; HANZO, Lajos. Joint active user detection and channel estimation in massive access systems exploiting Reed-Muller sequences. **IEEE Journal of Selected Topics in Signal Processing**, v. 13, n. 3, p. 739–752, 2019.

WEI, Zhiqiang; DAI, Linglong; NG, Derrick Wing Kwan; YUAN, Jinhong. Performance analysis of a hybrid downlink-uplink cooperative NOMA scheme. *In*: **2017 IEEE 85th Vehicular Technology Conference (VTC Spring)**. *[S.l.: s.n.]*, 2017. p. 1–7.

XIE, Xianbin; LIU, Jinhua; HUANG, Jiawen; ZHAO, Shiwei. Ergodic capacity and outage performance analysis of uplink full-duplex cooperative NOMA system. **IEEE Access**, v. 8, p. 164786–164794, 2020.

XU, C.; HU, Y.; LIANG, C.; MA, J.; PING, L. Massive MIMO, non-orthogonal multiple access and interleave division multiple access. **IEEE Access**, v. 5, p. 14728–14748, 2017. ISSN 2169-3536.

YANG, Z.; XU, P.; HUSSEIN, J. Ahmed; WU, Y.; DING, Z.; FAN, P. Adaptive power allocation for uplink non-orthogonal multiple access with semi-grant-free transmission. **IEEE Wireless Communications Letters**, v. 9, n. 10, p. 1725–1729, 2020.

YIN, H.; GESBERT, D.; FILIPPOU, M.; LIU, Y. A coordinated approach to channel estimation in large-scale multiple-antenna systems. **IEEE Journal on Selected Areas in Communications**, v. 31, n. 2, p. 264–273, 2013.

YU, H.; FEI, Z.; ZHENG, Z.; YE, N. Finite-alphabet signature design for grant-free NOMA: A quantized deep learning approach. **IEEE Transactions on Vehicular Technology**, v. 69, n. 10, p. 10975–10987, 2020.

YU, Nam Yul. Binary golay spreading sequences and Reed-Muller codes for uplink grant-free NOMA. **IEEE Transactions on Communications**, v. 69, n. 1, p. 276–290, 2021.

ZHANG, C.; LIU, Y.; YI, W.; QIN, Z.; DING, Z. Semi-grant-free NOMA: Ergodic rates analysis with random deployed users. **IEEE Wireless Communications Letters**, p. 1–1, 2020.

ZHANG, C.; QIN, Z.; LIU, Y.; CHAI, K. K. Semi-grant-free uplink NOMA with contention control: A stochastic geometry model. *In*: **IEEE International Conference on Communications Workshops (ICC Workshops)**. *[S.l.: s.n.]*, 2020. p. 1–6.

ZHANG, Ningbo; WANG, Jing; KANG, Guixia; LIU, Yang. Uplink nonorthogonal multiple access in 5G systems. **IEEE Communications Letters**, v. 20, n. 3, p. 458–461, 2016.

ZHANG, Q.; JIN, S.; ZHU, H. A hybrid-grant random access scheme in massive MIMO systems for IoT. **IEEE Access**, v. 8, p. 88487–88497, 2020.

ZHANG, Wanqing; LI, Shuaichao; CUI, Ying. Jointly sparse support recovery via deep auto-encoder with applications in MIMO-based grant-free random access for mMTC. *In*: **2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)**. *[S.l.: s.n.]*, 2020. p. 1–5.

ZHENG, Beixiong; WU, Qingqing; ZHANG, Rui. Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA? **IEEE Communications Letters**, v. 24, n. 4, p. 753–757, 2020.

ZUO, Jiakuo; LIU, Yuanwei; QIN, Zhijin; AL-DHAHIR, Naofal. Resource allocation in intelligent reflecting surface assisted NOMA systems. **IEEE Transactions on Communications**, v. 68, n. 11, p. 7170–7183, 2020.