DOCTORAL THESIS

Scheduling in Radio Resource Management for Massive Machine-Type Communications

Collins Burton Mwakwata

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2022

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
32/2022

Scheduling in Radio Resource Management for Massive Machine-Type Communications

COLLINS BURTON MWAKWATA

PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies
Thomas Johann Seebeck Department of Electronics

The dissertation was accepted for the defence of the degree of Doctor of Philosophy in Information and Communication Technology on 10 May 2022

Supervisor: Professor Muhammad Mahtab Alam,
Thomas Johann Seebeck Department of Electronics,
School of Information Technologies,
Tallinn University of Technology
Tallinn, Estonia

Co-supervisor: Professor Yannick Le Moullec,
Thomas Johann Seebeck Department of Electronics,
School of Information Technologies,
Tallinn University of Technology
Tallinn, Estonia

Opponents: Professor Luca Reggiani,
Politecnico di Milano,
Milano, Italy

Professor Qammer Abbasi,
University of Glasgow,
Glasgow, Scotland

Defence of the thesis: 21 June 2022, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been submitted for any academic degree elsewhere.

Collins Burton Mwakwata

signature

ISSN 2585-6898 (publication)
ISBN 978-9949-83-845-5 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9949-83-846-2 (PDF)
Printed by Koopia Niini & Rauam

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
32/2022

Raadioressursside planeerimine massiivse masin-masin tüüpi kommunikatsiooni korral

COLLINS BURTON MWAKWATA

KIRJASTUS

Contents

List of Publications and Patents

The present Ph.D. thesis is based on the following publications that are referred to in the text by Roman numbers.

Publications

I **C.B. Mwakwata**, H. Malik, M.M. Alam, Y. Le Moullec, S. Parand, S. Mumtaz. Narrowband Internet of Things (NB-IoT): From physical (PHY) and Media Access Control (MAC) Layers Perspectives. MDPI Sensors, 19(11), 2613 2019.

II **C.B. Mwakwata**, M.M. Alam, Y. Le Moullec, H. Malik, S. Pärand. Cooperative Interference Avoidance Scheduler for Radio Resource Management in NB-IoT Systems. European Conference on Networks and Communications (EUCNC), IEEE, 2020.

III **C.B. Mwakwata**, O. Elgarhy, Y. Le Moullec, M.M. Alam, S. Pärand, I. Annus. Intercell Interference Reduction Scheme for Uplink Transmission in NB-IoT Systems. 2021 International Wireless Communications and Mobile Computing (IWCMC), 400-405. IEEE, 2021.

IV **C.B. Mwakwata**, O. Elgarhy, M.M. Alam, Y. Le Moullec, S. Pärand, K. Trichias, K. Ramantas. Cooperative Scheduler to Enhance Massive Connectivity in 5G and Beyond by Minimizing Interference in OMA and NOMA. IEEE Systems Journal. (Early Access). IEEE, 2021.

Paper under review

V **C.B. Mwakwata**, M.M. Alam, Y. Le Moullec. Performance enhancement of RAN Slice scheduling by using Machine Learning Techniques adapted for beyond 5G wireless Networks. IEEE Systems, 2022.

Patent Application

I Interference minimizing cooperative scheduler for orthogonal multiple access (OMA) and non-orthogonal multiple access (NOMA) wireless communications; Owners: Tallinn University of Technology ; Authors: M.M. Alam, **C.B. Mwakwata**; Priority number: FR2103489; Priority date: 6.04.2021.

Author's Contributions to the Publications

I **Publication** I: I was the **first author**; I provided a comprehensive overview of the design changes brought in the cellular IoT standardization along with the detailed research developments from the perspectives of physical and MAC layers. Furthermore, I elaborated an overview of Evolved Packet Core (EPC) changes to support the Service Capability Exposure Function (SCEF); additionally, I proposed and discussed the possible deployment scenarios of mMTC (i.e., NB-IoT) in future Heterogeneous Wireless Networks (HetNet). Finally, I discussed the existing and emerging research challenges to motivate the planned research activities. I drafted the paper, prepared all figures and tables, and revised the manuscript based on reviewers' feedback.

II **Publication** II: I was the **first author**; I proposed an interference avoidance scheduling algorithm for NB-IoT systems. I simulated the proposed cooperative strategy to mitigate interference. I utilized the computed interference values as input to individual base station schedulers to perform scheduling. I performed extensive simulations to analyze the performance of our proposed algorithm and compared it to the conventional Round-Robin scheduling scheme. I drafted the paper, prepared all figures and tables, and revised the manuscript based on reviewers' feedback.

III **Publication** III: I was the **first author**; I proposed an inter-cell interference (ICI) minimization scheme for uplink transmission focusing on massive machine-type communications (mMTC) and specifically the narrow-band internet of things (NB-IoT) systems. I firstly established the theoretical ICI problem formulation and proposed its corresponding solution for the orthogonal multiple access (OMA) NB-IoT system. Based on the theoretical formulation, I designed a cooperative radio resource scheduler to reduce the impact of ICI and allocate the transmit powers to reduce the energy consumption to the scheduled users in a multi-cell scenario. Lastly, I ran extensive simulations to compare the performance of the proposed scheme with that of some benchmark OMA schedulers. I drafted the paper, prepared all figures and tables, and revised the manuscript based on reviewers' feedback.

IV **Publication** IV: I was the **first author**; I studied the impact of interference caused by mMTC connections. Additionally, I theoretically modeled the inter-cell interference (ICI) minimization problem for the existing orthogonal multiple access (OMA) technique and proposed its corresponding solution. Furthermore, I implemented the simulations to map the proposed solutions for both OMA and NOMA schemes. Finally, I evaluated the performance enhancements of the designed scheduler and compared them with the state-of-the-art frameworks, I drafted the paper, prepared all figures and tables, and revised the manuscript based on reviewers' feedback.

V **Publication** V: I was the **first author**; I implemented the data-driven model by using machine learning to enhance the scheduling of RAN slicing for beyond 5G networks. The author utilized the machine learning output parameters as input to the RAN scheduler to improve the slice scheduling, hence I performed the simulation based on real-time data collected from the 5G network, to evaluate the performance enhancements in comparison to the proposed schedulers from the literature. I drafted the paper, prepared all figures and tables, and revised the manuscript based on reviewers' feedback.

Abbreviations

3D	Three Dimension
3GPP	the 3rd Generation Partnership Project
4G	Fourth generation
5G	Fifth Generation
6G	Sixth Generation
AI	Artificial Intelligence
AP	Access Point
APSO	Adaptive particle swarm optimization
AVs	Autonomous Vehicles
B5G	Beyond 5G
BER	bit error rate
BS	Base Station
CCI	Co-Channel interference
CoMP	Coordinated Multi Point
CQI	Channel Quality Indicator
DCI	Downlink Control Information
DMRS	Demodulation Reference Signal
eDRX	extended Discontinuous Reception
eICIC	enhanced inter-cell interference coordination
eMBB	enhanced Mobile Broad Band
EPA	Equal power allocation
EPC	Evolved Packet Core
ESA	Equal sub-carrier allocation
FDR	Full Duplex Relaying
FFR	Fractional Frequency Reuse
FG-ML5G	Focus Group on Machine Learning for Future Networks including 5G
GF-NOMA	Grant Free-Non-Orthogonal Multiple Access
GSM	Global System for Mobile Communication
HARQ	Hybrid Automatic Repeat Request
HetNet	Heterogeneous Networks
HD	high definition
HDR	Half Duplex Relaying
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
HSUPA	High speed Uplink Packet Access
IAB	Integrated Access Backhaul
ICI	Inter-Cell Interference
IRS	Intelligent reflecting surfaces
IoT	Internet of Things
KKT	Karush Kuhn Tucker
KPI	Key Performance Indicators
LoRa	Long Range
LOS	line of sight
LPWAN	Low Power Wide Area Networks
LTE-M	Long Term Evolution for MTC
M2M	Machine to Machine

MBMS	Multimedia Broadcast Multicast Services
MCL	Maximum Coupling Loss
MIB	Master Information Block
ML	Machine Learning
mMTC	massive Machine Type Communications
MU MIMO	Multi User Multiple Input Multiple Output
NB-CIoT	Narrowband Cellular IoT
NB-M2M	Narrowband Machine to Machine
NB-IoT	Narrow-band Internet of Things
NFV	network functions virtualization
NOMA	Non Orthogonal Multiple Access
NPBCH	Narrowband Physical Broadcast Channel
NPDCCH	Narrowband Physical Downlink Control Channel
NPDSCH	Narrowband Physical Downlink Shared Channel
NPRACH	Narrowband Physical Random Access Channel
NPUSCH	Narrowband Physical Uplink Shared Channel
NPSS	Narrowband Primary Synchronization Signal
NR	New Radio
NRS	Narrowband Reference Signal
NSSS	Narrowband Secondary Synchronization Signal
NWDAF	Network Data Analytics Function
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
OTDOA	Observed time difference of arrival
PD	Power Domain
PSM	Power Saving Mode
QML	Quantum machine learning
QoE	Quality of experience
QoS	Quality of Service
RACH	Random Access Channel
RAN	Radio Access Network
RF	Radio Frequency
RIS	Reconfigurable Intelligent Surfaces
RRC	Radio Resource Control
RRM	radio resources management
RU	Resource Units
SC	Superposition Coding
SCEF	Service Capability Exposure Function
SCFDMA	Single Carrier Frequency Division Multiple Access
SC MCCH	Single-cell MBMS Control Channel
SC MTCH	Single-Cell MBMS Traffic Channe
SC-PTM	single-cell point-to-multipoint
SDR	software-defined networking
SIB	System Information Block
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference plus Noise Ratio
SLA	service level agreement
TBS	Transport Block Size
TSC	Time Sensitive Communication

TSN	Time-Sensitive Networking
UE	User Equipment
uRLLC	Ultra-Reliable Low Latency Communications
V2X	Vehicle to everything
VR	virtual reality

Symbols

$SINR^{z}_{x_c}$	the SINR of user x_c attached to cell c at sub-carrier z
$a^{z}_{x_c}$	allocation matrix of user x_c at sub-carrier z
I^{z}_{c}	interference on resource unit z in cell c
$P^{z}_{q_l}$	is the transmission power of the interfering user q_l at resource unit z
$P^{z}_{x_c}$	is the transmission power of user k_c's at resource unit z
$h^{z}_{x_c,c}$	is the channel gain of user q_l to the base station c at sub-carrier z
σ_N	is the receiver's noise power
$\vartheta_{x_c,lim}$	is the $SINR$ threshold for user k_c to satisfy its QoS
$h_{q_l,c}$	is the channel gain user q_l at cell c
P_{maxm}	is the maximum power that can be used by the user for its transmissions
$h_{c,c}$	is the channel gain of the transmitting user c at base station c
$h^{j}_{l,c}$	is the channel gain from user j, belonging to cell l and the NOMA group M_l within the cell, on cell c
$SINR^{z}_{x_c,NOMA}$	is the user x_c's SINR at sub-carrier z under NOMA approach

List of Figures

List of Tables

1 Introduction

1.1 Background and Motivation

By 2030, the 6th Generation (6G) of mobile networks is expected to integrate the terrestrial, aerial, and maritime communications into a heterogeneous network that is robust, reliable, with ultra-low latency for massive connectivity. In this regard, cutting edge technologies such as tera-Hertz and millimeter waves communication, Intelligent reflecting surfaces (IRS), artificial intelligence (AI)/machine learning (ML), non-orthogonal multiple access (NOMA), block-chain, tactile Internet, small cells communication, fog/edge computing, quantum communication/quantum machine learning (QML), etc., are the prime key technologies in the realization of beyond 5G (B5G) and 6G communications.

The massive deployment of devices to support the future networks' use-cases will realize advanced services such as smart traffic, smart homes, environment monitoring and control, virtual reality (VR)/virtual navigation, telemedicine, digital sensing, high definition (HD), condition-based maintenance, and full HD video transmission in connected drones and robots. IoT devices are predicted to reach 25 billion by the year 2025, therefore, to create such IoT environments, wireless connectivity is not an option, it is imperative. Therefore, the number of wireless IoT devices sustaining the continued growth of the IoT ecosystem requires more spectrum or efficient ways of using the available spectrum. However, it is very challenging for the existing multiple access techniques to accommodate such dense deployment of complex systems [2].

The fifth generation of mobile networks brings into play several service verticals such as Massive Machine Type Communications (mMTC), Ultra-Reliable Low Latency Communications (uRLLC), and enhanced mobile broadband (eMBB) communications. However, contrary to previous generations of mobile technology where the primary focus was to enable human-to-human communications, these verticals focus on enabling industrial communications [3]. In this regard, the demand for dense deployment, reliable, secure, low latency connectivity becomes obvious.

For example, uRLLC is one of the biggest game-changers which serve the new applications that require a response in fractions of a second. Examples of such applications include autonomous vehicles where the vehicles will be able to respond 10 to 100 times faster than over the existing cellular networks. Vehicle to everything (V2X) is the term used to describe the communication network where the vehicles will be able to meet such demands. V2X will be enabled along with intelligent transportation systems that will enhance smart traffic management. Other applications under uRLLC include industrial automation etc [4]. eMBB represents the evolution of the current 4G networks but with faster data rates and hence better user experience. eMBB is considered the enabler of immersive VR and AR applications. eMBB is intertwined with uRLLC as it enhances the real-time traffic alerts, high-speed internet access, streaming real-time video as well as playing games involving 3D 4K video [5]. Similarly, the mMTC aims at increasing the number of connected devices per human. Contrary to 4G LPWAN that could support up to $60,680$ devices per square kilometer, mMTC is estimated to support a minimum connection density of 1 million devices per square kilometer. In this thesis, the focus is given to mMTC due to its capacity to support massive IoT applications [6, 7].

mMTC is enabled by both licensed IoT technologies (e.g., Narrow-Band IoT (NB-IoT) [6] and unlicensed technologies (e.g., LoRa). Both types are categorized as low power wide area networks (LPWAN), aiming at servicing devices located in hard-to-reach areas, with minimum human intervention. However, in contrast to unlicensed technologies, licensed technologies reuse the existing cellular infrastructure and are, therefore, more economical and advantageous for cellular telecommunication operators and thus also for the end customers.

The current 5G deployments implement orthogonal multiple access (OMA) schemes which provide orthogonality in terms of frequency resources. However, for massive IoT technologies (i.e., NB-IoT and LTE-M), these OMA schemes are not able to reach the capacity demand for supporting $52,000$ devices per cell. This is because the meager spectrum is expected to accommodate a such number of transmitting devices per base station [8–10]. Consequently, uplink inter-cell interference is a major challenge for such a massive number of devices competing in a dense network, while ensuring that at the same time the required quality of service (QoS) to NB-IoT UEs can be delivered.

Additionally, the 5G broadband and 5G new radio (NR) capabilities bring the possibility of massive connectivity support of up to $1,000,000$ devices per square kilometer [11, 12]. In this regard, proactive scheduling and advanced multiple access techniques to support such dense deployment become of great significance. Like in LTE numerology, NB-IoT supports Orthogonal Frequency Division Multiple Access (OFDMA) and Single Carrier Frequency Division Multiple Access (SC-FDMA) in downlink and uplink, respectively. NB-IoT can be deployed in-band or in the guardband of the LTE spectrum; it can also occupy the spectrum of legacy Global System for Mobile Communication (GSM) technology (i.e. standalone). In these three modes, the NB-IoT system bandwidth is of a maximum of $200\ kHz$. In this limited bandwidth, NB-IoT supports $15\ kHz$ or $3.75\ kHz$ spacing on the uplink, in which different user equipment (UEs) can occupy single or several frequency slots (tones) [13]. This narrow bandwidth motivates the need for implementing novel radio resources management (RRM) techniques to provide maximum-possible data rates with minimum possible bit error rates (BER) performance. In particular, network slicing is proposed to support the futuristic diverse and heterogeneous architectures needed for robust dense networks where the network operators will be able to deploy only the functions necessary to support particular customers and particular market segments. In general, network slicing is similar to software-defined networking (SDN) and network functions virtualization (NFV) which facilitate networks toward software-based automation. In this regard, the one-size-fits-all design philosophy applied in existing networks is not viable anymore; therefore, the SDN and NFV enhance the network partitioning into virtual elements that enable the flexible orchestration of physical networks for a particular service segment [14].

1.2 Challenges in Wireless Communication

It can be noted that key players in the telecommunications ecosystem especially those representing IoT network providers, device manufacturers, users, and federal regulators identified spectrum-related challenges such as ensuring the availability spectrum and counteracting the impact of interference that is caused by the massive IoT deployments. Additionally, the modern communication network is shifting to new design paradigms driven by the strong need to improve Key Performance Indicators (KPIs) required by emerging wireless applications, specifically, the quality of user experiences (QoEs), and proactively avoid any disturbances that affect the service availability.

For example, the NLOS due to multipath, users' mobility and inter-cell interference from the users from adjacent cells affect the quality of the desired received signal, similarly, the orthogonal multiple access techniques fail to reach the expected cell capacity performance, therefore, it is necessary to study the novel radio resources management and the non orthogonal multiple access (NOMA) techniques to increase the users' performance and spectrum efficiency. An overview of the factors affecting the quality of the received signal hence requiring the novel radio resources management approaches for efficient spectrum management is depicted in Fig. 1 and are discussed in what follows.

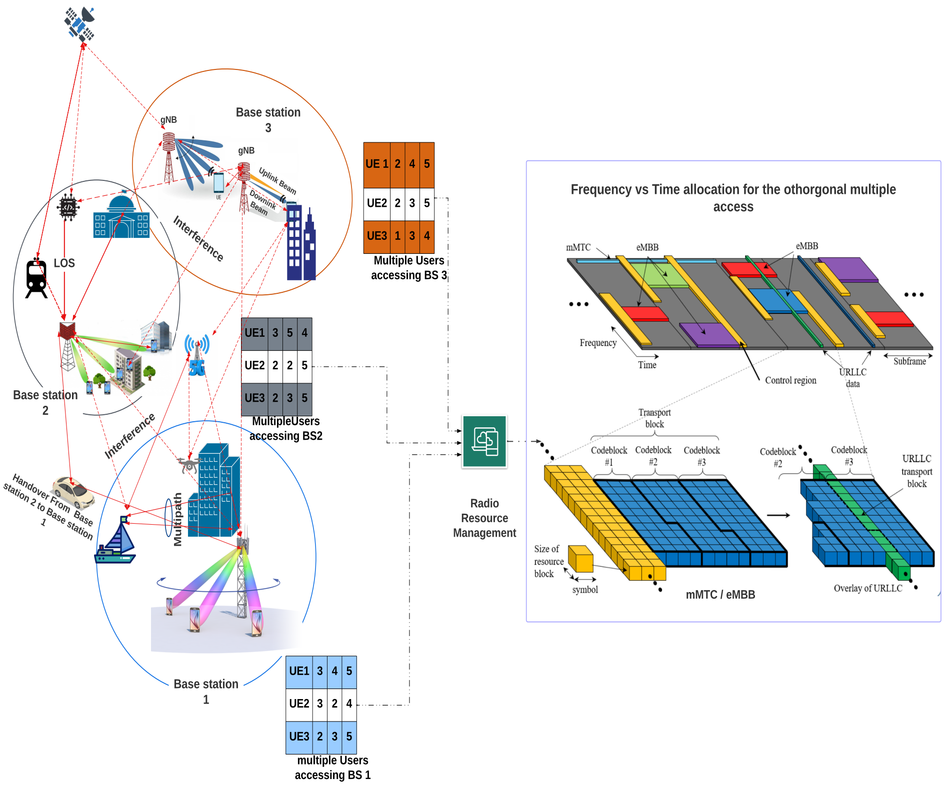

Figure 1: Radio Resource Management especially interference management as enabler for efficient spectrum utilization

1.2.1 Limited Energy

One of the major challenges of wireless communication is energy efficiency. This is because the wireless connected devices are mainly deployed for use cases that involve mobility, therefore, the devices need to be powered by a battery when transmitting through the access points such as base stations. Similarly, for mMTC applications, the rapidly changing of channel conditions, and the massive connectivity affect the overall energy consumption because the devices are forced to utilize the maximum transmit power, or re-transmissions to counteract the bad channel conditions, and interference. Furthermore, the power amplifiers, and signal processing cause energy consumption costs; in this regard, it is essential to propose energy-efficient mechanisms to enhance battery life.

1.2.2 Multipath Propagation

Multipath propagation happens when the transmitted signal from the source is spread in several paths due to objection between the transmitter and the receiver, i.e., as indicated at Base Station 1, in Fig. 1. In this regard, it is imperative to ensure proper mechanisms at the receiver that mitigate the impact of multi-path transmissions by compensating loss on the transmitted signal and equalizing the multiple-received signals to recuperate the desired information. In this regard, novel techniques such as intelligent reflecting surfaces (IRS) also termed Reconfigurable intelligent surfaces (RISs) are being used to enhance the link performance [15].

1.2.3 Radio Spectrum Limitations

One of the common resources that enable wireless communication is the radio spectrum. The licensed spectrum is expensive, hence the network operators are obligated to perform efficient spectrum management approaches to guarantee that regardless of the limited frequency resources, the available frequencies are utilized with minimum possible interference between the same operator's adjacent base stations, as well as between different operators. Therefore, it is crucial to implement spectrum sharing schemes that are resource-efficient and also guarantee the expected quality of service requirements. For example. mMTC, eMBB and uRLLC need efficient spectrum management to allocate the radio resources accordingly as indicated on the right side in Fig. 1.

1.2.4 Interference-Limited Systems

In scenarios where the interference is so high and dominant that the noise is almost negligible, the overall Signal-to-Interference plus Noise Ratio (SINR) at the receiver is so small that it is impossible to decode the signal and reconstruct the desired information. Therefore, it is necessary to characterize all forms of interference that can degrade the quality of communications by guaranteeing the SINR maximization through the design of the interference mitigation techniques to satisfy the expected transmission KPI. For example, the IoT user in Base Station 2 experiences interference (dashed lines), from adjacent base station 3 in Fig. 1, similarly, a given user can experience multiple interfering signals from several users from adjacent cells.

1.2.5 User Mobility

On top of multipath fading, the support for user mobility is very important to enhance the quality of experience; however, when a mobile user moves from one base station to another, there should be a seamless handover process between the consecutive serving base stations. Therefore, it is necessary to propose efficient schemes that take into account the channel variations created by user mobility to guarantee the user's expected quality of service and quality experience for the corresponding applications and use cases. For example, the user from Base station 2 to Base station 1 in Fig 1 should experience seamless handover process with expected quality of service.

1.2.6 Noise-Limited Systems

To guarantee the expected quality of service, the transmitting user must be in an acceptable SINR condition. In this regime, the system is required to have a minimum noise level which can then guarantee higher SINR, hence better transmission possibilities. The source of noise can either be internal or external, therefore, it is critical to design systems that are robust to all forms of noise to satisfy the required KPI.

1.2.7 Multiple Access Techniques

In wireless communications, it is vital to propose efficient multiple access techniques that can meet stringent requirements such as low latency, massive connectivity, spectral efficiency, fairness, required high throughput, reliability, etc. Therefore, it is necessary to map the physical resource block or the resource units (i.e., for NB-IoT), as shown with the Radio Resource Management Block in Fig.1, to guarantee the maximum possible number of connected users. Both the orthogonal and non-orthogonal multiple access techniques need to be well characterized to respond to the growing demand for massive connectivity for the current 5G and future 6G networks.

It should be noted that there are even more challenges (i.e., related to security and privacy, flexible duplexing, directional beamforming, passive inter-modulation, etc.) however, these are outside of the scope of this thesis. In this regard, it is important to note that in complex wireless systems especially in heterogeneous networks architecture, several factors can simultaneously affect the overall cell performance, i.e., the users can simultaneously experience multipath and inter-cell interference, in this regard, it is necessary to design intelligent systems that take into account several factors while decoding the desired transmitted signal.

1.3 Problem Statement and Research Questions

The growing demand for dense connectivity and higher data rates to satisfy the expected required connectivity for all the service verticals that are brought by the 5G and beyond 5G networks raises the issue of how to efficiently use the meager spectrum resource to satisfy the above.

In essence, the direction towards which the future of wireless communication is moving leads to an unprecedented level of complexity; our vision to overcome this situation is to enable the data-driven paradigm for the design of such networks. In this regard, the main question is not whether the machine learning approaches will be integrated into future wireless networks, but rather, how and when this integration will take place. Several leading academic, research, and industry communities support this statement and activated the ITU focus group named Focus Group on Machine Learning for Future Networks including 5G (FG-ML5G) [16].

Therefore, the main hypothesis formulated for this Ph.D. work is **"How the novel data-driven radio resource management approaches, especially, for the interference management techniques, would facilitate the autonomous network management with efficient utilization of the available spectrum in 5G, and beyond 5G heterogeneous network architectures?"**.

Even though the conventional approaches to resource management, especially for interference mitigation, are based on optimization theory techniques, it is necessary to implement suitable mathematical models for a particular problem, i.e. tractable to best characterize the proposed solutions based on information theory. However, even when this is done, the optimal solution will inevitably depend on the system parameters, i.e., the receiver's sensitivity, the users' location, the connection density, slow/fast fading channel variations, etc. When any of these parameters change, the optimization problem needs to be reformulated again, in this regard, it becomes very challenging to implement in real-time applications, especially in heterogeneous wireless network architectures. Furthermore, with dense deployments of wireless networks to support mMTC applications, there is an opportunity to utilize the amount of available network data to implement the novel approaches that utilize artificial intelligence (AI) to enable the data-driven real-time radio resource management techniques especially to counteract the impact of interference for the dense deployment of mMTC in wireless networks.

Considering the heterogeneous network architectures, the limited radio resources i.e., power, spectrum for several service verticals with diverse performance requirements in a given 5G physical network, in co-existence with legacy networks, this Ph.D. thesis intends to answer the following research questions:

- **RQ1: How are the orthogonal multiple access techniques intended to accommodate the massive connectivity of IoT use cases impacted by inadequate interference management techniques in heterogeneous network architectures?**

- **RQ2: Is it possible to proactively perform the data-driven radio resource management techniques to enhance the cell performance, with user's minimum-possible energy consumption while guaranteeing the expected quality of service requirements?**

- **RQ3: Are the novel approaches such as non-orthogonal multiple access (NOMA) sufficient to reach the expected cell performance for mMTC applications, considering the impact of inter-cell interference from the neighboring cells, and the complexity of successive interference cancellation (SIC) at the receiver?**

- **RQ4: Since radio access network (RAN) slicing proves to be efficient for sharing and orchestration of networks, how is it possible to accommodate the massive connectivity of mMTC connections with diverse service requirements in a given slice while guaranteeing the expected performance as per service level agreement (SLA) templates?**

1.4 Contributions of the thesis

To respond to the above questions, the following contributions are presented in this thesis:

- An in-depth literature review: It is performed to understand the state of the art in both telecommunication standards, and the research in the direction of radio resource management, especially the interference management techniques for massive connectivity in legacy, 5G, and beyond 5G wireless networks [6] (**Publication I**).

- The design of a novel interference management scheduler: A novel interference management scheme that utilizes a cooperative strategy in a multi-cell network has been studied, different network topology (i.e., micro-cell and Macro-cell architectures) have been explored and the corresponding interference impact is mitigated accordingly. Contrary to existing works, the proposed scheme computes the interference weights between users from adjacent cells and selects the best group of users with minimum impact of inter-cell interference, hence allocating the available radio resources for the corresponding transmissions [17] (**Publication II**).

- The novel inter-cell and intra-call interference management scheme for the orthogonal multiple access (OMA) techniques is proposed: With the proposed cooperative scheduler, the theoretical framework along with resource management scheme is developed to mitigate the impact of inter-cell interference for dense connectivity. Furthermore, the power allocation strategy is implemented to minimize the overall user's energy consumption while guaranteeing the SINR threshold to satisfy the expected quality of service. In constrast to existing works, the proposed scheme theoretically analyses the system modeling by taking into account the impact of massive connectivity and interference constraints for OMA systems, hence scheduling the resources to optimize the overall user's performance [18] (**Publication III**).

- Interference-aware scheme: Considering the limited computation complexity of mMTC users, the novel interference-aware scheme is proposed for the non-orthogonal multiple access (NOMA): The novel interference management technique is developed for the NOMA scheme to counteract the impact of both co-channel and inter-cell interference to enhance the UE performance. Additionally, power allocation was performed to further reduce the energy consumption. Contrary to existing works, the proposed NOMA scheduler takes into account the inter-cell interference and the reduced user's complexity that hinders the successive interference cancellation (SIC) at the receiver, hence utilizing different power coefficients to facilitate the decoding of the desired information [19] (**Publication IV**) .

- The radio access network (RAN) slice scheduler is proposed: The novel proposed scheduler utilizes machine learning to classify the users according to their channel conditions and predict their future transmission patterns that enhance the scheduling performance of a given RAN slice. In contrast to existing cooperative schedulers that share the scheduling tables for each scheduling instants hence causing the overhead on X2 interface, the proposed scheduler utilizes the users classification and prediction, hence reducing the number of shared scheduling tables for users' future transmissions (**Publication V**, Under review).

1.5 Thesis Organization

Chapter 2 presents the related works, it details standard evaluation, limitations, and the proposed solutions to the interference management problems. **Chapter 3** presents the proposed Cooperative scheduler to mitigate the impact of massive interference in the OMA scheme. **Chapter 4** presents the interference problem formulation and proposed solution for the NOMA scheme to enhance the cellular connectivity. **Chapter 5** presents the machine learning (ML)-enabled scheduler to maximize the overall throughput of the RAN slice for beyond 5G networks. Finally, the last chapter concludes the thesis by answering the research questions and suggesting directions for future work.

2 State of the Art

This section is based on our work **"Narrowband Internet of Things (NB-IoT): From Physical (PHY) and Media Access Control (MAC) Layers Perspectives**; CB Mwakwata, H Malik, MM Alam, Y Le Moullec, S Parand, S Mumtaz; Sensors 19 (11), 2613", [6].

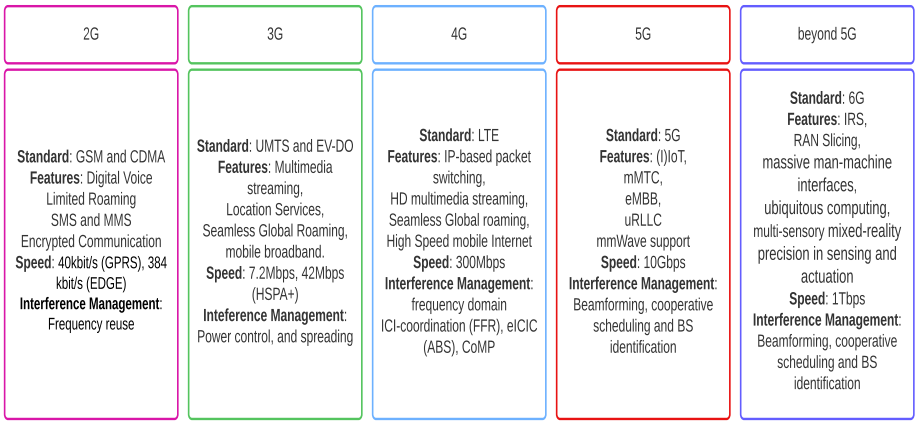

Figure 2: Evolution of 3GPP's wireless networks, from 2G to beyond 5G networks

2.1 A Brief Overview of the Evolution of Cellular Communication

Wireless communication systems evolved through several standards in the past few decades from the introduction of the first generation (1G) mobile network in the early 1980s. The rapid developments are due to the growing demand of connected users for different use-cases. Unlike in 1G where the technology involved only analogy switching to support only voice services, the second generation (2G) of mobile communication systems introduced a new digital technology for wireless transmission also known as Global System for Mobile communication (GSM). In the early 1990s, with relatively higher data rates as compared to 1G, 2G was able to support the newer services such as short message services (SMS) and email [20].

In the early 2000s, further enhancements were done on top of the existing 2G that yield to the third generation (3G) of mobile networks. Technologies such as Universal Mobile Terrestrial / Telecommunication Systems (UMTS) and the support for video calling were introduced. Furthermore, High-Speed Downlink Packet Access (HSDPA), and High-Speed Uplink Packet Access (HSUPA), were later introduced to enhance the achieved data rates. In the 2010s, the IEEE introduced the fourth generation (4G) of mobile networks which is the enhanced version of 3G networks, with even higher data rates, and capable to handle more advanced multimedia services. 4G design supports the backward and forward compatibility with legacy and even newer generations, thus facilitating easier deployment and upgrades. Contrary to previous generations, 4G supports the simultaneous transmission of voice and data which significantly improves data rate. Moreover, complex modulation schemes and carrier aggregation techniques are also possible to multiply uplink/downlink capacity [21].

The fifth-generation (5G) and beyond 5G mobile networks increase the demand for connectivity thanks to the available use cases to support both human-to-human and machine-to-machine communications. However, unlike in legacy technologies, 5G and beyond 5G networks bring into play service verticals such as massive machine-type communications (mMTC), enhanced mobile broadband (eMBB), and the ultra-reliable low latency communications (uRLLC). These verticals have different quality of service (QoS) requirements but utilize the same physical infrastructure [22].

For example, according to Information Handling Services (IHS) technology forecast, the Internet of Things (IoT) market is expected to grow to billions of devices by 2020. Massive connections are expected to respond to different IoT use cases such as smart city, smart wearable, smart home, etc. For these applications, latency-insensitive devices can be positioned in hard-to-reach areas and do not require high throughput or frequent reporting. Therefore, to cope with such tremendous IoT trends, the Third-Generation Partnership Project (3GPP) introduced the Narrow-band Internet of Things (NB-IoT) standard as a communication technology enabler. NB-IoT is categorized as one of the licensed Low-Power Wide-Area Networks (LPWAN) cellular technologies based on Long-Term Evolution (LTE) with long-range and low cost. In the LPWAN category, there exist other licensed technologies, i.e., Long-Term Evolution Category M1 (LTE-M), and unlicensed technologies, i.e., Long Range (LoRa), SigFox, Ingenu, etc., but they are not the focus of this thesis since they are not based on licensed cellular technology.

From all the previous and current telecommunication standards, interference has proved to be the major contributor to the performance degradation [23]. Furthermore, due to the co-existence of different technologies, the need to re-use frequencies among different cells becomes indispensable hence the inter-cell interference (ICI) among adjacent cells becomes a critical problem. Generally, there are three ways to manage interference. The first one is by providing orthogonality to the radio resources used by interfering users so that interference can be avoided [24]. The second one is by treating the interference as noise which is effective and practical only when weak interference is considered. And the third mitigation technique involves decoding interference by exploiting the structure of interference signals, which is rarely used in current systems due to its high computational complexity [25]. The overview of the wireless standard evolution from 2G to beyond 5G networks as well as the proposed interference management techniques is presented in Fig. 2

2.2 Challenges in Wireless Communication for Dense mMTC Deployment

Realizing the need and potential for new communication ways, 3GPP started a feasibility study on cellular system support for ultra-low complexity and low throughput IoT solution referred to as cellular IoT. In May 2014, Huawei and Vodafone proposed the Narrowband Machine to Machine (NB-M2M) to 3GPP as a study item to cope with the IoT market needs. Additional telecommunication industrial players got interested and later in the same year, Qualcomm proposed narrowband orthogonal frequency division multiplexing (NB-OFDM). In May 2015, 3GPP merged the two proposals (i.e., NB-M2M and NB-OFDM) and formed the Narrowband Cellular IoT (NB-CIoT). Eight months later, Ericsson proposed the Narrowband Long-Term Evolution NB-LTE. In September 2015, 3GPP included all proposals as a work item for Release 13. The key difference between NB-CIoT and NB-LTE was the number of reused legacy LTE network resources to support interoperability. In June 2016 NB-IoT was recognized as a new clean slate radio access technology (RAT). Only further improvements were allowed and implemented thereafter. An overview of mMTC standard evolution, where several key players of telecommunication contributed to freezing NB-IoT in 3GPP Release 13 for massive connectivity is shown in Fig. **??**.

In this regard, this section presents the mMTC technology, specifically detailing NB-IoT design changes from Release 13 until today that enabled the massive IoT connections with the corresponding solutions to respond to the adopted NB-IoT objectives. In general, the analysis is also valid for LTE-M but with small specifications when different. The enhancement features are classified following the objectives that are presented in the releases which would make it easier for the readers to refer back to the official 3GPP documents [26–32].

2.2.1 Release 13 Enhancements

3GPP introduced the following techniques in NB-IoT Release 13 to enable cellular massive IoT deployment for diverse use cases with low power, low complexity, and hence low cost. The introduced features and their corresponding objectives are as follows.

2.2.1.1 Mode of Operation With the limited bandwidth requirement, NB-IoT can be deployed in three different modes i.e., standalone, in-band, and guard-band. In in-band and guard-band modes, NB-IoT occupies one PRBs of 180 kHz in the LTE spectrum both in the downlink and uplink. It can also be allocated as standalone where it occupies the 200 kHz bandwidth by "refarming" the GSM spectrum. These flexible deployment possibilities enable fast integration and coexistence with legacy LTE and GSM systems.

2.2.1.2 Multi-Tone Transmission Support To reach the massive device deployment objective, NB-IoT introduces the allocation of Resource Units (RU) to multiple User Equipment (UE) contrary to LTE where the whole resource block is allocated to a single UE in the uplink. In this regard, tones (frequency domain) with different duration are allocated to UEs. For the uplink transmission, each tone may either occupy 3.75 kHz or 15 kHz of transmission bandwidth based on the SC-FDMA scheme; for downlink NB-IoT uses 15 kHz of transmission bandwidth with OFDM scheme as LTE. With 15 kHz spacing, NB-IoT can dedicate either single-tone (8 ms) or multi-tone (3 tones, 6 tones, and 12 tones) to different UEs with the duration of 4 ms, 2 ms, and 1 ms, respectively. On the other hand, the 3.75 kHz spacing supports only single-tone allocation to different users with 48 subcarriers of 32 ms duration [33–35].

2.2.1.3 Complexity and Cost Reduction Techniques NB-IoT is required to have low complexity to reach the low-cost objective to facilitate massive connections. The features that were implemented to reach this objective include relaxed base-band processing, low memory storage, and reduced radio-frequency (RF) components. In this regard, the system bandwidth is set as narrow as 180 kHz with reduced frequency and time synchronization requirements. Also, NB-IoT uses the restricted BPSK and QPSK modulation schemes with only one antenna support both in uplink and downlink transmission.

2.2.1.4 Power Reduction Method NB-IoT devices are intended to have a 10 years battery life to support massive deployment with limited human intervention. In this regard, two features i.e., Power Saving Mode (PSM), (from Release 12), and extended Discontinuous Reception (eDRx) (new feature from Release 13) were supported. These features are intended to extend the UE's battery longevity as follows:

In PSM, the NB-IoT device is configured to completely sleep while remaining registered online but cannot be reached by the base station signaling. In Release 13, the device can be in PSM mode for approximately up to about 413 days. In eDRX, the device is in an inactive mode for a few minutes to a few hours only. In both cases, the partial or complete inability to receive and send different signals enhance the battery life longevity; however, choosing either PSM, eDRX or both depends on the corresponding use-case requirement. In this regard, the device can be synchronized to wake up from these modes by either Real-Time Clock (RTC), triggering from sensors, or both.

2.2.1.5 Physical Channels and Signals NB-IoT adopts the same frame structure as LTE, with 1024 hyper frames, consisting of 1024 frames that contain 10 sub-frames of two slots with a duration of 0.5 ms each in the time domain. Similarly, in the frequency domain, NB-IoT contains 12 sub-carriers of 7 OFDM symbols mapped in each slot. In addition to that, when NB-IoT uses the 3.75 kHz spacing on the uplink, 48 sub-carriers are used with a slot duration of 2 ms.

The following channels and signals are used in the uplink:

- Narrowband Physical Random Access Channel (NPRACH).
- Narrowband Physical Uplink Shared Channel (NPUSCH).
- Demodulation Reference Signal (DMRS).

And the following are in the downlink frame:

- Narrowband Physical Downlink Shared Channel (NPDSCH).
- Narrowband Physical Downlink Control Channel (NPDCCH).
- Narrowband Reference Signal (NRS).
- Narrowband Primary Synchronization Signal (NPSS).
- Narrowband Secondary Synchronization Signal (NSSS).
- Narrowband Physical Broadcast Channel (NPBCH).

In general, NPRACH is used by UEs to perform initial access to the network, request transmission resources, and reconnect to the base station after a link failure. NPDSCH and NPUSCH are used to carry the downlink and uplink data packets transmissions, respectively. DMRS is used for uplink channel estimation accuracy. The UE acquires Master Information Block (MIB) from NPBCH and System Information Block (SIBs) from the NPDCCH. The defined MIB and SIB are broadcasted once during 640 ms and 2560 ms intervals, respectively. The timing of the remaining SIBs is configured in SIB1-NB. NRS is used for cell search and initial system acquisition. NPSS and NSSS are used by the UE for their frequency and timing synchronization with the base station. Due to overhead scheduling gaps in NPDCCH, the downlink and uplink peak data rates are 250 kb/s and 2267 kb/s, respectively, [28, 35–38].

2.2.1.6 Coverage Enhancement Method NB-IoT is designed to enhance coverage for the applications that are in hard-to-reach areas such as deep indoors and basements. In this regard, NB-IoT delivers an additional coverage of 20 dB as compared to the legacy LTE system. This corresponds to 164 dB of MCL. To enhance its coverage, NB-IoT uses up to 128 and 2048 re-transmissions in uplink and downlink, respectively. Hence, this makes NB-IoT suitable for use cases that are latency insensitive as it can tolerate up to 10 seconds of transmission delay.

2.2.2 Release 14 Enhancements

After the implementation of Release 13 features, studies erupted along with field trials that revealed the need for further enhancements to improve the quality of service as well as user experience. In this regard, 3GPP introduced further enhancement features to NB-IoT.

The enhancements features in Release 14 include positioning update, multicast services, and a new UE output power class in which the NB-IoT system throughput, mobility, service continuity, and non-anchor carrier operation are improved [39, 40].

2.2.2.1 Improved Positioning Technique 3GPP Release 14 introduces an indoor advanced positioning method of observed time difference of arrival (OTDOA) for NB-IoT to enhance UE position measurement of cell identity (CID). In the OTDOA method, the UE measures the times of arrival (ToAs) of positioning reference signals (PRSs) received from different transmitters to a reference node's PRS transmission to form the reference signal time difference (RSTD) measurements. In enhanced CID, the measurement requirements include the base station receive (Rx) and transmit (Tx) time difference, reference signal received power (RSRP), and reference signal received quality (RSRQ).

2.2.2.2 Multicast Services The main objective of this mechanism is to optimize resources as well as transmission latency by addressing the data to a group of UEs at the same time rather than sending it multiple times to separate devices.

Therefore in Release 14, Multimedia Broadcast Multicast Services (MBMS) is supported through single-cell point-to-multipoint (SC-PTM). In general, SC-PTM is an efficient dynamic mechanism for optimal radio resource usage as it allows broadcast or multicast services to a specific group based on real-time traffic load and user requirements. SC-PTM uses NPDSCH by mapping Single-cell MBMS Control Channel (SC-MCCH) and Single-Cell MBMS Traffic Channel (SC-MTCH) that carry control and data traffic to the physical layer scheduled by using the downlink control information (DCI).

2.2.2.3 New Power Class for Narrowband-IoT User Equipment Instead of the two power classes of Release 13 (i.e., 20 dBm and 23 dBm), in Release 14, the maximum allowed device's output power is reduced to 14 dBm. This has led to coverage relaxation of 9 dB that corresponds to 155 dB MCL as compared to 164 dB MCL and hence reduces the drained current. Technically, the use of the new power class facilitates the use of small coin-cell batteries and hence can be suitable for limited-size devices and applications that need a small battery. The compensation of the reduced NB-IoT power is achieved by increasing the NB-IoT transmission time to maintain the same energy per bit as the UE in Release 13 achieves. The newly introduced power class allows the serving base station to acquire the device power class during the establishment of the connection.

2.2.2.4 New Transport-Block-Size Support Contrary to Release 13 where NB-IoT supports relatively low data rates (250 kb/s and 226.7 kb/s in downlink and uplink, respectively), 3GPP Release 14 introduces a new NB-IoT device category which supports the improved data rates by enhancing the Transport Block Size (TBS) to 2536 bits. These data rates can be reached thanks to the ability to support a second Hybrid Automatic Repeat Request (HARQ) process. This second HARQ is useful for enhancing the reliability of the link for the UEs that experience favorable channel conditions. Implementation of this optional second HARQ process results in throughput gain as it reduces the overhead caused by NPDCCH scheduling gaps.

2.2.2.5 Multicarrier Operation To enable the massive NB-IoT deployment, in Release 14, NB-IoT can monitor paging and perform random access on non-anchor carriers. With this feature, one or more non-anchor carriers are added to the anchor carrier to carry out the synchronization and mobility measurements by using the NRS. Non-anchor carriers should also perform random access or paging when needed. Therefore, paging occasions and hence paging load will be spread over the anchor and non-anchor carriers, and all carriers can then monitor paging.

2.2.2.6 User Equipment Mobility Enhancement For the use cases that involve mobility, the temporary loss of radio interface impacts the system to a degree that can degrade link performance in terms of transmission errors. In this regard, 3GPP Release 14 introduces the possibility of Radio Resource Control (RRC) re-establishment for NB-IoT UE that supports data transfer via the control plane, i.e., the UE will try to re-establish the connection on that cell and resume the data transfer. This new RRC re-establishment feature hides the temporary loss of the radio interface to the upper layers.

2.2.3 Release 15 Enhancements

On top of all the enhancements that were introduced in Releases 13 and 14, the following improvements were introduced in Release 15 to satisfy the fast adoption of massive deployment with further improved quality of service.

2.2.3.1 Latency Reduction In Release 15, NB-IoT supports new features to further reduce the transmission delay as well as to further reduce the power consumption dissipated during long transmission requirements.

In this regard, the NB-IoT UE is now able to support the physical layer Scheduling Request (SR) which is a special physical layer message to request the network to send the access grant (DCI format 0) so that the UE can transmit the uplink data. Also, NB-IoT uses a wake-up (Wu) signal to wake up the main receiver. This signal is transmitted in idle mode only when the UE is required to decode the physical downlink control channel on paging occasions. Therefore, power consumption reduction with the wake-up signal technique is larger when the UE wakes up from deep sleep more frequently (i.e., for shorter DRX/eDRX cycles). Also, significant power consumption reduction is achieved even when a common wake-up signal is used for a group of UEs. Quick RRC release and early data transmission during random access channel (RACH) procedure are supported to reduce the UE transmission latency and hence power consumption.

2.2.3.2 Semi-Persistent Scheduling To enable better support of voice messages for the corresponding use cases, in Release 15, the Semi-Persistent Scheduling (SPS) feature is introduced. In general, SPS is comprised of persistent scheduling for initial transmissions and dynamic scheduling for re-transmissions. The base station assigns specific resource units to be used for NB-IoT UE voice messages with specific intervals to save control plane overhead and hence optimize the radio resource usage. By principle, the base station pre-configures the UE with the Radio Network Temporary Identifier (SPS-RNTI) which is used to specifically differentiate one NB-IoT UE from another or one radio channel from another. This SPS enables the NB-IoT data reception at a regular configured periodicity.

2.2.3.3 Small Cell Support To further improve the capacity as well as coverage, in Release 15, NB-IoT supports small cell deployments. The downlink power to be reused for

NB-IoT small cells is specified in section 16.2.2 of TS 36.213 [41]. In general, NB-IoT UE is not allowed to transmit more power than the configured maximum power, even if the configured power is lower than UE's maximum capability. This is done to avoid interference.

On the other hand, to extend the IoT connectivity, especially in remote and rural areas for use cases such as agriculture, logistics, and environmental monitoring, NB-IoT is now able to support up to 100 km range. According to Ericsson, this could be achieved with a software upgrade only, without any changes in the existing NB-IoT hardware [42].

2.2.3.4 Enhanced User Equipment Measurements Like in legacy LTE systems, UE measurements are critical since the corresponding reporting is mainly used to characterize the reference signal of a given bandwidth.

In Release 15, UE measurements are improved in a way that only NSSS additionally to NRS is defined for radio resource management measurement enhancement. This means that NRS is determined by the resource elements that carry NSSS in the NSSS occasions that the UE measures, through which the cell search and initial cell acquisition are improved.

2.2.3.5 Time Division Duplex (TDD) Support In Release 15, a new feature of TDD support is introduced with a new TDD frame structure (type 2). For both 3.75 kHz and 15 kHz spacing, some specified restrictions are introduced i.e., only a normal cyclic prefix is supported for NB-IoT transmission. To support some of the TDD configurations with few downlink sub-frames, some of the system information (SI) can be transmitted on non-anchor carriers. In this way, the UE will have reduced system information acquisition and search time, and hence reduced UE differentiation and access control [1, 6, 40].

2.2.4 Release 16 Enhancement

3GPP and many industrial players are involved in ongoing discussions for Release 16 enhancements. The agenda includes the following objectives with their corresponding solutions.

2.2.4.1 Grant-Free Access Most of the power consumption takes place during the NB-IoT UE active time, i.e., during Tx and Rx. In Release 16, the UE will be expected to transmit during RRC-Idle mode through Msg3 (RRC connection request) without an access grant. A UE in RRC connected mode can transmit data without grant or with the simplified control-less grant. A further enhancement is on reducing NB-IoT signaling overhead while guaranteeing the needed quality of service. These features will reduce both power consumption and latency. In Release 16, it is also proposed to further study other signal waveforms (i.e., FDMA) that require less orthogonality with more relaxed timing advance (TA) alignment as compared to SC-FDMA.

2.2.4.2 Simultaneous Multi-User Transmission The introduction of new schemes will enable simultaneous multi-user transmissions by using a shared resource in the time and frequency domains, such as Code division multiplexing (CDM), and multi-user multiple inputs multiple outputs (MU-MIMO), without increasing the number of antennae at the UE. In this regard, more dynamic access can also be achieved through enhanced base station receiver for detection of multiple users that are using the same resource unit as cluster and hence be able to schedule them effectively. This is because, for the last releases, NB-IoT

UE uses the static or semi-static configuration of more resources for the unexpected application traffic handling. Similarly, the introduction of NB-IoT transmission without grant will cause a collision of data packets so dynamic handling of multiplexing is necessary.

2.2.4.3 Enhanced Group Message Mechanism In Release 16, there should be more enhancements to support downlink commands between user groups and group RNTIs. This is because MBMS which was proposed in Release 14 is only efficient for large size downlink command message transmission and requires many UEs to be deployed. For example, the application layer common message can be very small but sent to many UEs under a small group of UEs hence making MBMS not efficient for such applications.

2.2.4.4 Inter-RAT Idle-Mode Mobility For applications such as smart tracking of logistics that involve mobility, the NB-IoT UE may still need to be accessible even when moved to the area served by other base stations.

In this regard, 3GPP should introduce the new feature for NB-IoT UE support for inter-RAT mobility during idle mode. The mentioned feature is introduced along with optional handover support during connected mode through procedure simplification i.e., without dedicated signaling for measurement control and report. This is because handover helps to reduce system information reading time.

2.2.4.5 Network Management Tool Enhancement to Improve UE Differentiation NB-IoT UE is expected to be able to perform differentiation according to maximal tolerable delay per service to optimize the radio resource usage. This is because, in the last release, the UE can be differentiated according to traffic model (periodic communication indicator, periodic time, scheduled communication time, traffic profile) and battery indication.

2.2.5 Release 17 enhancements

The primary objective of Release 17 is to improve the 5G system performance, increase the support for new use cases and verticals, and provide ubiquitous connectivity in diverse deployment conditions and scenarios. In general, Release 17 is the gateway towards 5G Advanced where AI/ML-based solutions will be used to introduce intelligent network management and solve multi-dimensional optimization issues concerning real-time and non-real-time network operation. The features include the following [43]:

2.2.5.1 Reduced-capability user equipment (RedCap UE) With additional benefits such as improved latency and the capability to operate in NR frequency bands ranging up to 52GHz, RedCap UE will fulfill service functionalities between the relaxed massive machine-type communication (mMTC) and highly stringent URLLC requirements,

2.2.5.2 Non-terrestrial networks (NTN) To complement the terrestrial networks (i.e., NB-IoT and LTE-M) for remote areas coverage over the sea and land where terrestrial coverage is absent, NTN topologies based on low Earth orbit (LEO) and geosynchronous orbit satellites are introduced. They will facilitate MBB and massive IoT services from Rel-17 onwards.

2.2.5.3 Enablers for network automation for 5G To improve the network automation, the 3GPP addressed several functionalities such as data collection from UE for analytics generation especially to ensure that slice SLA is guaranteed. Additional functionalities such as Multiple network data analytics function (NWDAF) Instances in one PLMN including hierarchies, enabling real-time or near real-time NWDAF communication, NWDAF-Assisted UP Optimization, Interaction between NWDAF and AI Models were addressed.

2.2.5.4 Support for Industrial IoT To cope with the growing demand for mMTC applications, the 3GPP introduced the support enhanced Time-Sensitive Networking (TSN), with features such as Uplink Time Synchronization, UE-UE time-sensitive communication (TSC), Exposure of Time Synchronization services for activation/deactivation, support for PTP time sync and use of Survival Time for Deterministic Applications in 5G

2.2.5.5 Enhancement of support for edge computing in 5GC The intended enhancements include the dynamic insertion of traffic offloading capabilities, seamless change of application server serving the UE, providing local applications with information on e.g. the expected QoS of the data path, supporting PSA change when the application does not support notifications of UE IP address change. These features are introduced to forward some UE application traffic to the applications/contents deployed in Edge Computing Environment hence improving the support for edge computing in 5G networks.

2.2.5.6 Architectural enhancements for 5G multicast-broadcast services 3GPP introduced the features to enhance the 5G architecture for general 5G multicast and broadcast communication services, supporting transparent IPv4/IPv6 multicast delivery, software delivery over wireless, group communications, and IoT applications, V2X applications, etc.

2.3 Radio Resource Management in mMTC

In 2G systems, the used frequencies were low band, hence the achieved coverage range was relatively large in this regard, the inter-cell interference was not a major concern hence, the frequency reuse technique was enough to mitigate the inter-cell interference. In 3G systems, the transmission technique is based on code division multiple access where the frequency reuse of factor one was allowed. In these systems, all the UEs use the same frequency resource but with different pseudo-random codes. The inter-cell interference is experienced by UEs that have the closest possible emitted powers, in this regard, power control and spreading were proposed as the interference mitigation technique.

In 4G cellular systems, the orthogonal frequency division multiplexing (OFDM) is used to efficiently utilize the available spectrum, however, the inter-cell interference is still a major challenge. In this regard, the coordination schemes such as fractional frequency reuse (FFR) are proposed to further reduce inter-cell interference. Additional inter-cell interference mitigation such as coordinated multipoint (CoMP) and almost blank subframe (ABS) was proposed as part of the enhanced inter-cell interference coordination (eICIC) technique where the radio resources as controlled employing time coordination between micro and macrocells.

The heterogeneous nature of the 5G and beyond networks and the need to support diverse use cases bring into play the advanced access techniques such as self-organized networks, directional beamforming, etc., however, despite these advancements, inter-cell interference is a major challenge, especially between the small cell mmWave radios and macro-cell radios. In this regard, it is trivial to propose the inter-cell interference management techniques that are robust and resource-efficient to guarantee the expected quality of service for all the co-existing technologies.

Several works have addressed the cell performance constraints and proposed the corresponding techniques to enhance the expected QoS. For example,

In [8] the authors studied the factors that affect cell data rate and proposed a radio resource allocation algorithm that takes into consideration the repetition factor for each user, time offset, and quality of service (QoS) constraints. In [44], the authors studied the impact of interference between NB-IoT and LTE in a coexisting network scenario; the presented analysis shows that the reduced complexity of NB-IoT user equipment (UE) makes them prone to carrier frequency offset, which significantly increases interference caused by radio frequency (RF) impairments.

In [45], the authors intended to minimize the total energy consumption subject to the computation capacity and execution latency limits. They obtained an optimal transmit power and computation resource allocation based on the Karush-Kuhn Tucker (KKT) conditions. Their results showed that the total energy consumption for both NOMA and OMA schemes increases with the number of NB-IoT user equipment (UEs). However, when compared to OMA, NOMA reduces the total energy consumption by 53.23%. Critically, it should be noted that the authors neglected the impact of inter-cell interference (ICI).

In [46], the authors derived an uplink system model for the NB-IoT IoT system. Their results reveal that the actual channel frequency response (CFR) is not a simple Fourier transform of the channel impulse response, due to sampling rate mismatch between the NB-IoT user and LTE base station. Consequently, they proposed a new channel equalization algorithm by deriving the effective CFR. In addition, they analytically derived interference to facilitate the co-existence of NB-IoT and LTE signals.

In [47], the authors investigated the downlink performance of NOMA with randomly deployed cellular users. From the presented analytical formulations, it is shown that the NOMA scheme leads to significant performance gains in terms of ergodic sum rate. However, the allocated power and the targeted data rate could directly influence the outage performance, i.e., if the allocated power is lower than the required power for successful transmission, the UE will suffer from the outage.

In [48], the authors dealt with the connection density maximization problem in NB-IoT networks by using NOMA. The authors used the bottom-up power filling algorithm and proposed item clustering heuristic approach which allows any number of devices to be multiplexed per sub-carrier. It should be noted that the authors suggested multiplexing any number per sub-carrier without considering the impact of ICI, which is a potential threat to meeting the performance requirements of NB-IoT's massive connectivity.

In [49], the authors proposed two cooperative relaying schemes i.e. ON/OFF - full-duplex relaying (ON/OFF - FDR), and ON/OFF - half-duplex relaying (ON/OFF - HDR) schemes. Either of the proposed schemes is applied to the cell center user (with good channel conditions) to help relay the direct NOMA transmissions on the downlink of cell edge users. In this regard, the ON/OFF relaying decision depends upon the quality of direct and relay links from the base station to the cell edge user. From the results, it is shown that the proposed cooperative scheme significantly improves the outage performance and the sum rate of both cell-center and cell-edge users. However, for mMTC devices such as in the LPWAN category, relaying of information leads to an increase in device complexity and cost, which is the limitation for most massive IoT use-cases.

In [50], the authors proposed a novel resource allocation technique for NOMA, based on cooperative cellular networks. In their proposed framework, the NOMA users with good channel conditions act as group heads, hence can relay information to NOMA users with bad channel conditions. Despite the gains of the proposed scheme for high complexity devices, it should be noted that the reduced complexity of NB-IoT devices, power-saving mode, and extended discontinuous reception (eDRx) make relaying of information (i.e. at the low complexity device) unfeasible.

In [51], the authors proposed a power-domain NOMA scheme for the NB-IoT systems to improve the massive connectivity by allowing UEs to simultaneously access one subcarrier. At the same time, the authors formulated a joint subcarrier and power allocation problem for both orthogonal and non-orthogonal transmissions to maximize the connection density. Their results showed that NOMA improved the number of connections by up to 87% compared to orthogonal schemes in the single-tone mode. Similarly, NOMA improved the number of connected devices as compared to multi-tone mode orthogonal schemes by 24%. Even though the NOMA approach provided enhancements over single-tone and multi-tone OMA schemes.

Similarly, radio access network (RAN) slicing is proposed as a technique to orchestrate the RAN network utilizing network virtualization and softwarization for 5G and beyond networks. In this regard, several works have studied different aspects as follows;

In [52], the authors studied RAN slicing for massive connectivity of IoT applications by optimizing the random access (RA) procedure to maximize the success probabilities and improve service multiplexing of mMTC and uRLLC traffic to save unnecessary energy consumption.

In [53], the authors adopted reinforcement learning to dynamically tune the discontinuous reception parameters to enhance the radio resource control (RRC) layer in the RAN environment by implementing their proposed architecture which is built on an open-source software platform (OAI) to create, modify and delete slices in the RRC layer to satisfy the diverse service requirements needed for IoT devices.

In [54], the authors investigated network slicing in virtualized wireless networks to solve the spectral efficiency problem by proposing a resource allocation algorithm to enhance uRLLC reliability. Even though the authors focused on eMBB and uRLLC, their work provides a framework suitable for allocation problems as compared to adaptive particle swarm optimization (APSO), equal power allocation (EPA), and equal sub-carrier allocation (ESA).

In [55], the authors revisited their previously proposed functional framework for the next generation RAN slice management to incorporate the recently specified principles and features of the new service-based management architecture in 3GPP Release 15 specifications. Furthermore, the authors presented the specific management object classes and attributes to enhance the provisioning of RAN slices and the applicability of the overall functional framework and information models in an illustrative next-generation RAN architecture. Specifically, the models are used to represent the manageable aspects of a sliced next-generation RAN operated by a neutral host provider and how the proposed functional framework operates through two examples: one illustrating the provisioning of a new RAN slice and another describing how the configuration of already activated RAN slices is modified in response to traffic demand variations.

In [56], the authors investigated the feasibility of the non-orthogonal RAN resource allocation on the uplink transmission of mMTC, eMBB, and URLLC from a common base station. Their study shows that the proposed heterogeneous non-orthogonal multiple access (H-NOMA) that involves UEs with heterogeneous service requirements can lead to significant performance improvement as compared to traditional orthogonal slicing. The enhancements are enabled by the capability of H-NOMA to provide service isolation, hence ensuring required performance thresholds for all services by leveraging their heterogeneous reliability requirements.

Moreover, several topics related to service level agreement (SLA) and the corresponding radio resource management techniques have been discussed in [57,58]. The possibility of using machine learning techniques to enhance the RAN slicing is presented in [59–61].

It should be noted that new advancements have been made to realize the goal of massive IoT under cellular technologies. For example, proactive techniques such as intelligent reflecting surfaces, that enhance the IoT links to the corresponding access point (AP) by counteracting the high pathloss, are introduced in [62]; the improved links are then exploited to better optimize the offloading of computations from the AP to the mobile edge computing (MEC) server. Similarly, proactive radio resource scheduling using machine learning techniques [63].

2.4 Scheduler Design for mMTC

Since it is trivial to guarantee the expected quality of service while providing fairness to all applications accessing the channel at a given frame in the wireless network, it is necessary to design the efficient radio resource schedulers that will be used according to the throughput performance, and intended priority by the service provider.

Scheduling algorithms, which distribute the available resources to the competing users that require to connect to the network have proved to be the key contributors to the quality of service provision in wireless networks, especially with the limited available radio resources to support a particular application. It should be noted that a large number of traffic scheduling algorithms have been proposed in the literature, however, the upcoming massive connectivity of IoT use-cases with different service requirements, the scarcity of resources, and critical delay requirements have rendered the adaptation of these schedulers very challenging.

Scheduling algorithms need to achieve fairness, efficiency and guarantee the expected quality of service for the scheduled users. Examples of classical schedulers are presented below.

2.4.1 Classical Schedulers

One of the prominent schedulers in wireless networks is Round Robin (RR). The RR assigns equal portions of packet transmission time to each user in a circular order. RR utilizes the first come first served principle, where all the users are assigned to radio resources regardless of their channel conditions, is on a first-come-first-served basis. Best channel quality indicator (CQI) scheduler is another scheduling scheme where the users with high CQI value have a higher chance to be served, and vice versa is also true. Unlike the best CQI scheduler, Proportional Fair (PF) Scheduler assigns the radio resources to users based on the level of desired fairness and experienced channel quality. PF aims at achieving a balance between Maximizing the cell throughput and fairness, by letting all users achieve a minimum quality of service. Max-Min scheduler targets at Maximizing the minimum of the UE throughput where it allocates the radio resources to guarantee the equal minimum throughput for all users. More available classical schedulers can be found in [64–66]

2.4.2 Cooperative Scheduler

In the coexistence of 5G, beyond 5G, and legacy wireless networks, the design of novel scheduling schemes that can guarantee the expected quality of service by taking into consideration not only fairness, the efficiency of the scheduled users, but also the overall cell and adjacent cells performance becomes trivial. This is because inter-cell interference from adjacent cell users has proved to be the cause of significant performance drop. Therefore, cooperative scheduling and data-driven scheduling that takes into account the inter-cell interference impact, and utilizes machine learning algorithms to classify the users accordingly, becomes of great necessity. In this regard, this thesis proposes novel cooperative scheduling frameworks to enhance the cell performance for the OMA and NOMA schemes accordingly.

2.5 Conclusion on the State of the Art

As it can be seen in the previous chapters, radio resource management in cellular mMTC is very critical, especially interference management for both orthogonal and non-orthogonal approaches to satisfy the corresponding expected quality of service requirements. Eventhough several techniques are proposed to enhance the overall cell performance, less attention has been given to mitigating the inter-cell interference and intra-call interference in heterogeneous network architecture, with diverse service requirements. In this regard, this Ph.D. thesis focuses on radio resource management, specifically, interference management techniques utilizing proactive scheduling to minimize interference to maximize cell performance.

The next chapter covers the proposed inter-cell interference management technique utilizing cooperative scheduling that is adapted for the massive connectivity of mMTC application. The overall focus is given on the interference reduction for the orthogonal multiple access techniques and adaptive power allocation to further reduce the unnecessary user's energy consumption.

3 Proposed Cooperative Scheduling in OMA Systems

This chapter presents our proposed cooperative scheduling method that can be utilized to mitigate the impact of inter-cell interference. Contrary to existing works, the cooperating base stations share their corresponding scheduling tables to be used to compute the interference weights, hence the radio resource scheduling is performed to the users with minimum impact of interference at a given transmission frame. In essence, the cooperative scheduling reduces the impact of inter-cell interference, and the performed power allocation based on users' channel conditions minimizes the unnecessary energy consumption.

This chapter is based on the following publications:

- C.B. Mwakwata, M.M. Alam, Y. Le Moullec, H. Malik, S. Pärand; Cooperative Interference Avoidance Scheduler for Radio Resource Management in NB-IoT Systems; 2020 European Conference on Networks and Communications.

- C.B. Mwakwata, O. Elgarhy, Y. Le Moullec, M.M. Alam, S. Pärand, I. Annus; Inter-cell Interference Reduction Scheme for Uplink Transmission in NB-IoT Systems; 2021 International Wireless Communications and Mobile Computing (IWCMC), 400-405

3.1 Proposed Cooperative Scheduling Method

Cooperative radio resource scheduling is considered in which three base stations are connected to communicate before the final resource allocation decision. The centralized cooperative scheduler is considered as the unit that receives the scheduling tables from cooperating base stations. At each base station, the UE channel parameters are observed, and for the cell edge users, their channel parameters are shared together with scheduling tables to the cooperative scheduler. The scheduler then computes the interference weights by taking into consideration i) one transmitting user and ii) one interfering user using the same radio resource from each base station. The users that have the minimum impact of interference are then selected and shared with the base stations to be scheduled at a given frame. When the base stations receive the list of these users and the available resources, power allocation is performed to reduce the unnecessary energy consumption due to excessive transmit power allocation for the cell edge users. The overview of the proposed scheme is shown in Algorithm 1. We also selected and implemented additional scheduling schemes i.e. proportional fair, max-min, best CQI, and round-robin as benchmarks. We fixed 10 UEs from each of the cooperating base stations and compare the performance of each scheduler.

For channel quality (CQI) estimation, our proposed scheme implements the Okumura-Hata channel model for small-medium cities. For power allocation, each base station assigns different transmit powers to their corresponding UEs such that good channel condition UEs are allowed to use a maximum transmit power of $14\ dBm$, UEs with moderate CQI are allowed to use a maximum transmit power of $20\ dBm$, and UEs with bad CQI are allowed to use the full maximum transmit power of $23\ dBm$. Compensating the reduced NB-IoT TX power (i.e. $14\ dBm$) is achieved by increasing the NB-IoT transmission time to maintain the same energy per bit like that of the UE with the maximum TX power (i.e. $23\ dBm$). Finally, power allocation is performed by considering the UE minimum and maximum power constraints as discussed in the proposed solution.

In our proposed cooperative scheduling method the base stations share (over the X2-interface) the channel quality information (CQI) (i.e. SNR, location, path-loss, cell ID, expected payload, etc.), for the devices (UE) to be scheduled for the next radio frame. The

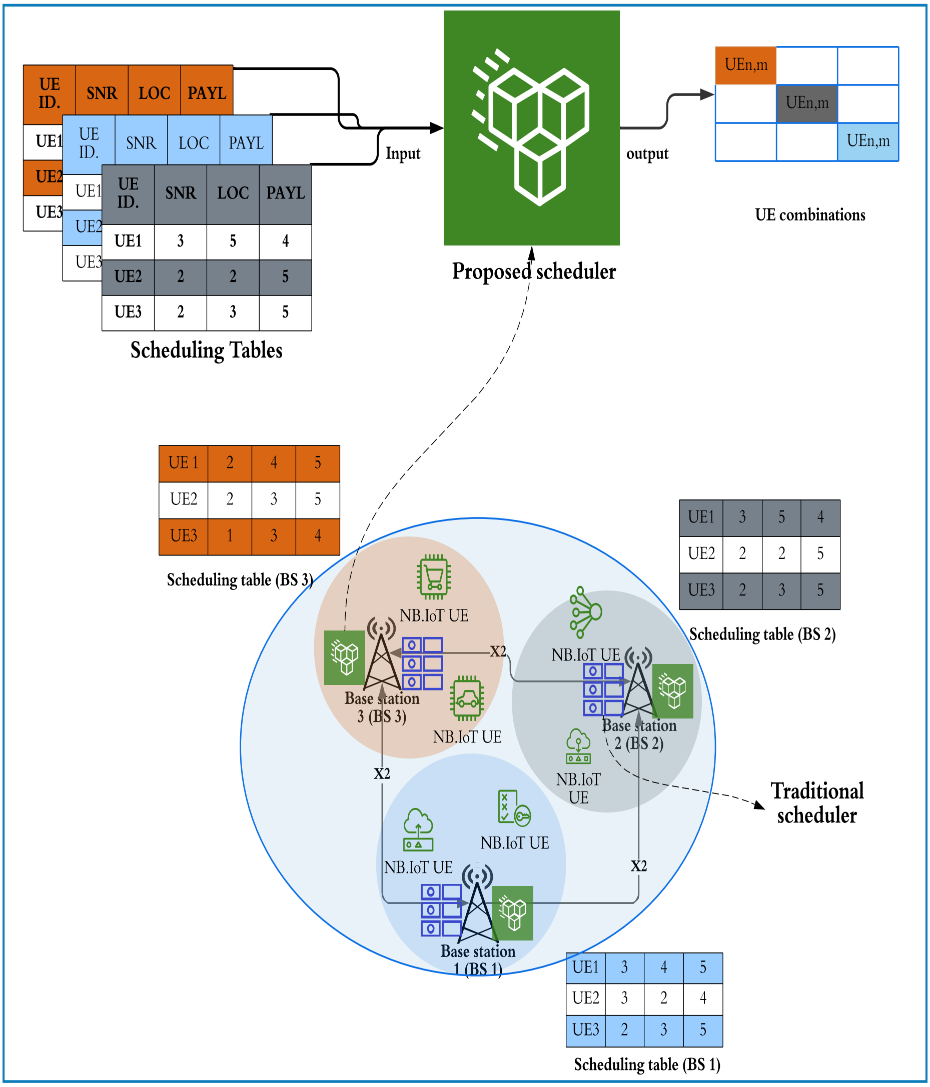

Figure 3: Proposed cooperative interference-avoidance strategy for NB-IoT system

proposed scheduler then uses this information to calculate the interference possibilities among shared UEs. The calculated interference values are hence used as input to the individual base station. Proactive scheduling is then performed by providing the available resources to UEs whose impact in terms of inter-cell interference is the lowest.

The proposed method comprises three main parts i.e. collection center, computing center, and scheduling center. The collection center receives and registers the scheduling tables from the individual base stations. The computing center computes the inter-cell interference between UEs. The scheduling center makes the final decision about the UEs that have the best throughput performance when scheduled in the same slots. The system under study is considered to be a small and medium sized city based on the Okumura-Hata channel model.

$$PL = A + B\ log(d) + C \tag{1}$$

where A, B and C depend on the antenna height and the frequency.

$$A = 69.55 + 26.16\ log(fc) - 13.82\ log(hb) - a(hm) \tag{2}$$

$$B = 44.9 - 6.55\ log(hb) \tag{3}$$

where fc and d are given in MHz and km, respectively, a and C depend on environmental factors, and hb and hm are heights for the base-station and UE, respectively.

The interference impact is based on the SINR which is calculated as shown in Equation (4).

$$SINR_{k,n}^{DL} = \frac{p_{k,n}|h_{k,n}|^2}{\sum_{m=1}^{M} \omega_{n,m} p_{m,n} |h_{m,n}^{C}|^2 + N_0 B} \tag{4}$$

where $SINR_{k,n}^{DL}$, k, $p_{k,n}$, and $|h_{k,n}|^2$ are the down-link signal to interference-plus-noise ratio, the transmit power, and channel response of user n from base station k, respectively. $\omega_{n,m}$, $p_{m,n}$, and $|h_{m,n}^{C}|^2$ are the power classes, transmit power, and channel response of user m from base station n, respectively. N_0B is the channel noise which is considered constant.

The proposed scheduler functions as follows: it receives the scheduling tables from the individual base stations and then checks for inter-cell interference; if there is interference, it checks the interference weight with all the other UEs to be scheduled in the same radio frame. If the UEs have the best throughput performance, it forwards that combination (UE identities) to be used by the individual base stations. The flowchart of the proposed inter-cell interference-avoidance algorithm for the NB-IoT system is as depicted in Fig. 4.

3.2 Simulation Setup

Extensive system-level simulations are performed to analyze the proposed method, as depicted in Fig.3. The simulation setup is considered close to the one presented in [1]; however, it is well adapted to fit the NB-IoT system. The NB-IoT UEs are considered fixed and hence the impact of Doppler spread on UE mobility is negligible. This suits well the use cases such as smart grid, smart water, smart gas metering, smart waste management, etc. [6].

The Round-Robin algorithm, as presented in [67], is used to compare the performance of the proposed algorithm. In Round-Robin, each eNB assigns the radio resources to UEs in a first-come-first-served way. That is, the first detected UE is given the available resources regardless of the impact of interference it may cause/face.

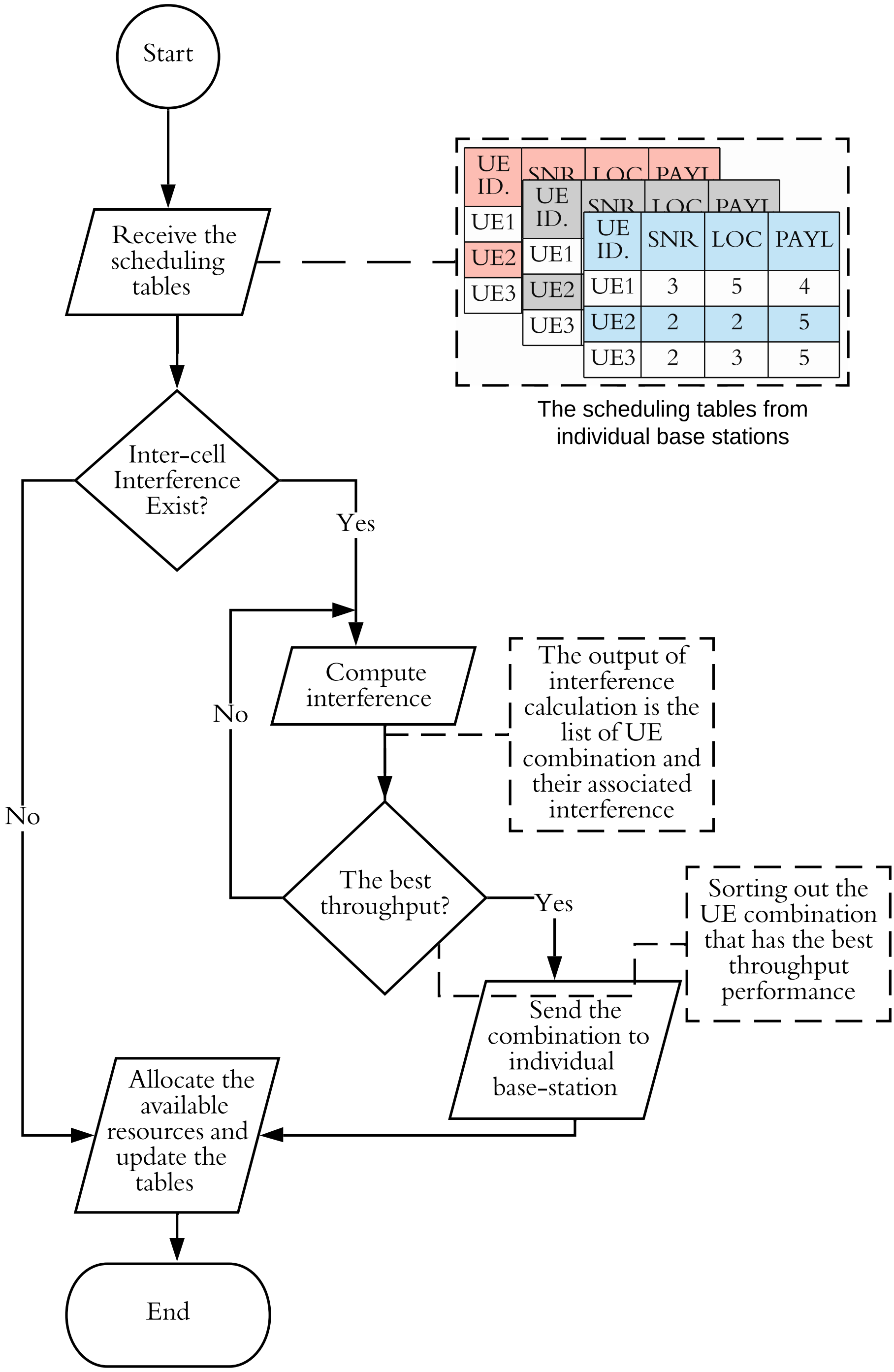

Figure 4: Flowchart of our proposed NB-IoT inter-cell interference avoidance algorithm

Table 2: Main simulation parameters for the proposed cooperative strategy for NB-IoT system [1]

Simulation Parameters	
Name	Value
(a) Transmit power of base station, UE (dBm)	46 , 23
(b) Modulation scheme	BPSK
(c) Carrier frequency (MHz)	900
(d) Receiver Thermal Noise density (dBm/Hz)	-174
(e) No. cooperating base station	3
(f) Interference Margin (dB)	0
(g) Channel model	Okumura Hata
(h) Effective Noise Power (dBm)	d + q + f + 10log(r)
(i) Required / calculated SINR (dB)	
(j) Receiver sensitivity	h + i
(k) MCL (dB)	a - j
(l) Modulation scheme	BPSK
(m) No. of antenna support per UE	1
(o) Height of base station, UE (m)	100, 1
(p) Radius of a cell	1 km
(q) Noise figure of base station, UE	9, 5 dB
(r) Occupied System bandwidth (kHz)	180

For the interference-aware scheduling algorithm to better function, it is crucial to perform proper estimation of channel parameters to effectively sort out the UEs under possible interference. This is achieved by fixing the base stations and UEs while evaluating their corresponding channel conditions based on the defined model. The important simulation parameters are as presented in Table 2 [42].

3.3 Cooperative Scheduling with Power Allocation to Mitigate ICI

3.3.1 Problem formulation

The NB-IoT uplink system has four possible resource unit configurations to choose from, here we employ the single tone resource unit configuration mode of deployment (i.e. one tone per user); however, the analysis can be replicated for other resource configurations. The tone bandwidth is given by $B_0 = B/X$, where B is the available system bandwidth and X is the resource unit spacing (i.e. for NB-IoT uplink, $B = 180\ kHz$, and $X = 15\ kHz$ or $X = 3.75\ kHz$). The index set for the available resource units is denoted as $z = \{1,2,\ldots,Z\}$. Let K_c be the set of users belonging to cell c, where the number of cells is C, i.e. $c = \{1,2,\ldots,C\}$, and k is the user index; thus, a user k in cell c will be denoted as k_c. The achievable rate of user k belonging to cell c on a given resource unit z is denoted by $R^z_{k_c} = B_0 \log_2(1 + SINR^z_{k_c})$ where $SINR^z_{k_c}$ is the signal to interference plus noise ratio experienced by user k belonging to cell c on a given resource unit z, and is given as:

$$SINR^z_{k_c} = a^z_{k_c}\left(\frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\sum_{l\neq c, l\in C}\sum_{j\in K_l}|h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} + P_n}\right) \quad (5)$$

where $|h^z_{k_c,c}|$ denotes user k_c's channel gain on resource unit z to its own base station in cell c, and $P^z_{k_c}$ denotes user k_c's transmission power on resource unit z. In the denominator, the interference term comes from other cells l, with a group of users K_l within the cell. We use j as the interference user index, thus a user j in cell l will be denoted as j_l. Moreover, $|h^z_{j_l,c}|$ denotes the channel gain of an interfering user, belonging to cell l, j_l on resource unit z to the base station of cell c, and $P^z_{j_l}$ denotes the interfering user j_l's transmission power on resource unit z. Binary variables are used for scheduling: variable $a^z_{k_c}$ denotes the resource unit occupancy coefficient such that $a^z_{k_c} = 1$ if the resource unit z is used by user k_c, and $a^z_{k_c} = 0$ otherwise. P_n is the noise power at the receiver.

The optimization goal is to minimize the inter-cell interference experienced by user k from adjacent cell users. The problem can either be modeled as the interference experienced by user k on a given resource unit, or, in order to avoid using the two binary variables, the problem can be modeled as minimizing the interference on the resource units of cell c, which is a realistic assumption since we adopted the full buffer model, where I^z_c is the interference on resource unit z in cell c. The objective function can be expressed as:

$$\min \sum_{c\in C} \sum_{z\in Z} I^z_c \tag{6}$$

Substituting the interference I^z_c, the objective function becomes:

$$\min \sum_{c\in C} \sum_{z\in Z} \sum_{l\neq c, l\in C} \sum_{j\in K_l} |h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} \tag{7}$$

subject to:

$$SINR^z_{k_c} \geq \vartheta_{k_c,min} \tag{8}$$

which is a constraint in order to satisfy the required quality of service (QoS), where $\vartheta_{k_c,min}$ is the minimum acceptable $SINR$ that user k_c can have to satisfy the QoS,

$$a^z_{k_c}\Big(\frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\sum_{l\neq c, l\in C}\sum_{j\in K_l} |h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} + P_n}\Big) \geq \vartheta_{k_c,min} \tag{9}$$

$$0 \leq P^z_{k_c} a^z_{k_c} \leq P_{max}, \forall c \in C, \forall k \in K_c, \forall z \in Z \tag{10}$$

where P_{max} is the maximum allowed transmit power per device.

$$\sum_{k\in K_c} a^z_{k_c} \leq 1, \forall z \in Z, c \in C. \tag{11}$$

The above constraint guarantees that, within the same cell, a resource unit can only be occupied by one user at a given time.

3.3.2 Problem solution

It can be noted that the proposed optimization problem in this work is a mixed binary integer non-linear programming (MBINP) problem. The variables to be optimized are $a^z_{k_c}$ and $P^z_{k_c} a^z_{k_c}$ which are very difficult to solve. In this regard, we apply a step-wise algorithm [58] for resource unit and power allocation. The algorithm follows three main steps: the first step initializes the transmit power, the second step performs the resource unit allocation, and the third step optimizes the power allocation.

- **Step One: transmit power initialization**

To perform the resource unit allocation, we need to establish an initial transmit power. For our analysis we choose one of the following two methods:

1. setting the initial transmit power equal to the maximum allowed power per device, i.e. P_{max}, or

2. we set the transmit power equal to a required power, calculated from the relation given by the minimum required SINR and average interference power, or maximum tolerated interference level if either is available for the use case or scenario.

Regarding the first option, the transmit power of the user is set to the maximum allowed power for the device.

For the second one, we have a known interference level, e.g average, tolerable, threshold, etc. We define this level at $In^z_{k_c}$. We can compute the transmit power by performing the following procedure:

The SINR for a generic user k_c with a known noise and interference level, $In^z_{k_c}$, is given by:

$$SINR^z_{k_c} = \frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{In^z_{k_c}} \tag{12}$$

The transmit power can be calculated as:

$$P^z_{k_c} = \frac{In^z_{k_c} SINR^z_{k_c}}{|h^z_{k_c,c}|^2} \tag{13}$$

By using the minimum acceptable $SINR$ per user, we have an inequality:

$$P^z_{k_c} \geq \frac{In^z_{k_c} \vartheta_{k_c,min}}{|h^z_{k_c,c}|^2} \tag{14}$$

Thus, we take the lowest acceptable transmit power, i.e. the equality case:

$$P^z_{k_c} = \frac{In^z_{k_c} \vartheta_{k_c,min}}{|h^z_{k_c,c}|^2} \tag{15}$$

- **Step two: resource unit allocation**

After step one, the optimization problem can be written as a function of the resource unit allocation binary variable only, for a fixed transmit power, thus no power constraint:

$$\min \sum_{c\in C} \sum_{z\in Z} \sum_{l\neq c, l\in C} \sum_{j\in K_l} |h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} \tag{16}$$

subject to:

$$SINR^z_{k_c} \geq \vartheta_{k_c,min} \tag{17}$$

$$\sum_{k\in K_c} a^z_{k_c} \leq 1, \forall z \in Z, c \in C. \tag{18}$$

It can further be seen that the optimization problem is a 0-1 assignment integer linear programming problem about $a^z_{k_c}$. To obtain the solution, we use the cooperative scheduling scheme which is discussed later on in Section 3.3.3.

- **Step three: transmit power allocation**

After performing the resource unit allocation, the result will have several interfering users on each resource unit. Moreover, these interfering users on a specific resource unit are not interfering with other users on any other resource unit because of the OMA scheme; i.e., intra-cell interference can be ignored. Thus, the transmit power allocation for the optimization problem, to minimize interference, can be solved for each resource unit separately because the solutions for each resource unit are independent from each other [58]. Thus, the problem will be solved per resource unit having one interfering user per cell, then this can be repeated for all the other resource units which are occupied by interfering users.

In order to follow the above approach, we reformulate the problem. The new problem has one user per cell interfering with each other. The optimization problem is to find the power allocation among interference users as in [68, 69].

The optimization problem, after the resource unit allocation step, can be written as:

$$\min \sum_{c \in C} \sum_{l \neq c, l \in C} |h_{l,c}|^2 P_l \tag{19}$$

subject to:

$$SINR_c \geq \vartheta_{c,min} \tag{20}$$

$$0 \leq P_c \leq P_{max}, \forall c \in C. \tag{21}$$

where $|h_{c,c}|^2$ is the channel gain of the user from cell c to its base station c, P_c is the transmit power of this user belonging to cell c, $|h_{l,c}|^2$ is the channel gain of the interfering user belonging to cell l to the base station c, and P_l is this interfering user's transmit power.

The $SINR$ for the user belonging to cell c on a given resource unit is given as:

$$SINR_c = \frac{|h_{c,c}|^2 P_c}{\sum_{l \neq c} |h_{l,c}|^2 P_l + P_n} \tag{22}$$

The constraint (20) is not linear. In the following steps, we linearize this constraint .

We start by substituting 22 in 20.

$$\frac{|h_{c,c}|^2 P_c}{\sum_{l \neq c} |h_{l,c}|^2 P_l + P_n} \geq \vartheta_{c,min} \tag{23}$$

where $\vartheta_{c,min}$ is the minimum required $SINR$ to satisfy the required quality of service for the user belonging to cell c.

By doing cross multiplication

$$|h_{c,c}|^2 P_c \geq \vartheta_{c,min} \Big(\sum_{l \neq c} |h_{l,c}|^2 P_l + P_n \Big) \tag{24}$$

and equivalently,

$$-|h_{c,c}|^2 P_c + \vartheta_{c,min} \Big(\sum_{l \neq c} |h_{l,c}|^2 P_l \Big) \leq -\vartheta_{c,min} P_n \tag{25}$$

By expanding for $c = 1, 2, \ldots, C$ the inequalities become:

$$\begin{aligned} c = 1 : -|h_{1,1}|^2 p_1 + \vartheta_1 |h_{2,1}|^2 p_2 + \vartheta_1 |h_{3,1}|^2 p_3 +, \\ \ldots, + \vartheta_1 |h_{C,1}|^2 p_C \leq \vartheta_1 P_n \end{aligned}$$

$$c=2: -|h_{2,2}|^2 p_2 + \vartheta_2 |h_{1,2}|^2 p_1 + \vartheta_2 |h_{3,2}|^2 p_3 +, \\ \ldots, + \vartheta_2 |h_{C,2}|^2 p_C \leq \vartheta_2 P_n$$

$$c=3: -|h_{3,3}|^2 p_3 + \vartheta_3 |h_{1,3}|^2 p_1 + \vartheta_3 |h_{2,3}|^2 p_2 +, \\ \ldots, + \vartheta_3 |h_{C,3}|^2 p_C \leq \vartheta_3 P_n$$

$\vdots$

etc, which can be written in matrix form such as

$$\tilde{A}\tilde{p} \leq \tilde{c} \tag{26}$$

This can be solved by linear programming solutions in Matlab.

3.3.3 Implemented OMA Algorithm

Algorithm 1 OMA

procedure Generate UE parameters ▷ using Okumura Hata channel model
    $k \leftarrow |h^z_{k_c,c}|^2$
    while $P^z_{k_c} \neq 0, j \neq i$ **do**
        $SINR^z_{k_c} \leftarrow a^z_{k_c} \left(\frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\sum_{l \neq c, l \in C} \sum_{j \in K_l} |h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} + P_n} \right)$
    return $SINR^z_{k_c}$ ▷ along with other channel parameters
procedure Share to the scheduler ▷ to compute interference weights
    while $In^z_{k_c} = \frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\vartheta_{i,min}} - P_n$ **do**
        $select\_the\_UEs\_from\_each\_cell$
    return k ▷ UE IDs for available resources
procedure Power Allocation
    $\frac{In^z_{k_c} \vartheta_{k_c,min}}{|h^z_{k_c,c}|^2} \leftarrow p$
    while constraint 14 is satisfied **do**
        $calculate\_Rate\_R_k$
    return R_k

For fairness analysis, we use the Jain's fairness index [70], which is given as:

$$f(R_1, R_2, \ldots, R_k) = \frac{\left[\sum_{k=1}^{K} R_k\right]^2}{K \sum_{k=1}^{K} (R_k)^2} \tag{27}$$

where K is the total number of UEs under analysis and R_k is the instantaneous UE data rate of user k. With this metric, the fairness is highest when the index value is equal to 1 and the lowest when it is equal to 0.

Additionally, the non-orthogonal multiple access (NOMA) scheme is considered to be the promising technique to provide capacity enhancement of above $100,000$ devices per cell Contrary to the OMA approach, the NOMA approach gives the possibility to simultaneously superpose multiple devices in a given available radio resource by allocating different power coefficients or codes to enable the successive interference cancellation (SIC) at the receiver. In this regard, NOMA brings an exponential increase in device support as compared to OMA, but at the cost of increased receiver complexity.

Despite the advantages that NOMA brings to 5G and B5G networks, it is still unclear if it can be implemented in low-power IoT devices. This is because NOMA involves superposition coding (SC) and SIC at the transmitter and receiver, respectively, which are highly computationally complex for mMTC applications.

Furthermore, for both OMA and NOMA approaches, if the radio resources are not well managed, the massive connectivity will lead to massive interference, which will severely degrade the performance of legacy, 5G, and B5G network systems. In this regard, the following chapter discusses our proposed approach adapted for NOMA systems.

4 Proposed Cooperative Scheduling in NOMA Systems

This chapter presents the proposed non-orthogonal multiple access (NOMA) scheduling for radio resource management in massive machine type communications. Contrary to existing works, the proposed NOMA scheduling method mitigates the impact of intra-cell and inter-cell interference through cooperation with adjacent base stations for radio resource management. Considering reduced complexity that hinders the performance of successive interference cancellation (SIC) at the receiver, we consider offloading the SIC at the corresponding base stations. Additionally, power allocation is performed to reduce the unnecessary user's energy consumption caused by exhaustive repetitions. Finally, the cell performance gains, complexity, and the fairness of the proposed scheduling method are discussed and included in the corresponding appendix.

This chapter is based on the following publication:

- C.B. Mwakwata, O. Elgarhy, M.M. Alam, Y. Le Moullec, S. Pärand, K. Trichias, K. Ramantas; Cooperative Scheduler to Enhance Massive Connectivity in 5G and Beyond by Minimizing Interference in OMA and NOMA; 2021 IEEE Systems Journal.

4.1 System Model and Problem Formulation for NOMA

We consider a system of x transmitting users served by cooperating base stations, and x={1,2, ..., X} be its index set of users. We consider M to be a positive, maximum number of devices that can be supported per sub-carrier. $z = \{1, 2, \ldots, Z\}$ represents the index of the resource units. x_c represents the cell c's UEs, and C, i.e $c = \{1, 2, \ldots, C\}$, represents the number of cells used in simulation. Therefore, the signal to interference plus noise ratio of the NOMA user x_c at unit z is given as:

$$SINR^z_{x_c,NOMA} = a^z_{x_c}\left(\frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{I^z_c + \sigma_N}\right) \tag{28}$$

where I^z_c is the total interference experienced by user x_c from the co-allocated interfering users i and users l from adjacent cells, which is given as

$$I^z_c = \sum_{i \neq k, i \in M} |h^z_{i,c}|^2 P^z_i a^z_{i_c} + \sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} \tag{29}$$

As was also the case for OMA, we aim to minimize the ICI at user x_c from users q_l, and interference from the NOMA users i of the same cell assigned to the same resource unit. The objective function can therefore be expressed as:

$$\min \sum_{c \in C} \sum_{z \in Z} \left(\sum_{i \neq k, i \in M} |h^z_{i,c}|^2 P^z_i a^z_{i_c} + \sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} \right) \tag{30}$$

subject to;

$$a^z_{x_c}\left(\frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{I^z_c + \sigma_N}\right) \geq \vartheta_{x_c,lim} \tag{31}$$

$$0 \leq P^z_{x_c} a^z_{x_c} \leq P_{maxm}, \forall c \in C, \forall x_c \in X, \forall z \in Z. \tag{32}$$

where P_{maxm} is the maximum allowed power per device.

$$\sum_{x_c \in X} a^z_{x_c} \leq 1, \forall z \in Z, c \in C. \tag{33}$$

$$\sum_{x_c \in X} a^z_{x_c} \leq M, \forall i \in M \forall z \in Z, c \in C. \tag{34}$$

It can be seen that the objective function is a combinatorial optimization problem and is hence difficult to solve. In this regard, the proposed solution is presented as follows.

4.2 Proposed Solution for NOMA in NB-IoT Systems

To solve the NOMA problem we follow the same steps as in OMA. Firstly, we set an initial interference power for all the users. Secondly, we perform the scheduling for all the users. Finally, we implement the power allocation to further reduce the interfering powers at the desired receiver. The initial interference power will be allocated as we did for OMA. However, the channel allocation problem in equation 30 will have two assumptions:

- The power is not a variable,
- There are no power constraints.

Therefore, we perform power allocation after the channel assignment. We rewrite the optimization problem in a similar way to that of OMA, i.e. by working per resource unit since there is no interference from adjacent resource units; however, we have to add the NOMA interference users in a given resource unit. Since in the OMA we had one user per resource unit per cell, there was no need to add a subscript for the resource unit. However, because of NOMA, we have more than one user, thus, we define M_c as the group of NOMA users per cell per resource unit, and x_c is a user in cell c that belongs to M_c, and we omit the resource unit index. In this regard, the optimization goal becomes:

$$\min \sum_{c \in C} \sum_{x_c \in M_l} \Big(\sum_{l \neq c, l \in C} \sum_{j \in M} |h^j_{l,c}|^2 P^j_l + \sum_{y \neq x_c, y \in M_c} |h^z_{c,c}|^2 P^y_c \Big) \tag{35}$$

where $h^j_{l,c}$ is the channel gain from user j, belonging to cell l and the NOMA group M_l within the cell, on cell c. P^j_l is the power of this user. These two terms represent the intercell interference from all the NOMA users of other cells. As for the NOMA part; $h^z_{c,c}$ is the channel gain of NOMA user y belonging to the same group M_c.

subject to:

$$SINR^z_{x_c,NOMA} \geq \vartheta_{c,lim} \tag{36}$$

$$0 \leq P^z_{x_c} \leq P_{maxm}, \forall c \in C. \tag{37}$$

The $SINR^z_{x_c,NOMA}$ is then given as:

$$SINR^z_{x_c,NOMA} = \left(\frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l \neq c} \sum_{j \in M_l} |h^j_{l,c}|^2 P^j_l + \sigma_N + \sum_{y \neq x_c, y \in M_c} |h^z_{c,c}|^2 P^y_c} \right) \tag{38}$$

which can be solved in the same way as for OMA. However, the number of inequalities will be larger. Moreover, this equation does not take into account the SIC effect on removing interference from other NOMA users within the same cell in the same resource unit. The effect of the SIC can be included in the inequalities by simply putting zero for the NOMA interference users within the same cell as the main user after passing through the SIC. The constraint (36) is not linear; in this regard, we start by substituting equation 38 into eqn. 36, hence linearize as follows;

$$\Big(\frac{|h^k_{c,c}|^2 P^k_c}{\begin{array}{c}\sum_{l\neq c}\sum_{j\in M_l}|h^j_{l,c}|^2 P^j_l+\sigma_N\\ +\sum_{y\neq x_c,y\in M_c}|h^z_{c,c}|^2 P^y_c\end{array}}\Big)\geq \vartheta_{c,lim} \tag{39}$$

$$|h^k_{c,c}|^2 P^k_c \geq \vartheta_{c,lim}\Big(\sum_{l\neq c}\sum_{j\in M_l}|h^j_{l,c}|^2 P^j_l+\sum_{y\neq x_c,y\in M_c}|h^z_{c,c}|^2 P^y_c+\sigma_N\Big) \tag{40}$$

$$\begin{aligned}&|h^k_{c,c}|^2 P^k_c-\vartheta_{c,lim}\Big(\sum_{l\neq c}\sum_{j\in M_l}|h^j_{l,c}|^2 P^j_l\Big)\\ &-\vartheta_{c,lim}\Big(\sum_{y\neq x_c,y\in M_c}|h^z_{c,c}|^2 P^y_c\Big)\geq \vartheta_{c,lim}\sigma_N\end{aligned} \tag{41}$$

equivalently,

$$\begin{aligned}&-|h^k_{c,c}|^2 P^k_c+\vartheta_{c,lim}\Big(\sum_{l\neq c}\sum_{j\in M_l}|h^j_{l,c}|^2 P^j_l\Big)\\ &+\vartheta_{c,lim}\Big(\sum_{y\neq x_c,y\in M_c}|h^z_{c,c}|^2 P^y_c\Big)\leq \vartheta_{c,lim}\sigma_N\end{aligned} \tag{42}$$

Substituting $c = 1,2,\ldots,C$ equation becomes:

$$\begin{aligned}c=1,k=1: -|h^1_{1,1}|^2 p^1_1+\vartheta_{1,min}\big(|h^1_{2,1}|^2 p^1_2+|h^2_{2,1}|^2 p^2_2+\ldots\\ +|h^1_{3,1}|^2 p^1_3+|h^2_{3,1}|^2 p^2_3+\ldots,\\ \ldots,+|h^1_{C,1}|^2 p^1_C+|h^2_{C,1}|^2 p^2_C+\ldots\big)+\\ \vartheta_{1,min}\big(|h^2_{1,1}|^2 p^2_1+|h^3_{1,1}|^2 p^3_1+,\\ \ldots,+|h^{M_1}_{1,1}|^2 p^{M_1}_1\big)\leq \vartheta_{1,lim}\sigma_N\end{aligned}$$

$$\begin{aligned}c=1,k=2: -|h^2_{1,1}|^2 p^2_1+\vartheta_{1,min}\big(|h^1_{2,1}|^2 p^1_2+|h^2_{2,1}|^2 p^2_2+\ldots\\ +|h^1_{3,1}|^2 p^1_3+|h^2_{3,1}|^2 p^2_3+\ldots,\\ \ldots,+|h^1_{C,1}|^2 p^1_C+|h^2_{C,1}|^2 p^2_C+\ldots\big)+\\ \vartheta_{1,min}\big(|h^1_{1,1}|^2 p^1_1+|h^3_{1,1}|^2 p^3_1+,\\ \ldots,+|h^{M_1}_{1,1}|^2 p^{M_1}_1\big)\leq \vartheta_{2,lim}\sigma_N\end{aligned}$$

$$\begin{aligned}c=2,k=1: -|h^1_{2,2}|^2 p^1_2+\vartheta_{2,min}\big(|h^1_{1,2}|^2 p^1_1+|h^2_{1,2}|^2 p^2_1+\ldots\\ +|h^1_{3,2}|^2 p^1_3+|h^2_{3,2}|^2 p^2_3+\ldots,\\ \ldots,+|h^1_{C,2}|^2 p^1_C+|h^2_{C,2}|^2 p^2_C+\ldots\big)+\\ \vartheta_{2,min}\big(|h^2_{2,2}|^2 p^2_2+|h^3_{2,2}|^2 p^3_2+,\\ \ldots,+|h^{M_2}_{2,2}|^2 p^{M_2}_2\big)\leq \vartheta_{3,lim}\sigma_N\end{aligned}$$

$\vdots$

etc. The above expansion can be shorten as a matrix of the following form:

$$\tilde{B}\tilde{q} \leq \tilde{v} \tag{43}$$

In this regard, equation 42 can be solved by linear programming solutions in Matlab. Algorithm 2 presents the proposed implementation of the NOMA approach; simulation parameters are presented in Table 2, unless specified otherwise.

4.3 Complexity Analysis

As seen in Algorithm 2, from line 1 to line 5 the algorithm computes the channel parameters for all users attached to the corresponding base stations. This operation has a computation cost of $O(n)$. Then from line 6 to line 14, there is the nested while or for-loop such that in the first loop, the interference weight is analyzed, and users (i.e., which have lower interference impact on each other) are superposed at a given sub-carrier. In the second loop, the transmit power is allocated to users to reduce unnecessary energy consumption. This operation has the computation cost of $O(n^2)$. From line 16 to the end of the algorithm, we evaluate the achieved user performance and the computation cost is $O(n)$. In this regard, the computation complexity becomes:

$$O(n+n^2+n) \tag{44}$$

Thus, the computational complexity of our proposed algorithm is $O(n^2)$, i.e. quadratic complexity.

If we analyze the computation complexity in the single form (i.e., without considering the interference impact), from line 1 to line 5 the algorithm computes the channel parameters for all users attached to the corresponding base stations. The operation still has a computation cost of $O(n)$. However, from line 6 to line 14, we will have only one whole or for-loop to allocate different power coefficients to NOMA users to enable the SIC at the receiver. This operation has a computation cost of $O(n)$. From line 16 to the end of the algorithm, we evaluate the achieved user performance and the computation cost remains $O(n)$. In this regard, if we do not consider interference reduction, then the computation cost becomes $O(n+n+n) = O(n)$.

Therefore, the complexity overhead of our proposed scheme ($O(n^2)$ vs. $O(n)$) is an acceptable trade-off, given the performance enhancements brought by the interference reduction.

The cooperative scheduling method follows the implementations as for OMA however, for the NOMA, each base station classifies the UEs into three groups based on their channel parameters, i.e., good, moderate, and bad UEs. We assume that we have two main sources of interference i.e., NOMA interference from users that are simultaneously allocated at a given resource unit at a given time slot, and the ICI from other users transmitting at the same resource unit but from adjacent cells. Then the scheduling tables from each base station are shared with the cooperative scheduler. After receiving the tables, the scheduler selects one UE from each group of users to be simultaneously superposed at a given resource unit.

In this regard, a maximum number of 3 UEs can simultaneously occupy a given resource unit at a given time slot. The scheduler computes the best combination of UEs for all the available resource units before sharing the respective allocation of slots within a frame to the base stations. Additionally, the scheduler performs the power allocation to reduce the impact of co-channel interference as well as ICI. During power allocation, we assume the power constraints for each group as follows: good channel users $P_{const} = 14$

dBm, moderate channel users $P_{const} = 20\ dBm$, and bad channel users $P_{const} = 23\ dBm$. Different power coefficients are assigned to users to successfully perform SIC at the receiving base station. We assume that the good channel users are close to their serving base station and hence can be given lower power constraints, and vice-versa is true for bad channel users. An overview of the proposed cooperative scheduling method for OMA and NOMA is presented in Fig. 5.

In general, unlike the joint processing in coordinated multi-point (CoMP) in LTE systems where a UE at the cell-edge is served by two or more base stations to improve signals quality and increase throughput [71], in our proposed cooperative scheduling method each base station serves its users. The simulation parameters are similar as presented in [18], with some modifications adapted for the NOMA. The overview of the followed steps is highlighted in Algorithms 1 and 2. We also selected additional scheduling schemes i.e. proportional fair (PF), max-min, and round-robin as benchmarks for comparison purposes.

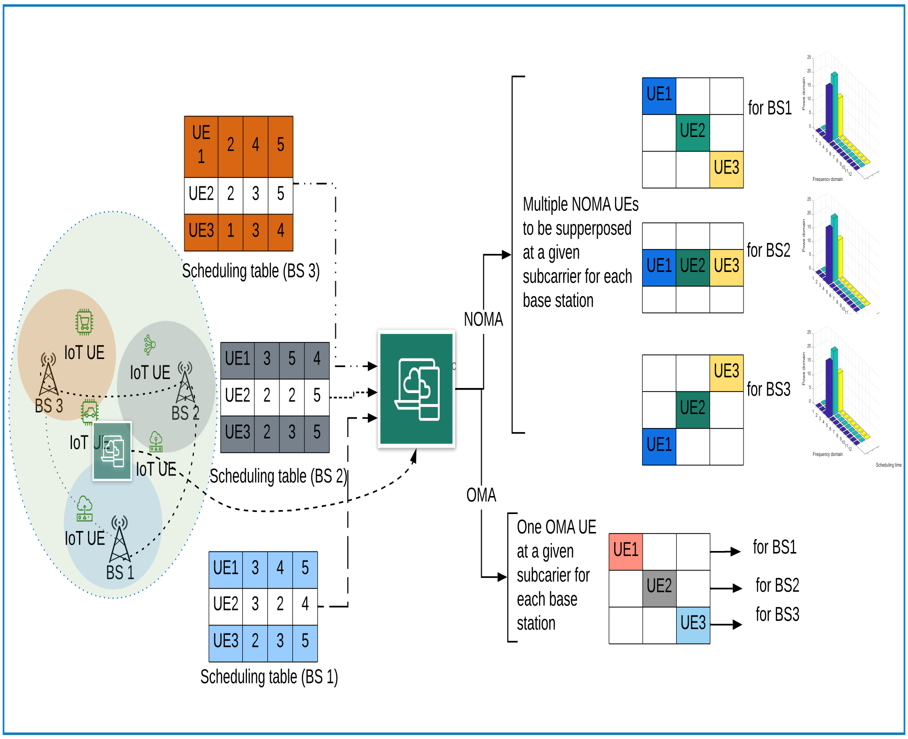

Figure 5: Proposed radio resource management scheduling exploiting the NOMA scheme in NB-IoT systems. Each cooperating base station (BS1 to BS3) share their respective scheduling tables for their future transmission. Then ICI is avoided by allocating resources to UEs whose impact in terms of interference is the lowest among the UEs. Then the base stations implement the OMA or NOMA scheme for their corresponding choice of strategy

One of the main goals of mMTC is to support a higher connection density of up to billions of devices per square kilometer [68,72] for beyond 5G networks. The aim of eMBB is to deliver peak data rates of up to 20 Gbps, peak spectral efficiency of up to 30 bps/Hz and 15 bps/Hz on the downlink and uplink, respectively, maximum tolerable latency of 4 ms, and high energy efficiency with seamless mobility support. Finally, uRLLC provides up

Algorithm 2 Implemented NOMA Algorithm

procedure User Equipment Creation ▷
  $x_c \leftarrow |h^z_{x_c,c}|^2$
  while $P^z_{x_c} \neq 0, j \neq i$ **do**
    $Equation(38)$
  return $SINR^z_{x_c}$
procedure Share the scheduling tables
  while $In^z_{x_c} = \frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\vartheta_{i,min}} - \sigma_N$ **do**
    $compute\_the\_best\_combination\_of\_UEs$
    $Divide\_the\_UEs\_in\_three\_groups$
    $Superpose\_One\_UE\_from\_each\ group\_in\_a\_given\_subcarrier$
  while $P^z_{x_c} \neq 0$ **do**
    $allocate\_power\_according\_to\_constraint$:
    $0 < P^z_{x_c} \leq P_{maxm}, \forall c \in C$
  return x_c
procedure Evaluate
  while $\frac{I^z_c \vartheta_{c,lim}}{|h^z_{x_c,c}|^2} \leftarrow p$ **do**
    $calculate\_Rate\_R_k$
    $calculate\_Energy\_Consumption$
  return $R_{x_c}, energy$

to only 1 ms user-plane latency, reliability of $1-10^{-5}$ success probability for transmitting 32 bytes in 1 ms with 0 ms mobility interruption time. In this regard, it is very challenging to support this diversity of service requirements under the same physical infrastructure without interrupting the services that are simultaneously running on the same network.

To cope with such a challenge, RAN slicing has been proposed as one of the enablers of network orchestration by flexibly customizing and managing the base stations utilizing softwarization and virtualization to support a multi-service-multi-tenant architecture [69]. In this regard, a RAN slice can be characterized by particular QoS requirements that necessitate a particular system behavior to support specific applications. For example, user equipment (UE) with reduced capabilities (REDCAP) operating under 5G NR Light [73], or narrow-band IoT (NB-IoT) can be served by a slice with radio access that is up to 10 seconds delay-tolerant with very limited or no mobility. On the other hand, Cat-M UEs can be served by a slice that guarantees mobility support for applications such as smart logistics [74, 75].

Radio resource scheduling is one of the most proposed approaches to enable optimal resource usage and to enhance the massive connectivity of mMTC UEs. However, if the radio resources are not well managed, the massive connectivity causes massive interference between UEs that are competing for the available radio resources in the heterogeneous network. One of the promising techniques to enable better scheduling by minimizing the inter-cell interference and guaranteeing the required QoS is the use of a cooperative scheduler as seen in previous chapters.

However, sharing of the scheduling tables increases the overhead in the X2 interface, and the shared data need the brute-force computation to select the best pair to be scheduled at given radio resources. Therefore it is necessary to study the proactive data-driven approaches that can reduce the computational complexity, proactively classify and predicts the next transmissions to enable proactive scheduling.

It should be noted that less attention is given to how the machine learning techniques can be used to classify and predict the users' transmission characteristics, hence enhancing the scheduling of RAN slices. In this regard, the next chapter studies the applicability of machine learning algorithms adapted to slice scheduling to increase the number of connected devices per slice, while providing the expected QoS requirements according to shared SLA.

5 Exploiting Machine Learning for RAN Slice Scheduling in Beyond 5G Communication

In this chapter, the radio access networks (RAN) slice scheduling is developed to maximize the cell performance and the number of the connected devices to support the massive connectivity in 5G and beyond wireless networks. The machine learning algorithms are used to perform the users' classification based on their channel conditions, and prediction for their future transmission patterns to enhance the scheduling. Unlike traditional cooperative scheduling which involves sharing of scheduling tables through X2 interface hence causing overhead, in this work, machine learning is used to predict the users' transmission patterns hence avoiding exchange of tables for every scheduling frame to reduce the overhead. Similarly, the classification and prediction helps to perform the radio resource scheduling to the users whose channel conditions can guarantee the successful transmissions in their next allocations.

This chapter is based on the following publication:

- C.B. Mwakwata, M.M. Alam, Y. Le Moulec; mMTC Users Classification Empowering Predictive Cooperative Scheduler in RAN Slicing for 5G and Beyond Networks; **Submitted**

5.1 Problem Formulation

We consider a multi-cell network structure where several UEs are transmitting to their corresponding base stations. In this scenario, adjacent base stations simultaneously receive the unwanted signals transmitted by adjacent cell UEs. In this regard, inter-cell interference in terms of transmit power is experienced. We assume the scheduling is performed for a given slice; however, the analysis can be replicated for several slices. Let $z = \{1,2,\ldots,Z\}$ be the index of the resource units. x_c represents the cell c's UEs, and C, i.e $c = \{1,2,\ldots,C\}$, is the number of cells used in the simulation. The achieved data rate of a given slice is derived from the Shannon formula of cell capacity given by

$$R_{SLC} = B_{SLC} log(1 + SINR_{SLC}) \tag{45}$$

where R_{SLC} is the achieved rate of a given slice based on the shared service level agreement (SLA). B_{SLC} is the allocated bandwidth for a given slice in order to satisfy the expected QoS, given by

$$B_{SLC} = \frac{B}{\mu} \tag{46}$$

where B is the overall system bandwidth, and μ is the bandwidth splitting coefficient that depends on the SLA. Finally, $SINR_{SLC}$ is the achieved signal to noise plus interference ratio of the allocated UE, given by

$$SINR_{SLC} = \frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N} \tag{47}$$

where $|h^z_{x_c,c}|^2$ represents the channel gain of the transmitting UE, and $P^z_{x_c}$ is the transmitting power of the allocated UE, which is subject to maximum allowed power constraint per each transmitting UE given by

$$0 \leq P^z_{x_c} \leq P_{maxm}, \forall c \in C. \tag{48}$$

Therefore, we aim to maximize the sum rate of the of the system by maximizing the rate of each allocated slice while guaranteeing the expected QoS of each slice. In this regard, the sum rate maximization problem can be represented as:

$$\max \sum_{c \in C} \sum_{z \in Z} a^z_{x_c} \log\big(1 + \frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N}\big) \tag{49}$$

Subject to:

$$SINR_{SLC} \geq \vartheta_{x_c,SLC} \tag{50}$$

i.e.,

$$a^z_{x_c}\big(\frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N}\big) \geq \vartheta_{x_c,SLC} \tag{51}$$

$$0 \leq P^z_{x_c} a^z_{x_c} \leq P_{maxm}, \forall c \in C, \forall x_c \in X_c, \forall z \in Z. \tag{52}$$

$$\sum_{x_c \in X_c} a^z_{x_c} \leq 1, \forall z \in Z, c \in C \tag{53}$$

where $\vartheta_{x_c,SLC}$ is the $SINR$ constraint to satisfy the required QoS of a given slice. It is considered that only the UEs above this threshold can guarantee successful transmission.

5.2 Proposed Users Classification and Prediction method to Enhance the Scheduler

As it is seen above, the problem is of a mixed binary integer non-linear programming nature, i.e. it is very challenging to maximize the sum-rate while minimizing the level of acceptable interference for the scheduled UE to satisfy the expected QoS requirement. In this regard, we apply the solutions to minimize the interference between allocated UEs. Then to optimize the allocation matrix, we implement the machine learning schemes, hence allocating the resources based on classification and prediction. Finally, we perform power allocation to further minimize the unnecessary energy consumption of the transmitting UEs. The corresponding discussion of the machine learning and the enhanced scheduler setup are presented in the following sub-sections.

To begin with, the IoT UEs' channel parameters data are collected from a live 5G network; in our case in Haapsalu, Estonia. Then data processing is performed to eliminate the coverage holes where no actual communication parameters data was collected. Next, we run different machine learning algorithms to classify the UEs according to their corresponding channel parameters. We deploy several machine learning algorithms on the processed data sets to classify the users in different clusters to enhance the UE scheduling; several output parameters such as minimum classification errors, true response vs prediction response, etc. are used to judge the quality of the classification performance. For the above specific data set, we present the best-performing algorithms as compared to all possible lightweight ML algorithms. From the analysis, it can be noted that rational quadratic Gaussian Process Regression (GPR) performs better than the fine Gaussian Support Vector Machine (SVM) when compared to the perfect prediction on the collected samples. Then, based on the collected insights from both classification and regression algorithms, the intra-slice scheduler was designed to predict the periodicity and buffer size for the next scheduling frames.

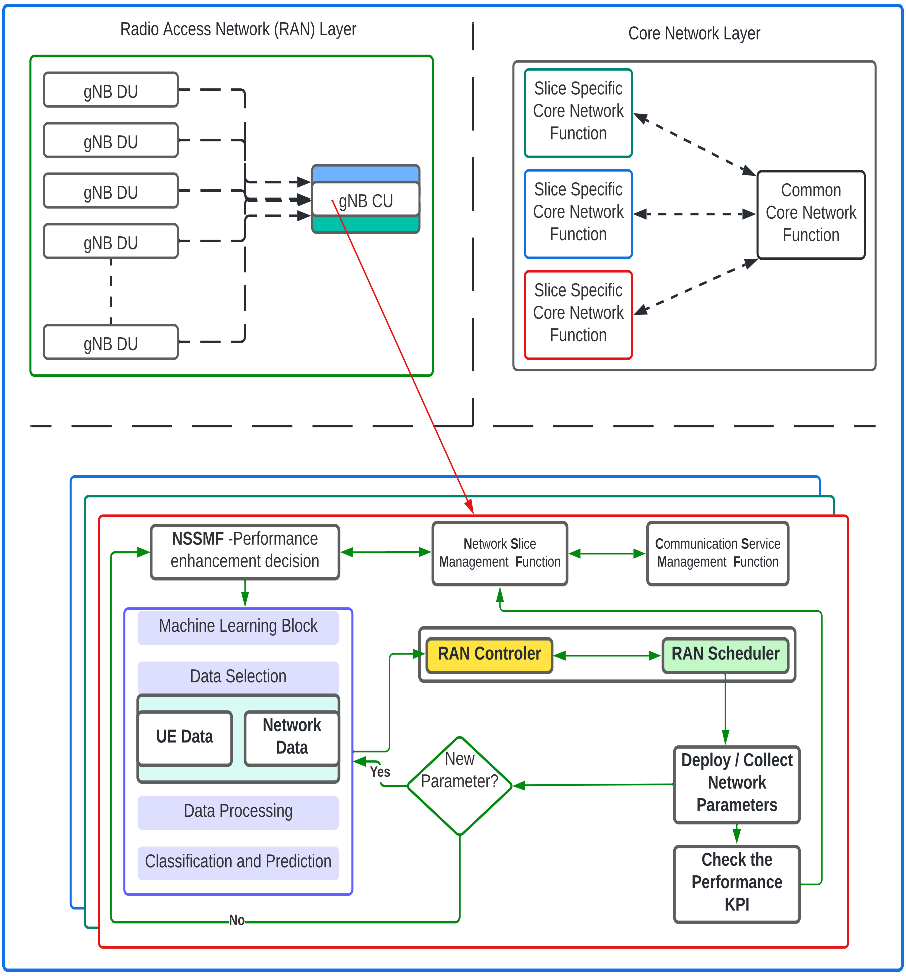

Figure 6: The proposed framework that utilizes the machine learning to perform the User clustering and Prediction to enhance the RAN slice scheduling for 5G and beyond networks

5.2.1 Intra-Slice Scheduling

Since our objective is to optimize the performance of massive IoT slices, the simulation is performed to map the collected real-time channel parameters to enhance the number of connected devices by predicting the UE periodicity and expected buffer size for the next frame. In this regard, unnecessary radio resources are released to support other slices that require higher bandwidth and/or transmission time slots. From the allocated and released resources, the performance of the network is analyzed to evaluate the effectiveness of the proposed algorithm in comparison to traditional scheduling algorithms that utilize fixed radio resources for a given expected QoS requirement. It should be noted that the inter-slice scheduling is out of scope for this work, but will be included in our future outlook.

The overall proposed framework is presented in Fig 6 and summarized below.

- The optimization parameter is selected in the Network Slice Sub-net Management Function (NSSMF) which acts as the brain of the slice, where slice function selection, configuration, and coordination are originated.
- The NSSMF decides to instantiate, scale, terminate, or move the slice based on the commands it receives from the Network Slice Management Function (NSMF), which receives the translation of related service requirements from the Communication Service Management Function (CSMF).
- Then the machine learning algorithm is applied to the collected data from either the 5G network. Based on the nature of the data, UE or network parameters cleaning is performed because some might be missing due to coverage holes or temporary UE disconnection from the network.
- The classification and prediction are performed on the data to give the knowledge on the behavior of UE hence the controller and the RAN scheduler coordinate and cooperate to allocate the optimal radio resources to the massive IoT slice to satisfy the required quality of service.
- When the allocated UEs are allocated to transmit on the network, either the UE or the network parameters are monitored and the performance is evaluated; in case of a completely new parameter or a significant change in current collected values, the changed parameter is injected back to be used in the machine learning to decide whether it can bring more enhancements to the slice performance in terms of UE energy consumption, throughput, number of connected devices etc.

5.2.2 Machine Learning Enabled Cooperative Scheduler

Our starting point is a cooperative scheduler designed to mitigate the impact of inter-cell interference caused by transmitting users from adjacent cells. The overall scheduling framework and its settings are presented in [29]. This scheduler comprised of sharing the scheduling tables between the base stations to mitigate the inter-cell interference by proactively allocating the radio resources to the users combinations that result in minimum possible interference to maximize the expected quality of service requirements. However, significant overhead was experienced in the X2 interface due to the need for information sharing between the base stations. In the current paper, we deploy the machine learning framework presented in Fig. 6 to perform the prediction of the base stations next transmission capabilities within a given slice. In doing so, the machine learning scheme is run in all the cooperating base stations for clustering not only the current transmissions but also the prediction of the corresponding upcoming transmissions of a given set of UEs in a given slice.

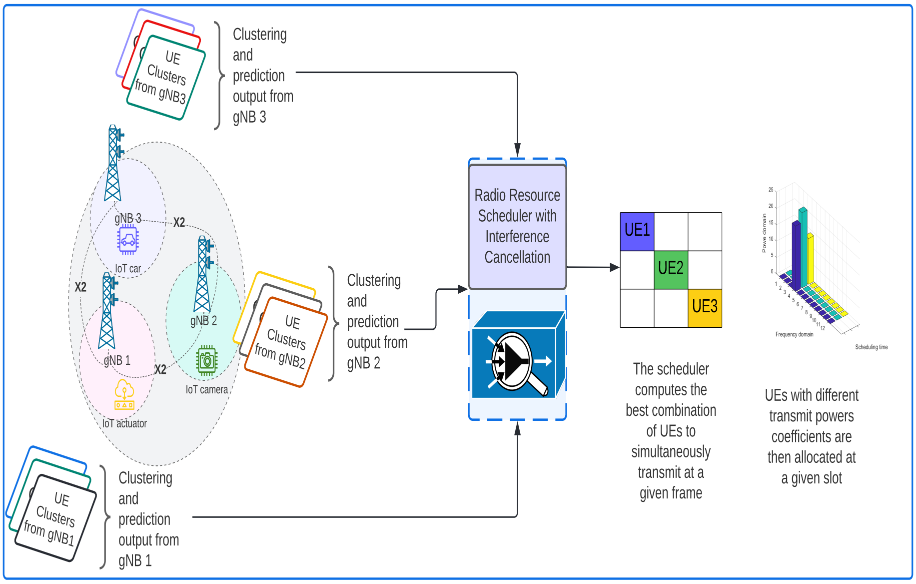

Figure 7: The proposed ML-enabled scheduler that utilizes the machine learning output (i.e. users classification and transmission prediction) from the cooperating base stations to allocate the radio resources to the UEs with minimum possible interference

For example, given the learned transmission pattern and the SINR distribution, a certain number of UEs can tolerate a 10 ms delay for a small packet size transmission that can occupy less bandwidth as compared to a set of UEs that require a 1 ms delay for the same packet size. In this regard, latency-sensitive UEs can occupy and release a given radio resource faster and let the slice be used by the latency insensitive UEs while the transmission periodicity is being monitored.

The simulation is performed in MATLAB to adapt different scheduling frameworks based on SINR, and the simulation parameters as used in [27]. We adapt the scheduling on the traditional and cooperative OMA schemes; however, for comparison purposes, the cooperative NOMA scheme is also considered to further compare the performance gains. In the current work, the machine learning framework is not adapted on the NOMA scheme due to its already existing computational complexity when performing the successive interference cancellation for the UE allocated at the same radio resources.

6 Conclusion

This Ph.D. thesis has answered the research question presented in chapter I as follows:

RQ1. How are the orthogonal multiple access techniques intended to accommodate the massive connectivity of IoT use cases impacted by inadequate interference management techniques in heterogeneous network architectures? By taking both intra-cell and inter-cell interference as constraints while formulating the objective functions, and designing the systems that utilize the real-time channel parameters during system modeling. In this thesis, we have studied the impact of massive connectivity in the interference-limited environment for the orthogonal multiple access techniques and proposed the cooperative scheduling framework that reduces the impact of both intra-cell and inter-cell interference to satisfy the expected users' quality of service. In this regard, the proposed framework improved throughput increased the number of connected devices per cell with reduced users' energy consumption.

RQ2. Is it possible to proactively perform the data-driven radio resource management techniques to enhance the cell performance, with the user's minimum-possible energy consumption while guaranteeing the expected quality of service requirements? Yes, it is possible. We implemented the data-driven scheduling, where the cooperating base stations share their corresponding scheduling tables before performing the final scheduling decision, this helped in avoiding unnecessary energy consumption to users whose channel conditions do not permit successful transmissions. While doing so, the available radio resources are given to users with better channel conditions hence reducing the number of repetitions that normally drain the devices' energy.

RQ3. Are the novel approaches such as non-orthogonal multiple access (NOMA) sufficient to reach the expected cell performance for mMTC applications, considering the impact of inter-cell interference from the neighboring cells, and the complexity of successive interference cancellation (SIC) at the receiver? The reduced receiver complexity to minimize the overall cost of mMTC devices hinders the performance of successive interference cancellations at the receiver. In this thesis, we centralized the scheduling by firstly grouping the users based on their channel conditions (i.e., bad, moderate, and good), then simultaneously allocating one user from each group at a given radio resource. Additionally, we implemented different power coefficients to users to help in decoding the respective powers hence retrieving the corresponding information.

RQ4. Since radio access network (RAN) slicing proves to be efficient for sharing and orchestration of networks, how is it possible to accommodate the massive connectivity of mMTC connections with diverse service requirements in a given slice while guaranteeing the expected performance as per service level agreement (SLA) templates? By using the machine learning algorithms, it is possible to cluster the users and predict their corresponding transmission patterns. In this regard, the SLA templates can be dynamically adapted by scheduling the users to satisfy the required quality of service of a given slice. It is necessary to choose the machine learning algorithms according to the nature of the collected data and adapt the scheduling framework accordingly. For example, if the data contains exhaustive data labels, it is better to use deep learning algorithms to find the relationship between different channel parameters and how these patterns affect the network performance. In our work, we used reinforcement learning because the amount of available data was limited. However, it can be noted that, depending on the nature of the data, possibility to interpret, speed of training, etc., different machine learning algorithms can be applied on either user or network data to enhance the RAN scheduling for massive connectivity support of massive IoT slice. It is observed that the ML-enabled scheduler outperforms the benchmark scheduling schemes significantly.

In general, the overall results in this thesis show that the proposed NOMA scheme is more spectrum efficient than OMA as it supports more than twice the number of connected devices for the same number of available resources. Furthermore, other network performance metrics such as throughput, user's energy consumption, and fairness are analyzed, discussed, and compared for both the OMA and NOMA schemes. Furthermore, the reduced impact of interference and the proposed power allocation techniques reduce the average energy consumption per device hence are more suitable for massive IoT deployments as it enhances the device's battery life longevity.

Our future outlook aims to implement advanced techniques such as intelligent reflecting surfaces (IRS), directional beamforming, and deep learning approaches to increase the cell capacity for the massive IoT devices in 5G and beyond networks. Furthermore, our future work includes inter-slice scheduling by proactive interference mitigation approaches to maximize the spectrum efficiency for massive IoT applications that require different QoS under heterogeneous slice configurations to be aligned with 6G flagship projects [76–78].

For example, the ubiquitous smart wireless connectivity is critical for future large-scale industrial tasks, services, assets, and devices. Very significantly improved connectivity needs to be unlocked through novel spectrum combinations and the fully autonomous management of the underlying network resources by applying online AI at multiple decision layers. Furthermore, novel data-driven techniques to support Grant Free Non-Orthogonal Multiple Access (GF-NOMA), multi-agent Deep Reinforcement Learning, End-to-End (E2E) Slicing, Integrated Access Backhaul (IAB), Industrial Virtual Assistant (IVA), Simultaneous Localization and Mapping (SLAM) need to be explored for beyond 5G networks.

References

[1] A. Müller and P. Frank. Cooperative interference prediction for enhanced link adaptation in the 3gpp lte uplink. pages 1–5, May 2010.

[2] Muhammad Waseem Akhtar, Syed Ali Hassan, Rizwan Ghaffar, Haejoon Jung, Sahil Garg, and M. Shamim Hossain. The shift to 6g communications: vision and requirements. *Human-centric Computing and Information Sciences*, 10(1):53, Dec 2020.

[3] Amitabha Ghosh, Andreas Maeder, Matthew Baker, and Devaki Chandramouli. 5g evolution: A view on 5g cellular technology beyond 3gpp release 15. *IEEE Access*, 7:127639–127651, 2019.

[4] Rashid Ali, Yousaf Bin Zikria, Ali Kashif Bashir, Sahil Garg, and Hyung Seok Kim. Urllc for 5g and beyond: Requirements, enabling incumbent technologies and network intelligence. *IEEE Access*, 9:67064–67095, 2021.

[5] Arjun Anand, Gustavo de Veciana, and Sanjay Shakkottai. Joint scheduling of urllc and embb traffic in 5g wireless networks. *IEEE/ACM Transactions on Networking*, 28(2):477–490, 2020.

[6] Collins Burton Mwakwata, Hassan Malik, Muhammad Mahtab Alam, Yannick Le Moullec, Sven Parand, and Shahid Mumtaz. Narrowband internet of things (nb-iot): From physical (phy) and media access control (mac) layers perspectives. *Sensors*, 19(11), 2019.

[7] Shiva Raj Pokhrel, Jie Ding, Jihong Park, Ok-Sun Park, and Jinho Choi. Towards enabling critical mmtc: A review of urllc within mmtc. *IEEE Access*, 8:131796–131813, 2020.

[8] H. Malik, H. Pervaiz, M. Mahtab Alam, Y. Le Moullec, A. Kuusik, and M. Ali Imran. Radio resource management scheme in nb-iot systems. *IEEE Access*, 2018.

[9] 3GPP. 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) (Release 13) . Technical Specification (TS) 36.331, 3rd Generation Partnership Project (3GPP), 08 2016. V13.1.0 (2015-11).

[10] E. M. Migabo, K. D. Djouani, and A. M. Kurien. The narrowband internet of things (nb-iot) resources management performance state of art, challenges, and opportunities. *IEEE Access*, 8:97658–97675, 2020.

[11] S. Böcker, C. Arendt, P. Jörke, and C. Wietfeld. Lpwan in the context of 5g: Capability of lorawan to contribute to mmtc. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 737–742, 2019.

[12] D. Loghin, S. Cai, G. Chen, T. Dinh, F. Fan, Q. Lin, J. Ng, B. Ooi, X. Sun, Q. Ta, W. Wang, X. Xiao, Y. Yang, M. Zhang, and Z. Zhang. The disruptions of 5g on data-driven technologies and applications. *IEEE Transactions on Knowledge and Data Engineering*, 32(06):1179–1198, jun 2020.

[13] Eshita Rastogi, Navrati Saxena, Abhishek Roy, and Dong Ryeol Shin. Narrowband internet of things: A comprehensive study. *Computer Networks*, 173:107209, 2020.

[14] Shunliang Zhang. An overview of network slicing for 5g. *IEEE Wireless Communications*, 26(3):111–117, 2019.

[15] Qingqing Wu, Shuowen Zhang, Beixiong Zheng, Changsheng You, and Rui Zhang. Intelligent reflecting surface aided wireless communications: A tutorial. *CoRR*, abs/2007.02759, 2020.

[16] ITU. Focus group on machine learning for future networks including 5g.

[17] Collins Burton Mwakwata, Muhammad Mahtab Alam, Yannick Le Moullec, Hassan Malik, and Sven Parand. Cooperative interference avoidance scheduler for radio resource management in nb-iot systems. In *2020 European Conference on Networks and Communications (EuCNC)*, pages 154–159, 2020.

[18] Collins Burton Mwakwata, Osama Elgarhy, Yannick Le Moullec, Muhammad Mahtab Alam, Sven Pärand, and Ivar Annus. Inter-cell interference reduction scheme for uplink transmission in nb-iot systems. In *2021 International Wireless Communications and Mobile Computing (IWCMC)*, pages 400–405, 2021.

[19] Collins Burton Mwakwata, Osama Elgarhy, Muhammad Mahtab Alam, Yannick Le Moullec, Sven Pärand, Konstantinos Trichias, and Kostas Ramantas. Cooperative scheduler to enhance massive connectivity in 5g and beyond by minimizing interference in oma and noma. *IEEE Systems Journal*, pages 1–12, 2021.

[20] Jack L. Burbank, Julia Andrusenko, Jared S. Everett, and William T.M. Kasch. *Second-Generation (2G) Cellular Communications*, pages 250–365. 2013.

[21] E. Ezhilarasan and M. Dinakaran. A review on mobile technologies: 3g, 4g and 5g. In *2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM)*, pages 369–373, 2017.

[22] Amit Kr. Jain, Rupesh Acharya, Saroj Jakhar, and Tarun Mishra. Fifth generation (5g) wireless technology "revolution in telecommunication". In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pages 1867–1872, 2018.

[23] Guang yong SUN, Fa jun YAN, and Wen bo WANG. Inter-cell interference cancellation algorithm for reverse link tdd-cdma systems. *The Journal of China Universities of Posts and Telecommunications*, 13(1):84–92, 2006.

[24] Fabio Fernandes, Alexei Ashikhmin, and Thomas L. Marzetta. Inter-cell interference in noncooperative tdd large scale antenna systems. *IEEE Journal on Selected Areas in Communications*, 31(2):192–201, 2013.

[25] Yiqing Zhou, Ling Liu, Hongyan Du, Lin Tian, Xiaodong Wang, and Jinglin Shi. An overview on intercell interference management in mobile cellular networks: From 2g to 5g. In *2014 IEEE International Conference on Communication Systems*, pages 217–221, 2014.

[26] A. Rico-Alvarino, M. Vajapeyam, H. Xu, X. Wang, Y. Blankenship, J. Bergman, T. Tirronen, and E. Yavuz. An overview of 3gpp enhancements on machine to machine communications. *IEEE Communications Magazine*, June 2016.

[27] J. Gozalvez. New 3gpp standard for iot [mobile radio]. *IEEE Vehicular Technology Magazine*, March 2016.

[28] A. D. Zayas and P. Merino. The 3gpp nb-iot system architecture for the internet of things. In *IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017.

[29] ERICSSON. Scope of Release 16-NB-IoT, RP-181187. `http://www.3gpp.org/ftp/TSG\_RAN/TSG\_RAN/TSGR\_80/Report/`, Available online 2018.

[30] Huawei. On NB-IoT Evolution in Rel-16, RP-180877, . `http://www.3gpp.org/ftp/TSG\_RAN/TSG\_RAN/TSGR\_80/Report/`, Available online 2018.

[31] 3GPP. Revision of WI on Enhancements of NB-IoT, RP-161901, . `https://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view\&contributionId=730352`, Available online 2018.

[32] 3GPP. TS 36.300 V13.8.0 (2017-06), . `https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2430`, Available online 2018.

[33] A. Adhikary, X. Lin, and Y. . E. Wang. Performance evaluation of nb-iot coverage. In *IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sep. 2016.

[34] T. Xu and I. Darwazeh. Non-orthogonal narrowband internet of things: A design for saving bandwidth and doubling the number of connected devices. *IEEE Internet of Things Journal*, June 2018.

[35] Y. D. Beyene, R. Jantti, O. Tirkkonen, K. Ruttik, S. Iraji, A. Larmo, T. Tirronen, and a. J. Torsner. Nb-iot technology overview and experience from cloud-ran implementation. *IEEE Wireless Communications*, June 2017.

[36] R. Ratasuk, N. Mangalvedhe, J. Kaikkonen, and M. Robert. Data channel design and performance for lte narrowband iot. In *IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sep. 2016.

[37] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J. Koskinen. Overview of narrowband iot in lte rel-13. In *IEEE Conference on Standards for Communications and Networking (CSCN)*, Oct 2016.

[38] 3GPP. specs. `http://www.3gpp.org/ftp/Specs/archive/45\_series/45.820/45820-d10.zip`, Available online 2019.

[39] A. Hoglund, X. Lin, O. Liberg, A. Behravan, E. A. Yavuz, M. Van Der Zee, Y. Sui, T. Tirronen, A. Ratilainen, and D. Eriksson. Overview of 3gpp release 14 enhanced nb-iot. *IEEE Network*, November 2017.

[40] R. Ratasuk, N. Mangalvedhe, Z. Xiong, M. Robert, and D. Bhatoolaul. Enhancements of narrowband iot in 3gpp rel-14 and rel-15. In *IEEE Conference on Standards for Communications and Networking (CSCN)*, Sep. 2017.

[41] 3GPP. TS 36.213 , . `https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2427`, Available online 2018.

[42] 3GPP. 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) (Release 13) . Technical Specification (TS) 36.331, 3rd Generation Partnership Project (3GPP), 08 2016. V13.1.0 (2015-11).

[43] 3GPP. Release 17.

[44] G. J. González, F. H. Gregorio, and J. Cousseau. Interference analysis in the lte and nb-iot uplink multiple access with rf impairments. In *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pages 1–4, 2018.

[45] L. P. Qian, Z. Zhu, N. Yu, and Y. Wu. Joint minimization of transmission energy and computation energy for mec-aware noma nb-iot networks. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7, Dec 2019.

[46] L. Zhang, A. Ijaz, P. Xiao, and R. Tafazolli. Channel equalization and interference analysis for uplink narrowband internet of things (nb-iot). *IEEE Communications Letters*, 21(10):2206–2209, 2017.

[47] Z. Ding, Z. Yang, P. Fan, and H. V. Poor. On the performance of non-orthogonal multiple access in 5g systems with randomly deployed users. *IEEE Signal Processing Letters*, 21(12):1501–1505, 2014.

[48] Shashwat Mishra, Lou Salaün, and Chung Shue Chen. Maximizing connection density in nb-iot networks with noma. In *IEEE VTC 2020, Antwerp Belgium*, pages 1–6, 05 2020.

[49] T. N. Do, D. B. da Costa, T. Q. Duong, and B. An. Improving the performance of cell-edge users in noma systems using cooperative relaying. *IEEE Transactions on Communications*, 66(5):1883–1901, 2018.

[50] A. Nazari, M. R. Javan, and S. S. Hosseini. Resource allocation in power domain noma-based cooperative multicell networks. *IET Communications*, 14(7):1162–1168, 2020.

[51] A. E. Mostafa, Y. Zhou, and V. W. S. Wong. Connection density maximization of narrowband iot systems with noma. *IEEE Transactions on Wireless Communications*, 18(10):4708–4722, Oct 2019.

[52] Petar Popovski, Kasper Fløe Trillingsgaard, Osvaldo Simeone, and Giuseppe Durisi. 5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view. *IEEE Access*, 6:55765–55779, 2018.

[53] Lian Cao, Rongpeng Li, Jon Crowcroft, Zhifeng Zhao, and Honggang Zhang. Intelligent slicing of radio resource control layer for cellular iot: Design and implementation, 2020.

[54] Tengteng Ma, Yong Zhang, Fanggang Wang, Dong Wang, and Da Guo. Slicing resource allocation for embb and urllc in 5g ran. *Wireless Communications and Mobile Computing*, 2020:6290375, Jan 2020.

[55] R. Ferrús, O. Sallent, J. Pérez-Romero, and R. Agustí. On the automation of ran slicing provisioning: solution framework and applicability examples. *EURASIP Journal on Wireless Communications and Networking*, 2019(1):167, Jun 2019.

[56] Peng Yang, Xing Xi, Tony Q. S. Quek, Jingxuan Chen, Xianbin Cao, and Dapeng Wu. Ran slicing for massive iot and bursty urllc service multiplexing: Analysis and optimization. *IEEE Internet of Things Journal*, 8(18):14258–14275, 2021.

[57] Pablo Caballero, Albert Banchs, Gustavo de Veciana, and Xavier Costa-Pérez. Multi-tenant radio access network slicing: Statistical multiplexing of spatial loads. *IEEE/ACM Transactions on Networking*, 25(5):3044–3058, 2017.

[58] Behnam Khodapanah, Ahmad Awada, Ingo Viering, David Oehmann, Meryem Simsek, and Gerhard P. Fettweis. Fulfillment of service level agreements via slice-aware radio resource management in 5g networks. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–6, 2018.

[59] Bin Han and Hans D. Schotten. Machine learning for network slicing resource management: A comprehensive survey, 2020.

[60] Muhammad Rehan Raza, Carlos Natalino, Peter Öhlen, Lena Wosinska, and Paolo Monti. Reinforcement learning for slicing in a 5g flexible ran. *J. Lightwave Technol.*, 37(20):5161–5169, Oct 2019.

[61] Yi-Jing Liu, Gang Feng, Yao Sun, Shuang Qin, and Ying-Chang Liang. Device association for ran slicing based on hybrid federated deep reinforcement learning. *IEEE Transactions on Vehicular Technology*, 69(12):15731–15745, 2020.

[62] Zheng Chu, Pei Xiao, Mohammad Shojafar, De Mi, Juquan Mao, and Wanming Hao. Intelligent reflecting surface assisted mobile edge computing for internet of things. *IEEE Wireless Communications Letters*, 10(3):619–623, 2021.

[63] F. AL-Tam, A. Mazayev, N. Correia, and J. Rodriguez. Radio resource scheduling with deep pointer networks and reinforcement learning. In *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pages 1–6, 2020.

[64] Alexandre Ragaleux, Sébastien Baey, and Cédric Gueguen. Adaptive and generic scheduling scheme for lte/lte-a mobile networks. *Wireless Networks*, 22(8):2753–2771, Nov 2016.

[65] Ramprasad Subramanian, Pantha Ghosal, Shouman Barua, Shiqi Xing, Sl Cong, H. Al Kim, and Kumbesan Sandy Sandrasegaran. Survey of lte downlink schedulers algorithms in open access simulation tools ns-3 and lte-sim. 2015.

[66] A. I. A. Jabbar and Fawaz Y. Abdullah. Long Term Evolution (LTE) Scheduling Algorithms in Wireless Sensor Networks (WSN). *International Journal of Computer Applications*, 121(10):12–16, July 2015.

[67] Nsiri Bechir, Mallouki Nasreddine, Ammar Mahmoud, Hakimi Walid, and Mhatli Sofien. Novel scheduling algorithm for 3gpp downlink lte cellular network. *Procedia Computer Science*, 40:116 – 122, 2014. Fourth International Conference on Selected Topics in Mobile and Wireless Networking.

[68] Olof Liberg, Mårten Sundberg, Y.-P. Eric Wang, Johan Bergman, Joachim Sachs, and Gustav Wikström. Chapter 16 - choice of iot technology. In Olof Liberg, Mårten Sundberg, Y.-P. Eric Wang, Johan Bergman, Joachim Sachs, and Gustav Wikström, editors, *Cellular Internet of Things (Second Edition)*, pages 687–707. Academic Press, second edition edition, 2020.

[69] Adlen Ksentini and Navid Nikaein. Toward enforcing network slicing on ran: Flexibility and resources abstraction. *IEEE Communications Magazine*, 55(6):102–108, 2017.

[70] S. A. AlQahtani and M. Alhassany. Comparing different lte scheduling schemes. In *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 264–269, 2013.

[71] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana. Coordinated multipoint transmission and reception in lte-advanced: deployment scenarios and operational challenges. *IEEE Communications Magazine*, 50(2):148–155, 2012.

[72] Saddam Alraih, Ibraheem Shayea, Mehran Behjati, Rosdiadee Nordin, Nor Fadzilah Abdullah, Asma' Abu-Samah, and Dalia Nandi. Revolution or evolution? technical requirements and considerations towards 6g mobile communications. *Sensors*, 22(3), 2022.

[73] Matteo Pagin, Tommaso Zugno, Marco Giordani, Louis-Adrien Dufrene, Quentin Lampin, and Michele Zorzi. Can 5g nr-light operate at millimeter waves? design guidelines for mid-market iot use cases, 2021.

[74] Xenofon Foukas, Mahesh K. Marina, and Kimon Kontovasilis. Orion: Ran slicing for a flexible and cost-effective multi-service mobile network architecture. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, MobiCom '17, page 127–140, New York, NY, USA, 2017. Association for Computing Machinery.

[75] Salah Eddine Elayoubi, Sana Ben Jemaa, Zwi Altman, and Ana Galindo-Serrano. 5g ran slicing for verticals: Enablers and challenges. *IEEE Communications Magazine*, 57(1):28–34, 2019.

[76] Mikko A. Uusitalo, Patrik Rugeland, Mauro Renato Boldi, Emilio Calvanese Strinati, Panagiotis Demestichas, Mårten Ericson, Gerhard P. Fettweis, Miltiadis C. Filippou, Azeddine Gati, Marie-Helene Hamon, Marco Hoffmann, Matti Latva-Aho, Aarno Pärssinen, Björn Richerzhagen, Hans Schotten, Tommy Svensson, Gustav Wikström, Henk Wymeersch, Volker Ziegler, and Yaning Zou. 6g vision, value, use cases and technologies from european 6g flagship project hexa-x. *IEEE Access*, 9:160004–160020, 2021.

[77] Mikko A. Uusitalo, Mårten Ericson, Björn Richerzhagen, Elif Ustundag Soykan, Patrik Rugeland, Gerhard Fettweis, Dario Sabella, Gustav Wikström, Mauro Boldi, Marie-Helene Hamon, Hans D. Schotten, Volker Ziegler, Emilio Calvanese Strinati, Matti Latva-aho, Pablo Serrano, Yaning Zou, Gino Carrozzo, Josep Martrat, Giovanni Stea, Panagiotis Demestichas, Aarno Pärssinen, and Tommy Svensson. Hexa-x the european 6g flagship project. In *2021 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit)*, pages 580–585, 2021.

[78] Wei Jiang and Hans Dieter Schotten. The kick-off of 6g research worldwide: An overview, Sep 2021.

Acknowledgements

Firstly, I would like to thank Tallinn University of Technology (Taltech), Thomas Johann Seebeck Department of Electronics for giving me the opportunity to pursue my Ph.D. studies. Secondly, I would like to thank my supervisors/advisors Prof. Muhammad Mahtab Alam, and Prof Yannick Le Moulec for supporting me in all aspects to make my Ph.D, successful. Thirdly I would like to thank university staff especially, Dr. Osama Elgarhy, Priit Roosipu, Frederick Rang, Eva Keerov, and Dr. Sven Pärand (Telia Eesti AS) for their support. Special thanks to my family for the love and support.

Lastly, I would like to thank the following funds that financed my Ph.D. studies.

- ICT Doctoral School at Tallinn University of Technology.
- Telia Eesti AS Scholarship (SA TTÜ ARENGUFOND)
- PRG667 : Platform for Decentralized Real-Time Management of Rainwater Systems in Climate-Resistant Smart Cities (DEPART)
- IT Academy Scholarship at Tallinn University of Technology.

Abstract

Scheduling in Radio Resource Management for Massive Machine-Type Communications

The increase in the number of connected devices to support different service verticals for legacy, 5G, and beyond wireless networks with limited radio resources necessitates the implementation of novel radio resource management techniques. This is because, as the number of users increases the impact of interference also increases, hence guaranteeing the expected quality of service under traditional approaches becomes impossible.

For example, the legacy and current 5G deployments utilize the orthogonal multiple access techniques that assign orthogonality in terms of frequency and time, in this regard, reaching the expected number of connected devices, especially for the massive machine-type communications became very challenging. Similarly, even in beyond the 5G networks, the non-orthogonal multiple access techniques fail to reach the expected quality of service requirements. The main reasons for limited cell performance are Limited Energy, multipath propagation, spectrum limitations, interference-limited systems, user mobility, security, privacy, support for multiple access techniques (i.e., duplexing), etc.

Furthermore, the current conventional approaches to resource management, especially for interference mitigation are based on optimization theory techniques, therefore, it is necessary to implement the suitable mathematical models for a particular problem, i.e. tractable to characterize the proposed solutions based on information theory. However, even when this is done, the optimal solution will inevitably depend on the system parameters, i.e., the receiver's sensitivity, the users' location, the connection density, slow/fast fading channel variations, etc. In this regard, when any of these parameters change, the optimization problem becomes obsolete, hence causing a significant complexity on the real-time realizations, especially in heterogeneous wireless network architectures.

Additionally, with dense deployments of wireless networks to support the massive machine-type communications, it is necessary to utilize the amount of available network data to implement the novel approaches that implement artificial intelligence (AI) to enable the data-driven real-time radio resource management techniques especially to counteract the impact of interference for the dense deployment networks. One of the most important mechanisms to manage the stringent radio resource management is through scheduling, therefore, it is necessary to design efficient radio resource schedulers that will guarantee the throughput performance and acceptable fairness by the service provider. It should be noted that scheduling algorithms, which distribute the available resources to the competing users that require to connect to the network have proved to be the key contributors to the quality of service provision in wireless networks, especially with the limited available radio resources to support a particular application. A large number of traffic scheduling algorithms have been proposed in the literature, however, the upcoming massive connectivity of internet of things with different service requirements and the scarcity of resources, have rendered the adaptation of these schedulers very challenging.

In this regard, this Ph.D. thesis focuses on advances of radio resource management, specifically, interference management techniques by employing proactive scheduling to minimize interference hence maximizing the overall cell performance. For example, we design the novel interference management scheduler that utilizes a cooperative strategy in a multi-cell network. Then, we propose the novel inter-cell and intra-call interference management scheme for the orthogonal multiple access techniques, additionally, by con-

sidering the limited computation complexity of connected users, the interference-aware scheme is proposed for the non-orthogonal multiple access with a power allocation mechanism to reduce the devices' energy consumption. Finally, we propose the radio access network slice scheduler that utilizes machine learning algorithms to classify the users according to their channel conditions and predict their future transmission patterns that enhance the scheduling performance of a given slice.

In general, the proposed scheduler enhances the cell performance accordingly. For the proposed OMA and NOMA schedulers, the results show enhancements up to 58%, 75%, and 100% in terms of user's data rates, energy consumption, and connection density, respectively.

Kokkuvõte
Raadioressursside Planeerimine Massiivse Masin-masin Tüüpi Kommunikatsiooni Korral

Aasta-aastalt suureneb sideseadmete hulk, mis võimaldavad nii viienda põlvkonna mobiilside (5G), kui ka sellele eelnevate ning järgnevate traadita side standarditel baseeruvate teenuste toimimist. Piiratud raadioressursside hulk koos aina suureneva sideseadmete arvuga tekitab vajaduse uudsete raadioressursi haldamise meetodite järgi. Kasvav raadioseadmete arv toob tahtmatult endaga kaasa ka suureneva häirete hulga, muutes traditsiooniliste meetoditega teenuse kvaliteedi tagamise võimatuks.

Kaasaaegsetes 5G- kui ka sellele eelnevates sidesüsteemide generatsioonides kasutusel olevad ortogonaalsed raadioressursi jaotamise tehnoloogiad (orgthogonal multiple access, OMA) kasutavad ortogonaalsuse printsiipi nii sageduslikus- kui ka aegruumis, et teenindada ära vajalikul hulgal kliente. Vaatamata antud meetodite kasutusele muudab suurenev sideseadmete, eriti massiivse masin-masin tüüpi suhtluse (massive machine-type communications, mMTC) seadmete arv traadita võrgu teenindamise aina keerukamaks. Sarnaselt eelnevale, ei suuda ka 5G'le järgnevates generatsioonides kasutatavad mitte-ortogonaalsed ühispöörduse tehnoloogiad (non-orthogonal multiple access, NOMA) tagada oodatud teenusekvaliteedi taset. Piiratud jõudlus tuleneb mitmekiirelisest levist, piiratud energiahulgast ja -raadiospektrist, vastastikuste häirete mõjust seadmetele, kasutajate liikuvusest, turvalisuse- ja privaatsusnõuetest, ühispöörduse tehnoloogiate toetusest jne.

Traditsionaalsed ressursihalduse meetodid põhinevad optimeerimisteoorial, nende kasutamisel rakendatakse sobilikke matemaatilisi mudeleid vastavalt konkreetse probleemi iseloomule. Sellisel juhul aga sõltub optimaalne lahend mitmetest erinevatest süsteemi parameetritest nagu näiteks vastuvõtja tundlikkus, kasutaja asukoht, sideseansside sagedus, sidekanali parameetrid jms. Mõne eeltoodud parameetri muutuse mõjul muutub ka antud optimeerimisprobleemi lahend mittekasutatavaks, see omakorda tõstab keerukust reaalajasüsteemides rakendamisel, esmajoones heterogeensetes traadita sidevõrkudes. Laiaulatuslikku masin-masin suhtlust võimaldavates traadita andmesidevõrkudes on seega vaja rakendada kogu sidevõrgu parameetrite kohta olemasolevaid andmeid, et arendada ja kasutada uudseid tehisintellektil (artificial intelligence, AI) baseeruvaid ressursihalduse meetodeid. Eelmainitud meetodid võimaldavad rakendada reaalajalist andmepõhist raadioressursside haldust eesmärgiga vähendada häirete mõju suure seadmetihedusega raadioside võrkudes.

Üks olulisimaid raadioressursside haldamise mehhanisme on sideseansside planeerimine/jaotamine (scheduling). Seetõttu on vajaliku läbilaskevõime ning sideseansside õiglase jaotamise saavutamiseks vaja luua efektiivseid raadioressursi plaaneerijaid (scheduler).

Ajakava planeerimise algoritmid on osutunud peamiseks osaks traadita sidevõrkude teenusekvaliteedi tagamisel. Esmajoones on need algoritmid olulised olukordades kus raadioressursid on tugevalt piiratud, tagades nõutud teenuse kvaliteedi läbi kasutajate raadioressursside ümber jaotamise. Kuigi erialakirjanduses on välja pakutud mitmeid erinevaid võrguliikluse planeerimise algoritme, on nende rakendamine raskendatud asjade interneti (internet of things, IoT) tulekuga, seda esmajoones suurenenud ressursside nappuse ning varasemast erinevatele, teenustele kehtestatud nõuete tõttu. Käesolev doktoritöö keskendub raadioressursside haldamise meetodite arendamisele. Täpsemalt käsitletakse häirete halduse meetoditele milledes rakendatakse ennetavat toimingute ajastust, et maksimeerida sidevõrgu üleüldist jõudlust.

Näiteks loodi uudne häirete haldamise planeerija, mis kasutab ära kärgvõrkude koostööstrateegiat. Järgnevalt pakuti välja uudne kärjesisese ning kärgede vahelise häirehalduse süsteem ortogonaalsete -ühispöördusmeetodite jaoks. Võttes arvesse ühenduses olevate seadmete piiratud arvutusvõimsust, pakkuti välja häireteadlik süsteem mitte -ortogonaalse ühispöörduse ja võimsuse jaotamise korral, et vähendada antud seadmete energiatarvet. Viimasena pakuti välja raadiojuurdepääsuvõrgu (radio access network, RAN) võrguviil plaanur, milles on rakendatud masinõppe algoritme ja mis klassifitseerib sidevõrgu kasutajaid vastavalt sidekanali tingimustele ning prognoosib nende tulevasi transmissiooniaegu ning -mustreid. Antud lahendus võimaldab suurendada konkreetse juurdepääsuvõrgu võrguviilu toimingute ajastamise sooritust.

Tulemustest saab järeldada, et väljapakutud plaanur suurendab sidevõrgu jõudlust: näiteks väljapakutud OMA ja NOMA plaanurid suudavad pakkuda kuni 58%, 75% ja 100% paranemist. Viimast siis vastavalt seadme edastuskiiruses, energiatarbes ning loodavate ühenduste ruumilises tiheduses.

Appendix 1

C.B. Mwakwata, H. Malik, M.M. Alam, Y. Le Moullec, S. Parand, S. Mumtaz. Narrowband Internet of Things (NB-IoT): From physical (PHY) and media access control (MAC) layers perspectives. MDPI Sensors, 2019.

sensors

MDPI

Article

Narrowband Internet of Things (NB-IoT): From Physical (PHY) and Media Access Control (MAC) Layers Perspectives

Collins Burton Mwakwata [1,*], Hassan Malik [1], Muhammad Mahtab Alam [1], Yannick Le Moullec [1], Sven Parand [2] and Shahid Mumtaz [3]

1 Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology (TalTech), Ehitajate tee-5, 19086 Tallinn, Estonia; hassan.malik@taltech.ee (H.M.); muhammad.alam@taltech.ee (M.M.A.); yannick.lemoullec@taltech.ee (Y.L.M.)
2 Telia Estonia Ltd., 10616 Tallinn, Estonia; sven.parand@telia.ee
3 Instituto de Telecomunicações, 1049-001 Aveiro, Portugal; smumtaz@av.it.pt
* Correspondence: collins.burton@taltech.ee; Tel.: +372-58-663569

Received: 27 April 2019; Accepted: 5 June 2019; Published: 8 June 2019

check for updates

Abstract: Narrowband internet of things (NB-IoT) is a recent cellular radio access technology based on Long-Term Evolution (LTE) introduced by Third-Generation Partnership Project (3GPP) for Low-Power Wide-Area Networks (LPWAN). The main aim of NB-IoT is to support massive machine-type communication (mMTC) and enable low-power, low-cost, and low-data-rate communication. NB-IoT is based on LTE design with some changes to meet the mMTC requirements. For example, in the physical (PHY) layer only single-antenna and low-order modulations are supported, and in the Medium Access Control (MAC) layers only one physical resource block is allocated for resource scheduling. The aim of this survey is to provide a comprehensive overview of the design changes brought in the NB-IoT standardization along with the detailed research developments from the perspectives of Physical and MAC layers. The survey also includes an overview of Evolved Packet Core (EPC) changes to support the Service Capability Exposure Function (SCEF) to manage both IP and non-IP data packets through Control Plane (CP) and User Plane (UP), the possible deployment scenarios of NB-IoT in future Heterogeneous Wireless Networks (HetNet). Finally, existing and emerging research challenges in this direction are presented to motivate future research activities.

Keywords: narrowband; IoT; PHY; NB-IoT; MAC; deployment; survey; mMTC; 5G

---

1. Introduction

According to Information Handling Services (IHS) technology forecast, the Internet of Things (IoT) market is expected to grow to billions of devices by 2020 [1]. Massive connections are expected to respond to different IoT use cases such as smart city, smart wearables, smart home, etc. [2]. For these applications, latency-insensitive devices can be positioned in hard-to-reach areas and do not require high throughput or frequent reporting. Therefore, to cope with such tremendous IoT trends, the Third-Generation Partnership Project (3GPP) introduced the Narrowband Internet of Things (NB-IoT) standard as a communication technology enabler. NB-IoT is categorized as one of the licensed Low-Power Wide-Area Networks (LPWAN) cellular technologies based on Long-Term Evolution (LTE) with long range and low cost. In the LPWAN category, there exist other licensed technologies, i.e., Long-Term Evolution Category M1 (LTE-M), and unlicensed technologies, i.e., Long Range (LoRa), SigFox, Ingenu, etc. [3–7], but they are not the focus of the current work since they are not based on cellular technology.

The term Narrowband refers to NB-IoT's bandwidth of maximum 200 kHz thanks to which it can coexist either in the Global System for Mobile Communications (GSM) spectrum or by occupying one of the legacy LTE Physical Resource Blocks (PRBs) as in-band or as guard-band. Since it coexists in the LTE spectrum, NB-IoT follows the legacy LTE numerologies as it uses Orthogonal Frequency Division Multiplexing (OFDM) and Single-Carrier Frequency Division Multiple Access (SC-FDMA) in the downlink and uplink transmission schemes, respectively. Some modifications in the physical (PHY) and medium access control (MAC) layers are implemented to support the long-range massive machine-type (mMTC) connections with low power, low data rates, low complexity, and hence low cost. However, despite its low complexity, this new radio access technology (RAT) delivers better performance in terms of the supported number of devices, and coverage enhancements for latency-insensitive applications with maximum coupling loss (MCL) of about 20 dB higher than LTE (i.e., 164 dB) [5–11].

With flexible deployment as well as the possibility to implement over-the-air (OTA) firmware upgrades, many telecommunication operators across the globe (as shown in Figure 1) deployed NB-IoT to test its practical feasibility on diverse use cases with real-life trials such as connected sheep in Norway [12], smart metering and tracking in Brazil [13], NB-IoT at sea in Norway [14], smart city in Las Vegas, USA [15], etc. The trials are enabled by different NB-IoT software and hardware solutions from different chip or module vendors such as Skyworks [16], Media tek [17], Neul (Huawei) [18], Quectel [19], Nordic Semiconductors [20], Intel [21], Sequans [22], Qualcomm [23], Siera wireless [24], Samsung [25], Altair [26], U-Blox [27], and so on.

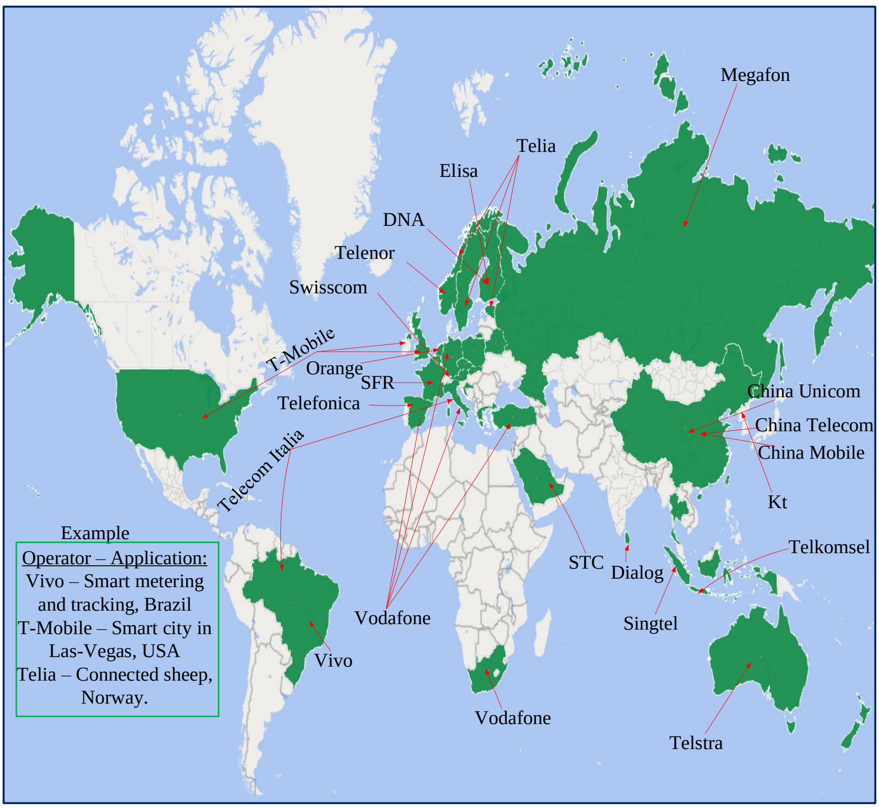

Figure 1. The geographical representation of countries with the ongoing NB-IoT real-life deployments for diverse use cases (May 2019).

The availability of such commercial off-the-shelf solutions speeds up the adoption of NB-IoT. For this reason, numerous studies addressing segmented enhancement criteria including survey articles emerged to analyze NB-IoT performance and implementation. Table 1 presents, in a nutshell, the main differences and similarities between this survey and the other existing ones by displaying the key focus features.

Table 1. Summarized comparison of this survey's contribution with respect to the existing surveys.

Survey	**The Third Generation Partnership Project**				**Layers**		**Deployment Strategies**
[Ref]	**Rel 13**	**Rel 14**	**Rel 15**	**Rel 16**	**Physical**	**Media Access Control**	
[28] 2017	✓						
[29] 2017	✓	✓					
[30] 2017		✓					
[31] 2017	✓						
[32] 2018	✓						
[33] 2019	✓	✓					
This survey	✓	✓	✓	✓	✓	✓	✓

For example, in [32], the authors surveyed the development path of MTC and elaborated the NB-IoT evolution in Release 13. Similarly, in [28], the authors discussed the Release 13 features and compared its performance with respect to other communication technologies such as LTE-M, SigFox, Lora and Wireless-Fidelity (WiFi), etc. In [29,30], the authors gave an overview of NB-IoT Release 14; however, in [30], the authors elaborated more on the expectations for NB-IoT Release 15 agenda. In [31], the authors presented a survey on the NB-IoT downlink scheduling issues by highlighting the associated scheduling process in terms of offset index selection. In [33], the authors surveyed the uplink and downlink performance evaluation of NB-IoT systems by analyzing the main causes of latency, trade-off between throughput and free resources, channel occupancy etc. with respect to Release 13 and Release 14 updates.

In contrast to the above surveys, this paper presents:

- A comprehensive survey of NB-IoT, from Release 13 to the ongoing Release 16 prospects.
- An all-inclusive overview of the state of the art of PHY and MAC layers by addressing the key improvement concerns in terms of challenges and the corresponding potential solutions.
- The possible NB-IoT deployment strategies for synchronous and asynchronous network structures in HetNet scenarios to foster the NB-IoT coexistence with legacy technologies as well as with the fifth generation (5G) networks.
- Discussion on the open research challenges to motivate future research directions.

To the best of the authors' knowledge, this is the first survey that covers broadly these above-mentioned contributions and hence will facilitate the reader's knowledge related to NB-IoT from standardization, ongoing research, and its practical implementation.

The rest of this paper is organized as follows: Section 2 discusses NB-IoT standards by elaborating the key design changes and the related ongoing enhancements. Section 3 presents the state of the art of NB-IoT protocol stack by detailing the PHY layer and MAC layer features. Section 4 discusses the open research questions and their potential solutions, and the conclusion is drawn in Section 5.

2. Narrowband-IoT Standard and Releases

Early in 2014, the LPWAN market rapidly developed thanks to the emergence of IoT. Realizing the need and potential for new communication ways, 3GPP started a feasibility study on cellular system support for an ultra-low complexity and low throughput IoT solution referred to as cellular IoT. In May 2014, Huawei and Vodafone proposed the Narrowband Machine to Machine (NB-M2M) to 3GPP as a study item to cope with the IoT market needs. Additional telecom industrial players got interested and later the same year Qualcomm proposed narrowband orthogonal frequency division multiplexing (NB-OFDM). In May 2015, 3GPP merged the two proposals (i.e., NB-M2M and NB-OFDM) and formed the Narrowband Cellular IoT (NB-CIoT). Eight months later, Ericsson proposed the Narrowband Long-Term Evolution NB-LTE. In September 2015, 3GPP included all proposals as a work item for Release 13. The key difference between NB-CIoT and NB-LTE was the number of the reused legacy LTE network resources to support interoperability. In June 2016 NB-IoT was recognized as a new clean slate RAT. Only further improvement changes were allowed and implemented thereafter.

In this regard, this section presents the main NB-IoT design changes from Release 13 until today that enabled the massive IoT connections with the corresponding solutions to respond to the adopted NB-IoT objectives. The enhancement features are classified following the objectives that are presented in the releases which would make it easier for the readers to refer back to the official 3GPP documents [8,9,34–38].

2.1. Release 13

3GPP introduced the following techniques in NB-IoT Release 13 to enable cellular massive IoT deployment for diverse use cases with low power, low complexity, and hence low cost. The introduced features and their corresponding objectives are as follows.

2.1.1. Mode of Operation

With the limited bandwidth requirement, NB-IoT can be deployed in three different modes i.e., standalone, in-band, and guard-band, as depicted in Figure 2. In in-band and guard-band modes, NB-IoT occupies one PRBs of 180 KHz in LTE spectrum both in the downlink and uplink. It can also be allocated as standalone where it occupies the 200 KHz bandwidth by "refarming" the GSM spectrum. These flexible deployment possibilities enable fast integration and coexistence with legacy LTE and GSM systems.

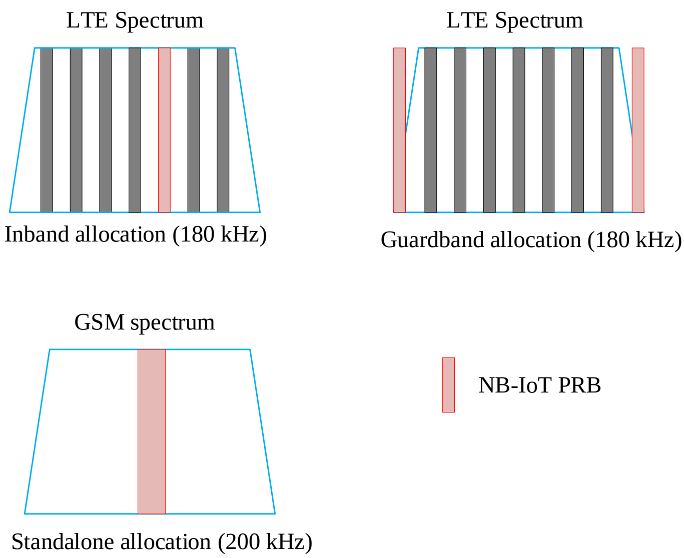

Figure 2. Narrow band Interet of Things (NB-IoT) Flexible Allocation inside Long-Term Evolution (LTE) spectrum (in-band and guard-band) and when refarming the Global System for Mobile Communications (GSM) spectrum (standalone).

2.1.2. Multi-Tone Transmission Support

To reach the massive device deployment objective, NB-IoT introduces the allocation of Resource Units (RU) to multiple User Equipment (UE) contrary to LTE where the whole resource block is allocated to a single UE in the uplink. In this regard, tones (frequency domain) with different duration are allocated to UEs. For the uplink transmission, each tone may either occupy 3.75 kHz or 15 kHz of transmission bandwidth based on the SC-FDMA scheme; for downlink NB-IoT uses 15 kHz of transmission bandwidth with OFDM scheme as LTE. With 15 kHz spacing, NB-IoT can dedicate either single-tone (8 ms) or multi-tone (3 tones, 6 tones, and 12 tones) to different UEs with the duration of 4 ms, 2 ms, and 1 ms, respectively. On the other hand, the 3.75 kHz spacing supports only single-tone allocation to different users with 48 subcariers of 32 ms duration [11,39,40].

2.1.3. Complexity and Cost Reduction Techniques

NB-IoT is required to have low complexity to reach the low-cost objective to facilitate massive connections. The features that were implemented to reach this objective include relaxed base-band processing, low memory storage, and reduced radio-frequency (RF) components. In this regard, the system bandwidth is set as narrow as 180 kHz with reduced frequency and time synchronization requirement. Also, NB-IoT uses the restricted BPSK and QPSK modulation schemes with only one antenna support both in uplink and downlink transmission.

2.1.4. Power Reduction Method

NB-IoT devices are intended to have a 10 years battery life to support massive deployment with limited human intervention. In this regard, two features i.e., Power Saving Mode (PSM), (from Release 12), and extended Discontinuous Reception (eDRx) (new feature from Release 13) were supported. These features are intended to extend the UE's battery longevity as follows:

In PSM, the NB-IoT device is configured to completely sleep while remaining registered online but cannot be reached by the base station signaling. In Release 13, the device can be in PSM mode for approximately up to about 413 days. In eDRX, the device is in an inactive mode for a few minutes to a few hours only.

In both cases, the partial or complete inability to receiving and sending different signals enhance the battery life longevity; however, choosing either PSM, eDRX or both depends on the corresponding use-case requirement. In this regard, the device can be synchronized to wake up from these modes by either Real-Time Clock (RTC), triggering from sensors, or both.

2.1.5. Physical Channels and Signals

NB-IoT adopts the same frame structure as LTE, with 1024 hyper frames, consisting of 1024 frames that contain 10 subframes of two slots with a duration of 0.5 ms each in the time domain. Similarly, in the frequency domain, NB-IoT contains 12 subcarriers of 7 OFDM symbols mapped in each slot. In addition to that, when NB-IoT uses the 3.75 kHz spacing on the uplink, 48 subcarriers are used with a slot duration of 2 ms.

The following channels and signals are used in the uplink:

- Narrowband Physical Random Access Channel (NPRACH).
- Narrowband Physical Uplink Shared Channel (NPUSCH).
- Demodulation Reference Signal (DMRS).

And the following are in the downlink frame:

- Narrowband Physical Downlink Shared Channel (NPDSCH).
- Narrowband Physical Downlink Control Channel (NPDCCH).
- Narrowband Reference Signal (NRS).
- Narrowband Primary Synchronization Signal (NPSS).
- Narrowband Secondary Synchronization Signal (NSSS).
- Narrowband Physical Broadcast Channel (NPBCH).

In general, NPRACH is used by UEs to perform initial access to the network, to request transmission resources, and to reconnect to the base station after a link failure. NPDSCH and NPUSCH are used to carry the downlink and uplink data packets transmissions, respectively. DMRS is used for uplink channel estimation accuracy. The UE acquires Master Information Block (MIB) from NPBCH and System Information Block (SIBs) from the NPDCCH. The defined MIB and SIB are broadcasted once during 640 ms and 2560 ms intervals, respectively. The timing of the remaining SIBs is configured in SIB1-NB. NRS is used for cell search and initial system acquisition. NPSS and NSSS are used by the UE for its frequency and timing synchronization with the base station. Due to overhead scheduling gaps in NPDCCH, the downlink and uplink peak data rates are ~250 kb/s and ~2267 kb/s, respectively, [34,40–43].

2.1.6. Coverage Enhancement Method

NB-IoT is designed to enhance coverage for the applications that are in hard-to-reach areas such as deep indoors and basements. In this regard, NB-IoT delivers an additional coverage of 20 dB as compared to the legacy LTE system. This corresponds to 164 dB of MCL. To enhance its coverage, NB-IoT uses up to 128 and 2048 retransmissions in uplink and downlink, respectively. Hence, this makes NB-IoT suitable for use cases that are latency insensitive as it can tolerate up to 10 seconds transmission delay.

2.2. Release 14 Enhancements

After the implementation of Release 13 features, studies erupted along with field trials that revealed the need for further enhancements to improve the quality of service as well as user experience. In this regard, 3GPP introduced further enhancement features to NB-IoT.

The enhancements features in Release 14 include positioning update, multicast services, and a new UE output power class in which the NB-IoT system throughput, mobility, service continuity and non-anchor carrier operation are improved [29,30].

2.2.1. Improved Positioning Technique

3GPP Release 14 introduces an indoor advanced positioning method of observed time difference of arrival (OTDOA) for NB-IoT to enhance UE position measurement of cell identity (CID). In OTDOA method, the UE measures the times of arrival (ToAs) of positioning reference signals (PRSs) received from different transmitters with respect to a reference node's PRS transmission to form the reference signal time difference (RSTD) measurements. In enhanced CID, the measurement requirements include the base station receive (Rx) and transmit (Tx) time difference, reference signal received power (RSRP), and reference signal received quality (RSRQ).

2.2.2. Multicast Services

The main objective of this mechanism is to optimize resources as well as transmission latency by addressing the data to a group of UEs at the same time rather than sending it multiple times to separate devices.

Therefore in Release 14, Multimedia Broadcast Multicast Services (MBMS) is supported through single-cell point-to-multipoint (SC-PTM). In general, SC-PTM is an efficient dynamic mechanism for optimal radio resource usage as it allows broadcast or multicast services to a specific group based on real-time traffic load and user requirement. SC-PTM uses NPDSCH by mapping Single-cell MBMS Control CHannel (SC-MCCH) and Single-Cell MBMS Traffic CHannel (SC-MTCH) that carry control and data traffic to the physical layer scheduled by using the downlink control information (DCI).

2.2.3. New Power Class for Narrowband-IoT User Equipment

Instead of the two power classes of Release 13 (i.e., 20 dBm and 23 dBm), in Release 14, the maximum allowed device's output power is reduced to 14 dBm. This has led to coverage relaxation of 9 dB that corresponds to 155 dB MCL as compared to 164 dB MCL and hence reduces the drained current. Technically, the use of the new power class facilitates the use of small coin-cell batteries and hence can be suitable for limited-size devices and applications that need a small battery. The compensation of the reduced NB-IoT power is achieved by increasing the NB-IoT transmission time to maintain the same energy per bit as the UE in Release 13 achieves. The newly introduced power class allows the serving base station to acquire the device power class during the establishment of the connection.

2.2.4. New Transport-Block-Size Support

Contrary to Release 13 where NB-IoT supports relatively low data rates (~250 kb/s and ~226.7 kb/s in downlink and uplink, respectively), 3GPP Release 14 introduces a new NB-IoT device category which supports the improved data rates by enhancing the Transport Block Size (TBS) to 2536 bits. These data rates can be reached thanks to the ability to support a second Hybrid Automatic Repeat Request (HARQ) process. This second HARQ is useful for enhancing the reliability of the link for the UEs that experience favorable channel conditions. Implementation of this optional second HARQ process results in throughput gain as it reduces the overhead caused by NPDCCH scheduling gaps.

2.2.5. Multicarrier Operation

To enable the massive NB-IoT deployment, in Release 14, NB-IoT can monitor paging and perform random access on non-anchor carriers. With this feature, one or more non-anchor carriers are added to the anchor carrier to carry out the synchronization and mobility measurements by using the NRS. Non-anchor carriers should also perform random access or paging when needed. Therefore, paging occasions and hence paging load will be spread over the anchor and non-anchor carriers and all carriers can then monitor paging.

2.2.6. User Equipment Mobility Enhancement

For the use cases that involve mobility, the temporary loss of radio interface impacts the system to a degree that can degrade link performance in terms of transmission errors. In this regard, 3GPP Release 14 introduces the possibility of Radio Resource Control (RRC) re-establishment for NB-IoT UE that supports data transfer via the control plane, i.e., the UE will try to re-establish the connection on that cell and resume the data transfer. This new RRC re-establishment feature hides the temporary loss of the radio interface to the upper layers.

2.3. Release 15 Enhancements

On top of all the enhancements that were introduced in Releases 13 and 14, the following improvements were introduced in Release 15 to satisfy the fast adoption of massive deployment with further improved quality of service.

2.3.1. Latency Reduction

In Release 15, NB-IoT supports new features to further reduce the transmission delay as well as to further reduce the power consumption dissipated during long transmission requirements.

In this regard, the NB-IoT UE is now able to support the physical layer Scheduling Request (SR) which is a special physical layer message to request the network to send the access grant (DCI format 0) so that the UE can transmit the uplink data. Also, NB-IoT uses a wake-up (Wu) signal to wake up the main receiver. This signal is transmitted in idle mode only when the UE is required to decode the physical downlink control channel in paging occasions. Therefore, power consumption reduction with the wake-up signal technique is larger when the UE wakes up from deep sleep more frequently (i.e., for shorter DRX/eDRX cycles). Also, significant power consumption reduction is achieved even when a common wake-up signal is used for a group of UEs. Quick RRC release and early data transmission during random access channel (RACH) procedure are supported to reduce the UE transmission latency and hence power consumption.

2.3.2. Semi-Persistent Scheduling

To enable better support of voice messages for the corresponding use cases, in Release 15, Semi-Persistent Scheduling (SPS) feature is introduced. In general, SPS is comprised of persistent scheduling for initial transmissions and dynamic scheduling for retransmissions. The base station assigns specific resource units to be used for NB-IoT UE voice messages with specific interval to save control plane overhead and hence optimize the radio resource usage. By principle, the base station preconfigures the UE with the Radio Network Temporary Identifier (SPS-RNTI) which is used to specifically differentiate one NB-IoT UE from another, or one radio channel from another. This SPS enables the NB-IoT data reception at a regular configured periodicity.

2.3.3. Small Cell Support

To further improve the capacity as well as coverage, in Release 15, NB-IoT supports small cell deployments. The downlink power to be reused for NB-IoT small cells is specified in section 16.2.2 of TS 36.213 [44]. In general, NB-IoT UE is not allowed to transmit more power than the configured maximum power, even if the configured power is lower than UE's maximum capability. This is done to avoid interference.

On the other hand, to extend the IoT connectivity especially in remote and rural areas for use cases such as agriculture, logistics, and environmental monitoring, NB-IoT is now able to support up to 100 km range. According to Ericsson, this could be achieved with a software upgrade only, without any changes in the existing NB-IoT hardware [45].

2.3.4. Enhanced User Equipment Measurements

Like in legacy LTE systems, UE measurements are critical since the corresponding reporting is mainly used to characterize the reference signal of a given bandwidth.

In Release 15, UE measurements are improved in a way that only NSSS additionally to NRS is defined for radio resource management measurement enhancement. This means that NRS is determined by the resource elements that carry NSSS in the NSSS occasions that the UE measures, through which the cell search and initial cell acquisition are improved.

2.3.5. Time Division Duplex (TDD) Support

In Release 15, a new feature of TDD support is introduced with a new TDD frame structure (type 2). For both 3.75 kHz and 15 kHz spacing, some specified restrictions are introduced i.e., only a normal cyclic prefix is supported for NB-IoT transmission. To support some of the TDD configurations with few downlink subframes, some of the system information (SI) can be transmitted on non-anchor carriers. In this way, the UE will have reduced system information acquisition and search time, and hence reduced UE differentiation and access control [30,46,47].

2.4. Release 16 Enhancement Prospects

3GPP and many industrial players are involved in ongoing discussions for Release 16 enhancements. The agenda includes the following objectives with their corresponding solutions.

2.4.1. Grant-Free Access

Most of the power consumption takes place during the NB-IoT UE active time, i.e., during Tx and Rx. In Release 16, the UE will be expected to transmit during RRC-Idle mode through Msg3 (RRC connection request) without access grant. A UE in RRC connected mode can transmit data without grant or with the simplified control-less grant. A further enhancement is on reducing NB-IoT signaling overhead while guaranteeing the needed quality of service. These features will reduce both power consumption and latency. In Release 16, it is also proposed to further study other signal waveforms (i.e., FDMA) that require less orthogonality with more relaxed timing advance (TA) alignment as compared to SC-FDMA.

2.4.2. Simultaneous Multi-User Transmission

The introduction of new schemes will enable simultaneous multi-user transmissions by using a shared resource in the time and frequency domains, such as Code division multiplexing (CDM), and multi-user multiple inputs multiple outputs (MU-MIMO), without increasing the number of antennae at the UE. In this regard, more dynamic access can also be achieved through enhanced base station receiver for detection of multiple users that are using the same resource unit as cluster and hence be able to schedule them effectively. This is because, for the last releases, NB-IoT UE uses the static or semi-static configuration of more resources for the unexpected application traffic handling. Similarly, the introduction of NB-IoT transmission without grant will cause a collision of data packets so dynamic handling of multiplexing is necessary.

2.4.3. Enhanced Group Message Mechanism

In Release 16, there should be more enhancements to support downlink command between user groups and group RNTIs. This is because MBMS which was proposed in Release 14 is only efficient for large size downlink command message transmission and requires many UEs to be deployed. For example, the application layer common message can be very small but sent to many UEs under a small group of UEs hence making MBMS not efficient for such applications.

2.4.4. Inter-RAT Idle-Mode Mobility

For applications such as smart tracking of logistics that involve mobility, the NB-IoT UE may still need to be accessible even when moved to the area served by other base station.

In this regard, 3GPP should introduce the new feature for NB-IoT UE support for inter-RAT mobility during idle mode. The mentioned feature is introduced along with optional handover support during connected mode through procedure simplification i.e., without dedicated signaling for measurement control and report. This is because handover helps to reduce system information reading time.

2.4.5. Network Management Tool Enhancement to Improve UE Differentiation

NB-IoT UE is expected to be able to perform differentiation according to maximal tolerable delay per service to optimize the radio resource usage. This is because, in the last release, the UE can be differentiated according to traffic model (periodic communication indicator, periodic time, scheduled communication time, traffic profile) and battery indication.

Section 2 has presented the NB-IoT standard and the corresponding enhancements from Release 13 until today. It has highlighted the main design changes and the corresponding further enhancements, i.e., deployment flexibility, physical channels and signals, positioning, multicast, new power classes, improved data rates, multicarrier operations, mobility support, improved scheduling, NB-IoT small cell support etc.

3. Narrowband-IoT: Protocol Stack

This section presents the NB-IoT protocol stack based on state of the art of the PHY and MAC layers to identify the knowledge gap and define future research directions. NB-IoT adopts the same protocol stack as the legacy LTE. However, some design changes in both PHY and MAC layers were introduced to support the massive long-range connections with up to additional 20 dB MCL than in legacy technologies such as LTE, GSM, and GPRS. Those changes are described in what follows.

3.1. Physical Layer

On the physical layer, NB-IoT adopts the same numerologies as legacy LTE along with OFDM and SC-FDMA signal waveforms in downlink and uplink, respectively. However, the resource scheduling unit in NB-IoT is the subcarrier (or tone) instead of PRB, to foster the network scalability by serving multiple UEs in a 180 kHz bandwidth. The downlink and uplink frame structures are as depicted in Figures 3 and 4, respectively.

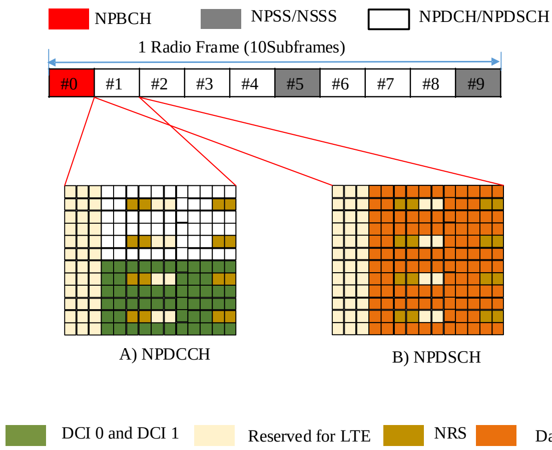

Figure 3. NB-IoT Downlink Frame Structure: subframe number 0 carries the Narrowband Physical Broadcast Channel (NPBCH), 1 to 4, and 6 to 8 carry the Narrowband Physical Downlink Control Channel (NPDCCH)/Narrowband Physical Downlink Shared Channel (NPDSCH), and 5 and 9 carry the Narrowband Primary Synchronization Signal (NPSS)/Narrowband Secondary Synchronization Signal (NSSS) (**A**) When the subframe is carrying control channels and (**B**) when the subframe is carrying data.

In general, the base station uses DCI to specify the scheduling information for a downlink/uplink transmission in NB-IoT. Then NB-IoT UE learns the deployment mode (standalone, in-band, or guard-band) as well as the cell identity through its initial acquisition, and it figures out which resource elements are already used by LTE. This is the way by which the UE can map NPDCCH and NPDSCH symbols to available resource elements. For example, in the downlink, NPDCCH is transmitted by aggregating the narrowband control elements (element 0 and element 1) where element 0 is occupied in subcarrier 0 to 5 and element 1 occupies subcarrier 6 to 11 in a subframe. The elements are determined by the type of DCI which is carried by NPDCCH to deliver scheduling command. Either two DCIs can be multiplexed in one subframe, or one DCI can be mapped in one subframe, corresponding to the aggregation level used [48]. However, NPDCCH, NPDSCH, and NRS cannot be mapped to the already occupied resource elements for LTE signals such as cell-specific reference symbols (CRS) and LTE physical downlink control channel (PDCCH). When NB-IoT UE receives NPDCCH which carries DCI, it decodes it and uses the device's scheduling feature (k0) to know the delay over which it will start to receive NPDSCH. The scheduling information is used to identify the allocated resources over NPDSCH and NPUSCH, respectively. In each NPDCCH, a maximum of two DCIs can be transported, and each UE can receive up to one DCI. The time interval between two successive NPDCCH opportunities is referred to as an NPDCCH period (PP) [48].

In the state of the art, different works have proposed solutions to the challenges that occur in PHY layer features, such as initial cell acquisition and synchronization, random access, channel estimation, error correction, and co-channel interference, as summarized in Table 2.

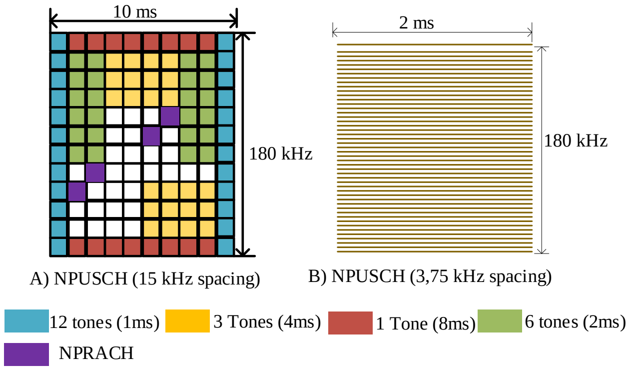

Figure 4. NB-IoT Uplink Frame Structure, (**A**) when 15 kHz spacing is used with different tone-allocation possibilities with slot duration of 0.5 ms and (**B**) when 3.75 kHz is used only single-tone allocation is supported with 4 times longer slot duration (2 ms).

Table 2. Articles on the proposed PHY layer enhancement techniques.

Feature	Article	Technique Used	Enhancement Criteria	Limitation
Cell Acquisition	[49]	Maximum-Likelihood (ML) NPSS detector	Average latency reduction for timing synchronization	It is a computationally complex detection method
	[50]	Cell search and initial synchronization algorithm	Time and frequency synchronization by using NPSS and NSSS with two-stage time domain NPSS correlation	mobility and new NB-IoT transmit power are not considered which have a direct impact on inter-RAT camping and the detected SNR, respectively
	[51]	Non-orthogonal spectral efficient frequency division multiplexing (SEFDM) waveform and an overlapped sphere decoding (OSD) detector	Resource optimization by the use of less bandwidth with better data rates compared to OFDM signal waveform	The proposed method would lead to sampling rate mismatch, carrier frequency offset and also will need to raise the computation complexity to NB-IoT UE
	[52]	New synchronization signal structure with Zadoff-Chu conjugates	Minimization of timing errors due to low-complexity NB-IoT frequency offset	If the same model is used for uplink synchronization it might lead to estimation errors if mobility is involved in NB-IoT
	[53]	NPRACH detection and time-of-arrival estimation for NB-IoT system	Enhancement on cell acquisition and channel estimation accuracy	The algorithm might not work for multi-tone allocation. Also, frequency hopping may raise power consumption as well as device complexity
	[54]	Receiver algorithm for NPRACH timing advance estimation and detection	Modeling the detection threshold to satisfy the NPRACH performance by lowering the probabilities of false alarm	The paper did not explain how receiver sensitivity can affect the NPRACH detection

Table 2. *Cont.*

Feature	Article	Technique Used	Enhancement Criteria	Limitation
Cell Acquisition	[55]	Mathematical modeling of NB-IoT performance	Throughput enhancement and NPRACH optimization by the use of repetition number, NPRACH preamble transmission per second and intersite distance	The work did not include some parameters such as the impact of mobility and how the achieved MCL for different coverage classes can impact the repetition assignment
	[56]	NPSS and NSSS frequency diversity reception	Time and frequency synchronization for cell search improvement	Alternative switching of NPSS and NSSS may require additional control commands which may lead to higher device complexity
Random Access	[57]	Configurable signal propagation model	System performance analysis in terms of number of supported devices, BER performance, preamble retransmissions, etc.	The impact of preamble retransmission on the overall transmission latency is not considered
	[58]	Mathematical evaluation of RACH preamble transmission	Analysis of NB-IoT transmission delay by using periodicity, start time, number of repetitions, number of preamble attempts and random access response window	Their model used minimum, intermediate, and maximum values for simulation which is so deterministic. However, it could be better to use random distribution to characterize NB-IoT realistic channel variations
	[59]	Random Access with differentiated barring (RADB) algorithm	Minimization of random access collision	Not resource efficient method since it does not include the impact of scheduling in different tone configurations
	[60]	New frequency hopping pattern of NPRACH preamble	Time-of-arrival estimation by the use of all the hopping distances	It only used a small cell scenario, if applied in dense NB-IoT network, estimation by considering all hopping distances may lead to system overhead and possible interference
Channel estimation	[61]	Frequency tracking algorithm	Frequency synchronization, as well as channel estimation for NB-IoT systems	More pilot signals, are used. This increases the overhead and hence can degrade the spectral efficiency
	[62]	Timing advance (TA) adjustment	Preamble sequence decoding by means of round trip estimation for coverage enhancements (on the sea)	It might not work for applications that do not involve a direct line of sight such as in dense urban environment
	[63]	MCS and coverage level optimization	Mobility effect on different coverage levels and how MCS affect paging performance	The channel model does not include other factors such as the effect of repetition, multipath, different Tx power for NB-IoT UEs as well as carrier frequency offset and inter-RAT operability
	[64]	New iterative algorithm for NB-IoT transmission scheme	NB-IoT error correction by using cryptographic redundancy and error correcting code	The channel estimation model to characterize NB-IoT transmission is not good, because some errors might be due to intersymbol interference and others due to intercarrier interference however the model does not explain

Table 2. *Cont.*

Feature	Article	Technique Used	Enhancement Criteria	Limitation
Interference mitigation	[65]	Channel Equalization algorithm	Intersymbol Interference mitigation by the phase-shifted channel frequency responses (CFR) to conquer the sampling mismatch between NB-IoT and base station	The proposed model did not consider the NPSS and NSSS impact ON time and frequency synchronization
	[66]	Mathematical model for sample duration in LTE and NB-IoT system	Interference and close-form interference analysis due to sampling mismatch between NB-IoT and base station	The model is computational complex when implemented in NB-IoT systems

3.1.1. Cell Acquisition and Synchronization

NB-IoT UE goes through the same process as LTE UE where to camp on a cell, it goes through frequency and timing synchronization to obtain the center carrier frequency as well as the allocated slot and frame timing used for the cell acquisition. In general, if MIB and SIB are properly decoded, cell ID, a subframe number, scheduling information, and system bandwidth can be detected successfully. In NB-IoT, the low complexity of devices may lead to poor synchronization and cell acquisition performance, especially due to carrier frequency offsets and poor channel estimation capacity. The following are the papers that have proposed different solutions to optimize the initial cell acquisition and initial synchronization procedure.

In [49], the authors presented a Maximum-Likelihood (ML) NPSS detector which is based on frequency domain cross-correlation metrics by using an overlap-save method. Their method achieves an average timing synchronization latency of 140 ms for the in-band deployed mode with SNR of −12.6 dB. Their proposed method showed a 34% reduction of the energy that is required for NPSS detection. However, their work showed only how much energy could be reduced with respect to the autocorrelation NPSS detection methods. It could be better to show how much of the total device's energy is consumed by their proposed computationally complex detector, i.e., it could be more realistic to include analysis in terms of reduction with respect to the energy consumed during time synchronization but also in terms of energy optimization over the total device consumption.

In [50], the authors presented an algorithm for initial synchronization and cell search. The proposed algorithm uses NPSS for timing acquisition and initial Carrier Frequency Offset (CFO) estimation called the two-stage time domain NPSS correlation. They also used NSSS sequences for the cell ID and frame timing. Their proposed algorithm showed that under extremely low SNR and different fading conditions, NB-IoT could provide the required performance and could also quickly camp on the cell, if any. However, practical experiments are still needed to prove the feasibility of these simulations especially on how the newly introduced NB-IoT power class and actual channel variations could have an impact on the detected SNR at the base station.

In [51], the authors presented an NB-IoT framework by using an advanced signal waveform called non-orthogonal spectral efficient frequency division multiplexing (SEFDM). This waveform uses less bandwidth as compared to OFDM waveform. The designed signal could improve the data rate without the need for more bandwidth. At the base station, the minimum Euclidian norm search detector is used for better error correction. The simulation results reveal that the proposed advanced signal waveform could achieve 25% improvement on data rate as compared to the OFDM signal waveform. The work also proposed an overlapped sphere decoding (OSD) detector which reduces the computation complexity as compared to the single sphere decoding detector while guaranteeing the needed performance. However, the model does not explain the impact of CFO due to the non-orthogonality of the subcarriers on the received signal.

In [52], the work investigated the downlink synchronization signal design and proposed the novel general synchronization signal structure with a couple of Zadoff-Chu (ZC) conjugated sequences in order to remove the potential timing errors caused by large frequency offsets. Their new synchronization signal structure demonstrated better functionality with the frequency offset tolerance of up to 40 kHz. However, the model does not explain the number of samples per symbol involved in synchronization operation and decision.

In [53], the random access preamble is discussed based on the design of NPRACH for single-tone frequency hopping only. It introduces the new single-tone frequency hopping random access signal used by NPRACH in NB-IoT systems. It further explains the design rationale and proposes some possible receiver algorithms for NPRACH detection and ToA estimation. The simulation results show that NPRACH performance is improved. However, the paper did not discuss the impact of massive interference which may result in lower received Signal Interference plus Noise Ratio (SINR) at the base station such that the lower the SINR the lower the detection probability of NPRACH. In addition to that the higher the number of devices the higher probability of NPRACH preamble collision. So lower SINR detected at base station and higher collision probability both affect NPRACH detection negatively.

In [54], the authors described a NPRACH design as specified in 3GPP in standard in Release 13. They proposed a receiver algorithm for the NPRACH timing advance estimation as well as detection. The simulation results for the NPRACH detection shows that if one preamble sequence is transmitted, the detection threshold should be set between 55% to 70% of the average value to satisfy the desired NPRACH performance at the lowest SNR. The results also showed that at 5 and 11 preamble sequence transmission, the detection threshold should be 50% and 35% of the average value, respectively. It is noted that increasing the detection threshold lowers the false alarm probability, which leads to an increased likelihood of misdetection.

In [55], the authors provided a mathematical model of an NB-IoT network in order to predict the optimum performance with a specific configuration of some design parameters (i.e., repetition, number of the preamble in NPRACH per second, coverage classes and intersite distance). The paper analyzes the effects of parameter choice in outdoor, indoor, and deep indoor. The work finally proposes how to choose the optimal configuration i.e., by providing the highest throughput, as well as success probability higher than minimal success probability with minimal one being of 90%. The work showed that even though the success probability has a maximum limit, it can still be altered by modifying the number of repetitions to enhance the coverage or the system capacity in terms of throughput.

In [56], the authors presented the NB-IoT frequency diversity (FD) reception for NPSS as well as NSSS. In the reception mode, the NB-IoT UE alternatively receives the NPSS and NSSS in time domain radio frame by switching the received signals transmitted in different resource blocks in the frequency domain. Their simulation results show that using the proposed FD reception could improve the detection probability by 16% more than without applying the frequency diversity. Additionally, using FD with precoding vector switching (PVC) transmit diversity, achieves 90% of physical cell ID detection (PCID) probabilities at the average SNR of 0 dB with maximum carrier offset of 70 kHz. The method also achieves 97% of PCID detection probability without consideration of frequency carrier offset.

3.1.2. Random Access Procedure

Like in LTE, NB-IoT random access (RA) is intended for initial UE uplink synchronization through which the UE acquires its unique UE ID used for communication with the base station. RA is also used to regain the lost UE access due to the long state of inactivity which has led to the loss in uplink synchronization. In NB-IoT, RA faces several challenges as seen on the research discussions; some solutions to improve the RA performance have been proposed as described in what follows.

In [57], the authors presented the random access procedure (RAP) model and analyzed the system performance by taking into consideration the configurable signal propagation model, a number of supported users per cell, and the RAP configuration parameters. The paper used the contention-based random access with Msg3 collisions instead of Msg1 collision (as multipath transmission) for the random access procedure. The proposed model results show the impact of the parameters (Msg3 transmission mode, Msg3 modulation and coding scheme (MCS), power control schemes and power ramping step) in the single-tone and multi-tone transmissions Bit Error Rate (BER) performance. The results are presented in terms of the total number of preamble transmission success, preamble retransmissions and lost preamble attempts. The work concludes that Msg3 must be considered in the random access procedure analysis, the transmission mode as well as MCS, and for better system performance and fairness distribution of UEs in the cell, it is better to configure power control correctly.

In [58], the authors analyze NB-IoT transmission delay as well as mathematical evaluation of the probability of success for the random access procedure preamble transmission. The analysis is based on three scenarios; scenario one uses minimum values of parameters, scenario two uses the intermediate values, and scenario three uses maximum parameter values. The used parameters are NPRACH periodicity, start time, number of repetitions, number of preamble attempts, and random access response window size. The average delay analysis was performed such that k preamble sequences are mapped in n subcarriers. The preamble collision occurs when multiple UEs send preamble sequences in the same subcarrier. A successful preamble attempt occurs when only one UE sends the preamble to a given subcarrier.

In [59], the authors investigated a random access optimization algorithm and summarized the NPRACH feature and hence designed random access with differentiated barring (RADB) for NB-IoT system. It is observed that the RADB could solve the preamble request conflict caused by massive NB-IoT UEs and hence provide reliable random access for latency-sensitive devices. However, the authors did not consider the problems of channel resource distribution and resource use rate.

In [60], the authors designed a new frequency hopping pattern of NPRACH preamble which uses all feasible hopping distances for a given number of subcarriers. It is seen that their proposed pattern was compatible with standards that is keeping the same NPRACH structure with only very small changes (hopping in the standard is allowed only between the subcarriers of the same resource group). Their simulation where they adopted their first traffic model which deploys 3000 devices, 48 ms and 40 ms of NPRACH preamble and periodicity, respectively, show that the proposed hopping pattern could improve the ToA estimation without additional system overhead.

3.1.3. Channel Estimation and Error Correction

Like in LTE systems, NB-IoT system performance depends to some extent on the quality of the channel estimation. However, for NB-IoT systems massive deployment, the poor quality of channel estimates is highly influenced by the low complexity of the UEs that can lead to misdetection of some signals, frequency offset, phase noise, passive intermodulation (PIM) on the device level, etc. To address the challenges that affect the channel estimation as well as to improve the quality of error correction to ensure the required performance with the low complexity, several works have proposed some solutions, as summarized in the following paragraphs.

In [61], the authors presented an NPSS detection method whose timing metric is composed of symbol-wise autocorrelation and a dedicated normalization factor in an in-band downlink NB-IoT system. The authors proposed a novel low-power algorithm for frequency tracking by the use of more pilot signals as compared to the LTE system. Their algorithm is implemented to compensate for the accumulated frequency offset during the NB-IoT transmission of NB-IoT. Their proposed frequency tracking algorithm delivers high estimation efficiency in terms of Minimum Mean Square Error (MMSE), the probability of correct cell acquisition, etc. However, their study did not elaborate on what could be the impact of mobility and inter-RACT support in the cell search procedure for NB-IoT.

In [62], the authors presented a practical coverage test on the ocean; it is shown that the proposed solution (where the base station decides whether the compensated round trip delay is short or long enough to decode the preamble sequence) was done by considering NPRACH design and hence the authors proposed their solution which considers the TA adjustment. Their proposed solution proved that NB-IoT coverage could reach as far as 35 km. However, the paper does not elaborate on the solution feasibility in environments without line of sight.

In [63], the work provided optimization cases for NB-IoT downlink in terms of MCS. The work also provided the optimization cases of coverage level (CL) by taking into consideration the RACH success rate with different driving speeds of NB-IoT devices in a commercially deployed network. Their results show that the base station paging success rate is decreased as the adjacent cell interference increases. However, the decrease in MCS improves paging performance. Coverage level 0 is the best choice for NB-IoT use cases that involve mobility, whereas coverage level 1 and 2 are mostly for fixed location NB-IoT use cases.

In [64], the author presented an iterative algorithm for NB-IoT transmission procedure. The simulation results in terms of BER and blocks error rate (BLER) show that by use of concatenated error correcting codes or cryptographic redundancy and error correcting code, the algorithm improves the NB-IoT coverage and reduces the overall NB-IoT power consumption. The modification of additional correction of low reliable bits could demonstrate the error correction of the damaged messages by the noisy transmission and hence can reduce the repetition number. However, this work did not discuss how effective the algorithm is when taking into consideration different channel conditions, payload sizes, as well as different repetition numbers with respect to device signal quality.

In [67], the authors considered the presence of random phase noise of the received signals mainly caused by oscillators impairments in both the transmitting and receiving sides and how to lower the mean square error (MSE) estimates. They presented the sequential MMSE channel estimation method that could be implemented in NB-IoT systems. Their model shows that if random phase noise is considered during channel estimation, it is possible to improve the detected SNR by up to 1 dB. However, the model is assumed to be uniformly distributed hence does not present the real-time channel which is randomly changing over time.

3.1.4. Co-Channel Interference

NB-IoT being deployed in the existing LTE spectrum, co-channel interference may occur between NB-IoT and LTE UEs. This is due to several reasons such as sampling rate mismatch, inter-PRB interference due to power leaking between NB-IoT and LTE PRBs, etc. To mitigate the impact of co-channel interference in the NB-IoT/LTE coexistence scenario, the following works have addressed the problems and proposed potential solutions.

In [65], the authors proposed the design guidance for channel equalization that can be used for 5G networks. The proposal set some assumptions such that the currently most used algorithms in cyclic prefix—OFDM system for pilot design, channel estimation, equalization, synchronization, and system performance analysis may no longer be applicable to NB-IoT systems. Their mathematical modeling demonstrated that channel equalization coefficients for NB-IoT UE are a set of phase-shifted CFR combination and not a simple Fourier Transforms of the channel impulse responses. This is the consequence of sampling rate mismatch between NB-IoT user and base station.

In [66], the authors established a comprehensive system model for in-band and guard-band NB-IoT by considering sample duration. They derived the mathematical expressions of received LTE and NB-IoT signals and analyzed the close-form interference power on LTE signal from adjacent NB-IoT signal. It is observed that the sample duration of NB-IoT significantly impacts the desired signal as well as interference on LTE UE; this is due to mismatched sampling rate between NB-IoT UE and the base station. Their proposed system model and derivations match the simulations, hence can be used for coexistence analysis for NB-IoT system.

Summary: This subsection has addressed the state of the art of NB-IoT PHY layer protocol. The main focus was set on different approaches to improve cell acquisition process, random access process, channel estimation, and interference mitigation. The next subsection focuses on MAC layer features by addressing the corresponding challenges and potential solutions.

3.2. Media Access Control Layer

Handling retransmissions (HARQ), multiplexing, random access, timing advance, choice of transport block formats, priority management, and scheduling are the tasks executed by the MAC layer. The discussion on this part focuses on features such as radio resource management, link adaptation, coverage, and capacity improvement, power, and energy consumption reduction, as summarized in Table 3.

Table 3. Articles on proposed MAC layer enhancement techniques

Feature	Article	Technique Used	Enhancement	Limitation
Resource allocation	[68]	Resource blanking	Interference cancellation by resource blanking	The proposed technique may lead to performance degradation in terms of spectral efficiency, especially for NB-IoT massive deployment.
	[69]	Iterative algorithm by a cooperative approach	Radio resource management in terms of scheduling index, repetition number and interference	The proposed solution is sub-optimal hence it does not provide maximum achievable performance in terms of maximum rate and capacity
	[70]	Scheduling algorithm	Efficient resource allocation by reducing the NPDCCH periods	Mobility is not considered and reducing NPDCCH period could lower the channel estimation quality hence may degrade the performance by unrealistic channel estimation
	[71]	Resource allocation technique by extending the specific PRB for paging traffic offload	power consumption reduction for NB-IoT UE during paging loading and offloading	The use of specific PRB for paging offloading is not an efficient use of the existing resource blocks. Also, the model is not applicable in standalone mode.
	[72]	NB-IoT scheduling algorithm	Interference analysis for 15 kHz LTE coexistence with 3.75 kHz guard-band NB-IoT	Emptying the LTE resource is not efficient resource use. Also, the model is not applicable for the standalone mode of deployment
Link adaptation	[73]	NB-IoT basic scheduler algorithm	Optimal resource usage by considering an average device delay and processing time	The scheduler did not consider semi-persistent scheduling, especially for inter-RAT networks
	[31]	Offset index selection and UE specific and common search spaces for NB-IoT dense networks	Cell capacity enhancement by means of optimal scheduling	Did not consider the number of sessions that each device has to transmit with respect to different requirements and use cases
	[74]	Link adaptation algorithm by using the mathematical expression of Shannon theorem	Coverage enhancement by characterizing SNR, repetition number and NB-IoT supported bandwidth	The work did not consider the impact of channel state information on UE link adaptation
	[75]	Two-dimensional NB-IoT dynamic link adaptation algorithm	Optimization of repetition number by dynamically adjusting MCS to ensure better BLER and BER performance	the model does not encompass the effect of speed and the deployment of the optional HARQ process to ensure better channel modeling

Table 3. *Cont.*

Feature	Article	Technique Used	Enhancement	Limitation
Coverage and capacity	[11]		NB-IoT coverage comparisons in different scenarios for 15 kHz and 3.75 kHz spacing	The channel estimation impairments, carrier offset as well as mobility with respect to different configurations are not considered for the claimed 170 dB of achieved MCL of NB-IoT
	[76]	Preconfigured access scheme and the joint spatial and code domain scheme	capacity and spectral efficiency improvement	It can only be applicable in small cell configurations when NB-IoT is deployed in large scale, preconfiguring access for different require
	[77]	Control plane small data transmission scheme	Effective data transmission enhancement by transmitting small packets in RRC connection set up	This scheme may results in NB-IoT signaling overhead due to Radio Resource Control (RRC) connection setup process encompassed with small data
	[10]	UE coverage and capacity simulation measurement based on real operators network parameters	NB-IoT enhanced coverage measurements by the use of real network configuration parameters	Optimal repetition number for NB-IoT devices is not considered, with additional penetration loss, it does not explain the additional repetition requirement to enhance the coverage while guaranteeing the required performance
	[78]	Low Earth Orbit (LEO) satellite to extend NB-IoT coverage	NB-IoT Coverage extension beyond LTE achieved link budget	The work did not consider the impact of repetition number on extended coverage as well as time and frequency synchronization that can lead to sampling rate mismatch as well as carrier frequency offset for low-end NB-IoT modules
Power management	[79]	Practical power measurement	Power consumption analysis for NB-IoT by varying payloads and repetition numbers, I-eDRx and PSM	Using two devices is not representative massive NB-IoT devices in the because different chips have different power consumption depending on the enabled features such as inter-RAT support that can affect the overall device consumption
	[80]	Prediction-based energy-saving algorithm	Reduction of power consumption by reducing the scheduling request procedure	The solution is not optimal because it reduces scheduling request without considering the device requirement with respect to channel parameters
	[81]	Semi-Markov chain for energy evaluation	Energy consumption and delay requirement evaluation for NB-IoT systems by considering the four states, namely power saving mode, idle mode, RACH procedure, and transmission mode	The model does not include the energy consumption during transition between the four mentioned modes and it does not include the impact of repetition on the device power consumption

3.2.1. Radio Resource Allocation

In NB-IoT, resource allocation is the key feature to ensure the expected massive connections in a cell. Tone allocations, PRBs, repetition number options, power configurations, subframes, or time slots, etc. must be optimized to maximize performance with minimum possible resources. Since NB-IoT is intended for low rate, less frequent time insensitive applications but with the required performance metrics, better radio resource management will ensure the optimal resource usage for expected throughput, spectral efficiency, and coverage enhancement.

In [68], the authors discussed the impact of interference for partial deployment of NB-IoT such that if one PRB is used for NB-IoT in some of the cells, that same PRB could be used for LTE in other cells for the in-band mode of NB-IoT operation. In such a deployment, possible co-channel interference may appear between NB-IoT UE and LTE UE. The authors modeled the partial deployment in percentile such that 100%, 75%, 50%, 25% represent the percentage of cells where NB-IoT enabled. Their results were analyzed by means of cumulative density functions (CDF) of respective SINR detected and maximum coupling loss achieved. The work demonstrates possible NB-IoT interference between NB-IoT and another NB-IoT UEs from adjacent cells and between NB-IoT and LTE from the adjacent cell. The simulation is performed for the in-band mode of operation where both NB-IoT and LTE UEs share the same PRB. They proposed the PRB blanking i.e., blanking the resources that are used by NB-IoT to not be used by LTE, not even being used for CRS. Blanking of these resources will omit the interference from LTE UEs and will result in NB-IoT only access to this PRB. However,the paper did not consider the performance degradation due to reduced available radio resources after when resource blanking is applied.

In [69], the authors formulated an analytical model to characterize the maximum achievable data rate, then investigated the impact of intercell interference in a multicell environment (for in-band and standalone scenarios), and finally proposed an iterative algorithm which uses cooperative approach which takes into consideration the overhead of control channels, repetition number, intercell interference as well as time offset. The proposed sub-optimal solution ensured better radio resource allocation, which raised the data rate by 8% and reduced the overall device energy consumption by 17% with respect to the non-cooperative approach.

In [82], the authors presented preliminary results of RSSI and detected SNR by developing a DORM (integrateD cOmpact naRrowband platforM) node which was deployed on a university campus to test its practical feasibility in different indoor scenarios.Their SNR and RSSI values were observed to be in the range of 18 dB to 23 dB and −65 dBm to −70 dBm, respectively, which shows its suitability for indoor coverage. The RSSI and SNR values variations are considered to be due to different elevations that the nodes are, with respect to the serving base station. However, the paper does not explain the channel estimation and measurements quality and their impact on the achievable throughput, moreover their paper does not cover the outdoor deployment and the impact of repetition on the overall devices' energy consumption when devices are located in different indoor environments.

In [70], the authors introduced the NB-IoT radio access strategy in detail and studied the NB-IoT scheduling problem. Their primary objective is to lower the number of used radio resource while each device's data requirement can still be satisfied. They furthermore formulated the NB-IoT scheduling problem and proposed an efficient algorithm to overcome such a problem. Their simulation results show that they could minimize the number of NPDCCH periods (NPs) used to satisfy each device's data requirement.

However, the repetition number is given according to the distance between the base station and the device. It could be better to use real-time channel parameters or MCS or BLER value to schedule the respective downlink channels to the devices. This is because, within the same distances, devices may experience different signal attenuation due to different factors such as fading, non-line of sight, line of sight, indoor placement, outdoor placement, underground placement, etc. Therefore, there is still an open space for practical deployment to analyze the effectiveness of the different downlink and

NB-IoT scheduling schemes which considers the device's simplicity, modulation schemes, channel conditions, and delay requirement for specific use cases.

In [71], the paper proposed a new resource allocation technique by extending the paging resource that will be specific for paging traffic offload. The authors noted that the new paging PRB could lower power consumption which is mostly used to load and offload the paging load. Also, the work proposed the selection scheme based on UE identity (ID) that is used to balance the load between the paging resource blocks. The simulation results show that power consumption reduction and resource optimal usage are of 80% and 30.5%, respectively. This work considered adding other PRBs for paging monitoring; however, the authors do not demonstrate the trade-off between the newly introduced scheme and the UE complexity requirements.

In [83], the authors proposed an enhanced access reservation protocol (ARP) that allows the device to transmit a fraction of a preamble sequence by providing an analytical model that captures the performance of ARP in terms of the false alarm, misdetection, and collision probabilities. They mathematically analyze the trade-off between the misdetection and the collision probabilities. The drawback of this protocol is that with massive NB-IoT deployment, altering the configuration of the protocol may result in detection performance degradation which can lead to huge packet loss.

In [72], the authors analyzed the impact of interference when the 15 kHz LTE system coexists with a guard-band NB-IoT system with 3.75 kHz subcarrier separation. Their simulation results demonstrated that it is desirable that the scheduler of the LTE system empties the neighboring RBs of the NB-IoT system and allocates resources if possible. The authors then proposed an NB-IoT scheduling method for the LTE system to improve the performance of the studied NB-IoT system. Their results showed that if emptying is not done, at 10^3 BER there is 1 dB drop of SNR as compared to when emptying of RBs is done.

3.2.2. Link Adaptation

Like in LTE, NB-IoT link adaptation involves adaptive modulation and coding schemes as well as adaptive power allocation. However, the modulation schemes are limited to QPSK to enable low complexity and hence reduce the overall power consumption. To extend the coverage and increase the link reliability, a repetition number of up to 128 times is introduced. In the literature, it is seen that NB-IoT link adaptation has several issues; potential solutions are also proposed, as summarized below.

In [31], the author formulated the scheduling issue such that the resource assignment must be in a specific format taking into an account reserved signaling resources and capabilities of the NB-IoT UE. They proposed a solution that incorporates two parameters which are (i) offset index selection (k0) and (ii) UE specific and common search space configuration. The offset index selection was chosen because with the limited k0 and varying size of payloads, it is critical to adapt the scheduling process for high resource use to accommodate more devices at the same time. Additionally, UE specific and common search space configuration were chosen because it decides the timing of NPDCCH and NPDSCH for different UEs, hence it can consequently improve the overall scheduling efficiency.

In [73], the authors presented a basic NB-IoT scheduler for NB-IoT system and analyzed the enhancements on average delay, optimal resource usage, and processing time. The proposed algorithm demonstrated that shorter NPDCCH period selection may reduce the UE average delay and optimize the overall system resource usage. Also, the model shows that the scheduling delay (k0) should be determined before the allocation of subcarriers. However, the model does not elaborate the type of configuration used since the choice of configuration such as single tone or multi-tones have a direct impact on periodicity and transmission delay and hence can directly impact the system performance.

In [74], the authors analyzed NB-IoT repetition number and bandwidth allocation and proposed analytic expressions based on SNR, bandwidth, and energy per bit that can be derived from Shannon theorem in order to characterize the impact of the repetition number as well as bandwidth allocation to different UEs. Additionally, their work proposed an algorithm for link adaptation. The algorithm exploits resource unit number, repetition as well as bandwidth. Their results show that reducing bandwidth and performing repetitions could enhance the coverage. However, the work did not consider the actual impact of channel parameters as well as NB-IoT UE impairments such as CFO which may lead to transmission errors.

In [75], the authors proposed a new NB-IoT link adaptation scheme with the consideration of the repetition factor. They claim that their proposed two-dimensional scheme is composed of Inner Loop Link Adaptation that copes with BLER by periodically adjusting the repetition number and outer loop link adaptation which coordinates the MCS and repetition number. This is because 20 dB coverage enhancement beyond LTE can be achieved by the repetition of transmitted data. So, in this work, they proposed an algorithm that dynamically chooses MCS and repetition number based on estimated real-time channel state information (CSI). However, their algorithm does not elaborate on the different NB-IoT power classes and to which range their respective coverage could be enhanced.

3.2.3. Coverage and Capacity

NB-IoT support for extended coverage of up to 164 dB of maximum coupling loss is to enable the technology to be used for cellular IoT services, especially for applications that are in hard-to-reach areas. Its narrow bandwidth and support for repetition are the key features to enable the enhanced coverage.

In [10], the authors simulated and analyzed the NB-IoT wide-area rural deployment and deep indoor urban deployment by using the network parameters of one metropolitan operator. Their work showed that NB-IoT devices could still transmit and receive data at an MCL of 167dB, which is 3 dB higher than the 3GPP's 164 dB of MCL limit set. Furthermore, in different indoor scenarios, even with an addition of 30 dB as penetration loss, NB-IoT had better outage probabilities as compare to another LTE LPWAN technology (eMTC). For outdoor and light indoor conditions with an additional 10 dB penetration loss and an average intersite distance of 2.8 km, NB-IoT had less than 0.1% of outage probability. However, despite the varying additional penetration losses of 10 dB, 20 dB and 30 dB, their simulation does not consider the impact of features such as mobility, CFO, lower power class on the achieved MCL.

In [11], the authors showed that for the maximum number of repetitions (128 times), with 15 kHz and 3.75 kHz subcarrier spacing, coverage of up to 170.2 dB MCL and 174.2 dB MCL could be achieved, respectively. The work concluded that the evaluations show that NB-IoT could provide up to 20 dB coverage enhancement in various deployment scenarios as compared to legacy LTE. Similarly, the work did not study the impact of mobility and weak channel estimation quality to the achieved MCL.

In [76], the authors proposed two less complex scheduling schemes (compared to brute-force) that can be used NB-IoT. The first proposed scheme is called the preconfigured access scheme and the second is the joint spatial and code domain scheme. Their simulation performance results (spectrum efficiency, number of active devices as well as low collision rate achieved) by the two low complex schemes were found better when compared to the ones that can be achieved by the brute-force scheme.

In [78], the authors proposed a specific unidirectional system to study the coverage enhancement by using satellite network i.e., LEO constellation. Their proposed model with the mathematical derivations shows that NB-IoT could achieve the 20 dB more than LTE achieved MCL and could still operate according to Release 13 standards. From their results, it is seen that the packet error rate (PER) of the transmitted signal is distorted by Doppler spread. However, the model does not consider the clock synchronization between NB-IoT device and satellite, which can lead to performance degradation especially caused by the CFO or the sampling rate mismatch. Additionally, the work did not consider the maximum achievable throughput when their system is employed to comment on the effectiveness of the techniques as compared to terrestrial NB-IoT deployment.

3.2.4. Power and Energy Management

The NB-IoT reduced complexity is intended to reduce the power consumption in different modes. PSM and eDRX are the implemented features dedicated to foster the long-lasting battery life.

In [79], the authors presented the NB-IoT power measurement. Their measurements were set in such a way that NB-IoT transmission consumes 716 mW when at 23 dBm with a power efficiency of 37%. DL control and data signals consume 213 mW, idle-mode-eDRx and PSM consumes 21 mW and 13 µW, respectively. In general, according to their empirical measurements, it is shown that the power consumption is 10% lower than the 3GPP estimates. During measurements, parameters such as time domain repetition, I-eDRx, and PSM were taken into consideration. To characterize each component in the proposed model, several test cases such as Tx power, UL, and DL data rates, I-eDRx, and PSM were used and all parameters except one at a time were fixed. Their results showed that the NB-IoT devices power consumption is independent of the subcarrier spacing. However, the total ON time of the devices is in many cases defining the overall battery life. As their remark, the data rates do not directly impact the power consumption, but it has a major direct impact because it defines the overall device ON time. If the transmitting interval is 1 h, the device achieves only 2.5 weeks of battery life. Increasing the duration to 24 h, the lifetime of the device increases to 12.8 years in PSM.

In [80], the authors proposed a prediction-based energy-saving mechanism to reduce energy consumption by decreasing the number of scheduling request procedures. Their proposed scheme showed that it could reduce the NB-IoT active time from 5% to 16% for the medium and bad channel quality and achieve from 10% to 34% battery saving in different scenarios as compared to 3GPP consumption simulation specifications in [43].

In [81], the authors developed a semi-Markov chain with power saving mode, idle mode, random access, and transmission mode to study the energy requirement and delay performance for NB-IoT. It is noted that for massive synchronous connections, extra power is drained in random access and transmission states due to collisions. The paper further proposes an energy optimization model based on a priori method that takes into consideration the PSM duration as well as power consumption. The results demonstrate that for optimal energy and delay requirement, it is important to set the higher RACH transmission number to accommodate more delay on the UE. However, their optimization model did not consider the power consumption during the transition of different states, because when the UE is required to perform several sessions per day, it might go through several transitions that have a significant effect on power consumption. Furthermore, the mode does not include the small data transmission scheme during RRC connection as proposed in the updated standards. However, with the introduction of the new power class in NB-IoT Release 14, there is a need for practical experiments to evaluate the new coverage classes. With lower transmit power, the SNR detected at the base station becomes lower hence the device will need to perform more repetitions to enhance coverage.

3.3. Upper Layers

Although the focus of this paper is mainly on the features regarding PHY an MAC layers, it is still imperative to address some enhancements, challenges, and potential solutions to the upper layers. Especially the changes that are implemented in Evolved Packet Core (EPC) by adding the Service Capability Exposure Function (SCEF) to manage both IP and non-IP data packets [84].

Control and User Plane Optimization

To support massive end-to-end device connectivity with extremely low complexity and reduce the transmission signaling, NB-IoT implements new small data transmission procedures based on Cellular IoT (CIoT) Evolved Packet System on both Control Plane (CP) and User Plane (UP). These transmission procedures support small bursts of data efficiently while guaranteeing the long-range coverage as compared to legacy GPRS [85,86]. In this regard, NB-IoT is can support more than one data path in CP for the transmission of user data which is carried by the signaling messages managed by the Mobile Mobility Entity (MME) as shown in Figure 5. The procedures are optimized to efficiently support the small data transfer as follows:

- Mandatory CP CIoT EPS;
- Optional UP CIoT EPS.

CP CIoT EPS optimization encapsulates the data packets in Non-Access Stratum (NAS) by using control plane signaling messages. In this regard, this procedure is mandatory. Compared to the conventional SR procedure, the NB-IoT UE skips some steps required for each data transfer hence this optimization procedure best fits the short data transmission or reception.

On the other hand, UP CIoT EPS optimization requires the RRC connected mode to get the scheduled radio resources as well as Access Stratum (AS) between the UE and the network. This mode uses the newly introduced connection to Suspend and Resume procedures. Connection suspend procedure helps to retain the network context so that the UE can resume the connection when traffic is available. Retaining the context helps the UE and the network to skip the AS and RRC reconfiguration in each data transfer. Since it uses user plane, the UP CIoT EPS is suitable for both small and large transactions.

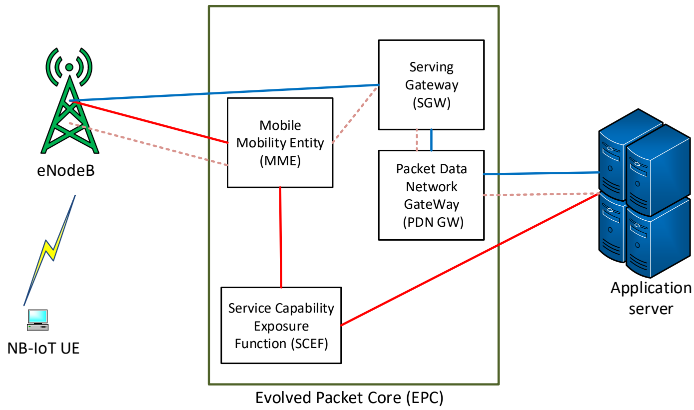

Figure 5. Representation of NB-IoT IP and Non-IP data path: Blue line displays the IP data path in UP mode (as Legacy LTE), Red line displays the non-IP data path in CP mode, and dashed-line displays the IP data path in CP mode.

Furthermore, the UE in Service Request procedure (an LTE procedure used by the UE and base station to transmit or receive data in RRC idle state) is required to be in a connected state in order for base station to allocate the radio resources. For NB-IoT this SR is optional; however, NB-IoT UE that supports UP optimization needs also to support SR. For example; if the NB-IoT UE wants to transmit the uplink data in idle state, it will send the random access preamble through which the base station and UE will establish RRC connection and UE will be allocated with the radio resources for data transfer. After a certain period of inactivity, the base station initiates the release procedure.

Similarly, for UE downlink data reception, if the UE is in DRX mode, the UE regularly listens to downlink signaling and if the UE notices the paging message, it will perform the SR procedure as described in uplink data transmission. Additionally, if the UE is in PSM mode, it will be completely inaccessible until it initiates the same SR procedure for the uplink grant or by using Tracking Area Update (TAU).

There are works that are addressing the upper layers such as [77], where the authors proposed an efficient small data transmission scheme by using CP procedure. The proposed scheme enables the devices to transmit data packets through the RRC connection setup procedure when the device is in idle mode. This process reduces the signaling overhead caused by the security setup process and data radio bearer setup process. However, a suggestion could be to analyze the power consumption during this small data transmission and compare its effectiveness to when the same data is transmitted during the UP procedure.

Summary: This section discussed PHY layer features, highlighting the corresponding enhancements on cell acquisition procedure, random access channel estimation, and interference mitigation. It then addressed the MAC layer enhancements regarding resource allocation, link adaptation, coverage and capacity, and power management. It further addressed the upper layers changes related to cellular IoT evolved packet system optimization through user and control planes to enhance the small data packets transmissions for end-to-end massive connectivity.

4. Narrowband-IoT Possible Deployment Strategies

This section proposes potential deployment strategies for NB-IoT massive deployments by considering the NB-IoT support for small cells in heterogeneous network scenarios.

HetNets are effective network deployment strategies in which small cells are incorporated in macrocells with the objective of improving performance in terms of capacity, coverage, and spectral efficiency. In general, the macrocells are characterized by higher transmit power and broader range as compared to small cells. When smalls cells are overlaid in macrocells, interference becomes a concern, especially to small cell edge users. Several techniques for interference cancellation, estimation and coordination that involve frequency hopping, frequency reuse, power control etc. have been proposed; however, the performance trade-offs for the proposed techniques for macrocells and small cells are still challenging [87–90].

Similarly, NB-IoT is expected to coexist with the currently deployed legacy LTE as well as the forthcoming 5G networks. This questions the existing interference management techniques i.e., are they applicable to the newly deployed technology since NB-IoT is expected to support different power classes while maintaining the low complexity which can severely affect the channel estimation quality and hence interference estimation quality. In NB-IoT coexistence with the legacy cellular networks, the possible deployment scenarios are as follows:

- Synchronous NB-IoT deployment in all small cells;
- Asynchronous NB-IoT deployment in all small cells;
- Synchronous NB-IoT deployment in small cells and Macrocells;
- Asynchronous NB-IoT deployment in small cells and LTE in macrocells.

These scenarios, as shown in Figure 6, are detailed in what follows.

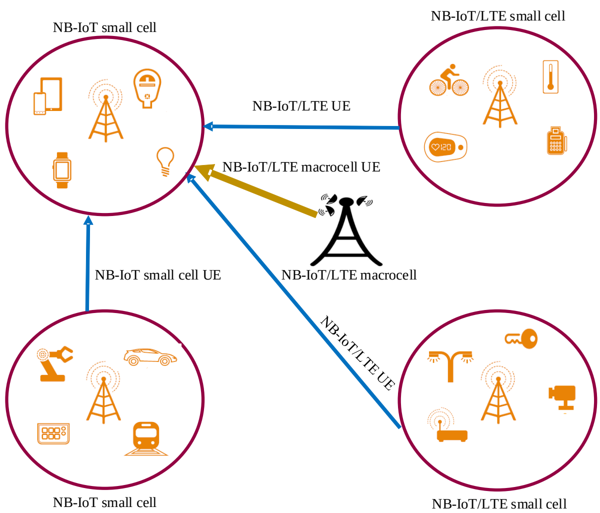

Figure 6. Summary of NB-IoT deployment strategies. For example, when NB-IoT is deployed in macrocell and LTE in small cell, when LTE is in macrocell and NB-IoT is in small cells, when NB-IoT is in macrocell and small cells support both NB-IoT and LTE, and when LTE is in macrocell and LTE/NB-IoT is in small cells

4.1. Synchronous NB-IoT Deployment in All Small Cells

This is the NB-IoT deployment strategy which is enabled in all the small cells by using the same physical resource blocks. All the small cells are synchronized in such a way that with the same PRBs, all the NB-IoT UEs are using the transmit power that is configured regardless of its maximum transmitting power capacity. This means that even though NB-IoT devices might support different power classes such as 14 dBm, 20 dBm, or 23 dBm, the NB-IoT devices will only be configured to use the minimum allowed transmit power in order to avoid causing the co-channel interference to other UEs using the same radio resources.In this strategy, power control may be the key feature to ensure the required performance. However, cell edge UEs may still suffer from the interference problem. This interference may highly be increased due to the low channel estimation quality of NB-IoT UEs associated by its reduced computational complexity.

4.2. Asynchronous NB-IoT Deployment in All Small Cells

This deployment strategy is employed in such a way that NB-IoT is enabled in all small cells by using different physical resource blocks. This implementation may avoid the interference between NB-IoT UEs from different small cells; however, this may result in co-channel interference between NB-IoT and LTE UEs that are using the same radio resources. When deploying under this strategy, it is imperative to implement proper frequency planning as well as proper power configuration for NB-IoT devices. As seen from the state of the art, some works have proposed blanking of the radio resources to the adjacent cells for the resources that are already occupied by NB-IoT even in the cells that NB-IoT is not enabled. However, blanking of the resources is a wastage of resources, so, there should be some other means such as frequency hopping to avoid wastage of resources (blanking) as well as to mitigate interference.

4.3. Synchronous NB-IoT Deployment in Small Cells and Macro Cells

In this strategy, NB-IoT is enabled in the small cells as well as in macro cell on the same PRBs. Macrocell UEs are configured to use higher transmit power as compared to small cell UEs while keeping the same PRBs for NB-IoT while others left for legacy LTE. Possible co-channel interference may occur in small cell edge UEs if the UEs are scheduled on the same resource units. The impact may further increase for UEs under mobility which might require the use of handover for smoothing the UEs transition from one serving cell to another. From our review, no work has addressed the interference cancellation mechanism for such a case. It is imperative to employ the existing geographical planning, frequency reuse, frequency hopping, and power control while considering the low complexity but high coverage range NB-IoT.

4.4. Asynchronous NB-IoT Deployment in Small Cells and LTE in Macrocells

In this strategy, NB-IoT uses separate PRBs between small cells and macro cells. This means that one or more PRBs are used for small cells and different PRB(s) for the macrocells. If the PRBs are not well planned, NB-IoT users from adjacent cells (using the same resource units) may suffer from interference. Also, LTE users that are using the same resource elements may interfere with small cell or macro cell UEs. Different transmit power control configurations may be used to control interference.

The choice of the deployment strategy depends on several factors such as use-case requirements, environmental conditions, equipment quality, etc. It is imperative to implement better interference estimation, mitigation or management techniques that will ensure better performance and spectral efficiency for the massive NB-IoT deployment in coexistence with other technologies.

Summary: This section has presented the possible NB-IoT deployment strategies by considering the NB-IoT support for small cells in coexistence with legacy LTE in HetNet scenario.

The following section presents the open research challenges to motivate future research directions.

5. Open Research Questions and Discussion

5.1. Battery Life

PSM and eDRx were introduced in NB-IoT Release 12 and 13 to lengthen the NB-IoT devices' battery life. Moreover, the most recent updates require the UE to be able to transmit during RRC-idle mode which will reduce the required ON time for data transmission. However, devices experiencing bad channel conditions due to hard-to-reach areas will require to perform several retransmissions per session, which will drain the device's energy and hence shortens the battery life. Similarly, devices that require a relatively large number of reporting sessions per day will consume more energy, which makes energy management a concern. As seen in Section 3, most of the proposed algorithms are power hungry because most of the power is consumed during transmission and reception. Therefore, energy harvesting alternatives such as solar, biogas, vibrations, etc. that will lengthen the NB-IoT device battery life should be introduced to complement or replace frequent battery charging.

5.2. Radio Resource Management

5.2.1. Tones Allocation.

As seen in the literature, most of the articles consider single-tone allocation for the simplicity in the simulation, thus, multi-tone allocation is not well studied. This causes a knowledge gap in the effectiveness of different tone-allocation possibilities. Moreover, for guard-band, in-band and standalone it is still not clear about the respective performance metrics that could be achieved in terms of throughput, coverage range, interference robustness etc. This restricts to a certain extent the optimal choice of deployment for a large number of devices with the required performance. Furthermore, different frame structures, especially for TDD configurations, are not discussed even though NB-IoT is required to support TDD. Therefore, optimal resource use techniques must be proposed that incorporate repetition, mobility, tones allocation, etc. for efficient spectrum usage.

5.2.2. Interference Mitigation

Interference prediction, estimation, cancellation, and coordination techniques for NB-IoT become a challenge. This is because of the sharing of spectrum resources between NB-IoT and legacy LTE. Similarly, with NB-IoT being deployed in a small cell or macrocell scenarios in heterogeneous networks, interference becomes a concern. Several works have tried to address this by means of resource blanking, power control, or better uplink and downlink scheduling schemes and frequency and timing synchronization, etc. However, it is still challenging to incorporate the NB-IoT features such as repetition, low complexity (which affects channel estimation quality), and mobility in deploying the already existing LTE interference management techniques. As seen in the possible NB-IoT deployment scenarios above, there is still a need for deploying effective schemes that will ensure better NB-IoT performance without degrading the LTE performance [91,92].

5.3. Mobility Management

As seen in Section 2, most of the simulation works have ignored the mobility impact of NB-IoT channel modeling. However, for use cases that involve movement, Doppler shift has to be taken into consideration during channel estimation, which might slightly increase the device complexity to support handover and other mobility features such as the support for inter-RAT mobility during idle mode [93,94]. The increase in NB-IoT UEs mobility makes the channel suffer from fast varying channel conditions, due to which adaptive transmission schemes that might involve channel estimation, error correction, etc. must be implemented.

Therefore, applying intelligent/adaptive algorithms that are low power and optimal for repetition number, yet mobility-aware, is of great importance. The algorithms could involve low-power frequent CSI reporting, early data transmission by using both user and control plane in either Msg 3 or Msg 4.

5.4. Latency

NB-IoT latency tolerance is set to 10 ms. This is due to its support for use cases of UEs that are in environments with bad channel conditions [95–97]. Initial cell acquisition, frequency, and timing requirements, RACH transmission, half duplex mode of transmission and several repetitions that are performed during transmission are some of the features that play part in the overall data transmission delay. Several works are trying to reduce the timing requirement so as to reduce transmission latency of devices; however, most of the works have not addressed delay by taking into consideration the massive congestion that is expected for the IoT networks, processing delays due to low complex devices, queuing delays, propagation delays especially with long-range feature, as well as errors and error recovery.

However, early data transmission schemes and the second NB-IoT HARQ process for devices that have good channel conditions are among the features that can be used to reduce the transmission latency and improving the transmission link performance. However, only a handful of research articles have discussed the effectiveness of these processes when applied in NB-IoT.

5.5. Semi-Persistent vs. Dynamic Scheduling

Most of the NB-IoT literature addresses dynamic uplink and downlink scheduling by studying the scheduling of logical channels and signals. There are still very few NB-IoT studies about the effectiveness of Semi-Persistent Scheduling schemes (SPS) even though SPS helps to reduce the NPDCCH overhead as compared to dynamic scheduling. It provides the NB-IoT UEs with longer allocated resources (more than one subframe) so that the NB-IoT device will not need the frequent downlink assignment as well as an uplink grant which is delivered by NPDDCH for each subframe. However, for applications that involve mobility or fast varying channel conditions, how is this scheduling scheme going to be effective knowing that NB-IoT has poor channel estimation capabilities as compared to LTE?

5.6. Random Access

Massive NB-IoT modules that try to request the radio channel resources at the same time for uplink data transmission may suffer from random access preamble collision. This is caused by several factors such as detection inaccuracy that may not satisfy the detection threshold, the high probability of false alarm, etc. Several works have proposed random access preamble detection algorithms (i.e., random access with differential barring etc.) and others have developed mathematical models to characterize the preamble transmissions in order to improve the NPRACH success rate and better time-of-arrival estimation and other NPRACH performance improvements. However, it is still unclear which scheme is effective for massive deployment, since most of the proposed schemes do not consider the heterogeneous network architecture, channel estimation impairments, or realistic channel conditions [98,99].

5.7. Timing Advance (TA)

When the base station responds to NB-IoT UEs about RRC connection request, it incorporates the TA command to be used for NB-IoT UE terminal data uplink transmission timing (i.e., to time-synchronize the UEs to the base station and help to compensate the propagation delays). However, for NB-IoT UE, the TA adjustment accuracy of the signaled timing advance with respect to the prior uplink transmission may highly be affected by the massive number of NB-IoT devices contending for the access. This is because the base station may need to correct some UE timing while for other NB-IoT UEs that had already transmitted NPRACH could receive the random access response which is not intended for them. Some works have addressed the receiver algorithms for NPRACH TA estimation as well as detection timing advance adjustment decoding schemes to improve the estimation but the NB-IoT receiver sensitivity and weak channel estimation quality still negatively affect the TA adjustment.

5.8. Cell Search and Initial Synchronization

NPSS and NSSS are two signals based on frequency domain Zadoff-Chu sequence that are used for NB-IoT time and frequency synchronization to the base station. According to NB-IoT standard, NPSS and NSSS may not be transmitted on the same antenna port hence NB-IoT initial synchronization may rely on NPSS only. The challenge is that the imperfect channel conditions may severely affect the cell camping procedure as a small CFO may result in a phase shift to a received frequency domain sequence which as a consequence may degrade the cell search and synchronization performance. To improve this, frequency diversity techniques should also be used for NPSS and NSSS reception improvement.

5.9. Unified NB-IoT Testing Tool

Since NB-IoT is a promising technology, there should be a unified testing tool used as a reference to verify if the produced products comply with the standards. Taking Bluetooth as an example, for better compatibility towards different available products from handsets to car kits, Profile Tuning Suite (PTS) software is used to automate the compliance testing to specific Bluetooth function. So, to support compliance with standards and hence backward compatibility and interoperability, what is the testing tool to validate if different available products will fit standards? Similarly, for simulation purposes, most of the works choose the parameters that can generate results easily. If there is a concrete simulation model that takes into account the major NB-IoT features and incorporating all the possibilities from repetition number allocation, mobility selection, modulation and coding scheme, real-time channel variations, etc. it would be easier to get realistic modeling for different scenarios.

5.10. Backward Compatibility and Interoperability

A ten-year telecommunication generation is characterized by different changes in releases and updates. In order to reach their lifespan as compared to what the standards stipulate, NB-IoT devices should operate for around ten years with a single battery charge. Whenever new releases or updates are introduced, backward compatibility and interoperability should be possible. Apparently, the device complexity is set as low as possible; will these simple devices (hardware) support hard and robust algorithms that will be implemented by over-the-air upgrades/updates to satisfy the demands of future NB-IoT use cases?

Summary: This section has presented the open research questions regarding battery life, radio resource allocation, cell search, and initial acquisition procedures, mobility management, latency, random access, etc., as summarized in Table 4, in order to motivate future research directions. The next section concludes the paper.

Table 4. Open Research Questions related to the physical layer, MAC layer, and standard.

Physical Layer	MAC Layer	Standard
Radio resource management	Timing advance adjustment	Support for small cell
Frequency and time synchronization	Dynamic scheduling and semi-persistent scheduling	TDD support
Random access	Latency	Antenna diversity
Channel estimation	Power management	Mobility and handover support
Error correction	Network throughput	More efficient group messages
Link adaptation	Control packet overhead	Multicarrier operation
Interference mitigation	Control plane small data transmission	Network management tool for UE differentiation

6. Conclusions

Due to the fact that most of the existing works are segmented and only consider one or two releases in their corresponding studies or simulations, this paper has presented a comprehensive overview of NB-IoT standard from Release 13 to Release 16 prospects to enhance and enable more realistic research. It further presented the detailed current state of the art of NB-IoT based on the ongoing discussion on NB-IoT protocol stack along with the related contributions and analyzed the knowledge gaps by using NB-IoT standard as a benchmark. It is observed that most of the articles focus on improving one or few features while neglecting others,it could be better to display the trade-offs between the improvement feature and the neglected ones, i.e., performance trade-off between PHY and MAC layer when one feature is changed in either of the layers, the impact of repetition on overall energy consumption, CFO on channel estimation quality etc. This paper also presented the NB-IoT deployment strategies to highlight the coexistence possibilities with other legacy technologies i.e., LTE, by considering the NB-IoT support for small cells in HetNet scenarios. Lastly, it discussed the open research challenges and the future common research focus on NB-IoT i.e., battery life, optimal resource usage, handover support during mobility, transmission latency, scheduling, etc. To the best of the author's knowledge, this is the first survey that covers broadly these mentioned contributions and hence this work will help the researchers get most of the needed information to accelerate their research by finding the relevant information and sources for deeper exploration of the research concepts as well as finding possible solutions.

Author Contributions: Conceptualization, C.B.M., H.M., M.M.A., Y.L.M., S.P. and S.M.; investigation, C.B.M., H.M., M.M.A., Y.L.M., S.P. and S.M.; resources, C.B.M., H.M., M.M.A., Y.L.M., S.P. and S.M.; writing—original draft preparation, C.B.M..; writing—review and editing, C.B.M., H.M.; formal analysis, C.B.M., H.M., M.M.A., Y.L.M., S.P. and S.M.; supervision, M.M.A., H.M., Y.L.M.

Funding: This research received no external funding.

Acknowledgments: This project has received funding partly from European Union's Horizon 2020 Research and Innovation Program under Grant 668995 and European Union Regional Development Fund in the framework of the Tallinn University of Technology Development Program 2016–2022. This material reflects only the authors' view and the EC Research Executive Agency is not responsible for any use that may be made of the information it contains.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lucero, S. IoT Platforms: Enabling the Internet of Things. 2019. Available online: https://cdn.ihs.com/www/pdf/enabling-IOT.pdf (accessed on 12 January 2019).
2. Salman, L.; Salman, S.; Jahangirian, S.; Abraham, M.; German, F.; Blair, C.; Krenz, P. Energy efficient IoT-based smart home. In Proceedings of the IEEE 3rd World Forum on Internet of Things (WF-IoT), Reston, VA, USA, 12–14 December 2016.
3. Bardyn, J.; Melly, T.; Seller, O.; Sornin, N. IoT: The era of LPWAN is starting now. In Proceedings of the ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference, Lausanne, Switzerland, 12–15 September 2016; pp. 25–30. [CrossRef]
4. Devalal, S.; Karthikeyan, A. LoRa Technology—An Overview. In Proceedings of the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Tamilnadu, India, 29–30 March 2018.
5. Mekki, K.; Bajic, E.; Chaxel, F.; Meyer, F. A comparative study of LPWAN technologies for large-scale IoT deployment. *ICT Express* **2019**, *5*, 1–7. [CrossRef]
6. Lauridsen, M.; Nguyen, H.; Vejlgaard, B.; Kovacs, I.Z.; Mogensen, P.; Sorensen, M. Coverage Comparison of GPRS, NB-IoT, LoRa, and SigFox in a 7800 square km Area. In Proceedings of the IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, Australia, 4–7 June 2017.
7. Sinha, R.S.; Wei, Y.; Hwang, S.H. A survey on LPWA technology: LoRa and NB-IoT. *ICT Express* **2017**, *3*, 14–21. [CrossRef]

8. Rico-Alvarino, A.; Vajapeyam, M.; Xu, H.; Wang, X.; Blankenship, Y.; Bergman, J.; Tirronen, T.; Yavuz, E. An overview of 3GPP enhancements on machine to machine communications. *IEEE Commun. Mag.* **2016**, *54*, 14–21. [CrossRef]
9. Gozalvez, J. New 3GPP Standard for IoT [Mobile Radio]. *IEEE Veh. Technol. Mag.* **2016**, *11*, 14–20. [CrossRef]
10. Lauridsen, M.; Kovacs, I.Z.; Mogensen, P.; Sorensen, M.; Holst, S. Coverage and Capacity Analysis of LTE-M and NB-IoT in a Rural Area. In Proceedings of the 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, Canada, 18–21 September 2016.
11. Adhikary, A.; Lin, X.; Wang, Y..E. Performance Evaluation of NB-IoT Coverage. In Proceedings of the IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, Canada, 18–21 September 2016.
12. Telia. Verdens Storste Iot Pilot. 2018. Available online: https://www.telia.no/magasinet/verdens-storste-iot-pilot/ (accessed on 23 December 2018).
13. U Blox. Press Releases. 2019. Available online: https://www.u-blox.com/en/press-releases/ (accessed on 26 January 2019).
14. Norway, T. Sub-Pump. 2018. Available online: https://www.teliacompany.com/en/news/news-articles/2017/sub-pump/ (accessed on 26 December 2018).
15. T Mobile. Smart Cities. 2019. Available online: https://iot.t-mobile.com/solutions/smart-cities/ (accessed on 10 February 2019).
16. Skyworks. Cellular IoT. 2019. Available online: http://www.skyworksinc.com/Products/1152/Cellular_IoT (accessed on 15 February 2019).
17. Mediatek. NB-IoT. 2019. Available online: https://www.mediatek.com/products/nbIot/mt2625 (accessed on 19 February 2019).
18. Huawei. 2019. Available online: https://e.huawei.com/my/solutions/technical/iot/nb-iot (accessed on 20 February 2019).
19. Quactel. LPWA IoT Module. 2019. Available online: https://www.quectel.com/product/list/LPWAIoTModule.htm (accessed on 15 March 2019).
20. Nordic Semiconductors. 2019. Available online: https://www.nordicsemi.com/News/2018/12/ (accessed on 10 February 2019).
21. Intel. Modem Solutions. 2019. Available online: https://www.intel.com/content/www/us/en/mobile/modem-solutions.html (accessed on 10 February 2019).
22. Sequans. 2019. Available online: http://www.sequans.com/products-solutions/ (accessed on 10 February 2019).
23. Qualcomm. IoT Modem. 2019. Available online: https://www.qualcomm.com/products/mdm9206-iot-modem (accessed on 10 February 2019).
24. Sierrawireless. Products and Solutions. 2019. Available online: https://www.sierrawireless.com/products-and-solutions/embedded-solutions/ (accessed on 10 February 2019).
25. Samsung. Samsung Exynos. 2019. Available online: https://news.samsung.com/global/samsungs-exynos-i-s111 (accessed on 10 February 2019).
26. Altair Semiconductors. 2018. Available online: https://altair-semi.com/ (accessed on 10 February 2019).
27. U blox. Cellular Modules. 2019. Available online: https://www.u-blox.com/en/cellular-modules (accessed on 10 February 2019).
28. Chen, M.; Miao, Y.; Hao, Y.; Hwang, K. Narrow Band Internet of Things. *IEEE Access* **2017**, *5*, 20557–20577. [CrossRef]
29. Hoglund, A.; Lin, X.; Liberg, O.; Behravan, A.; Yavuz, E.A.; Van Der Zee, M.; Sui, Y.; Tirronen, T.; Ratilainen, A.; Eriksson, D. Overview of 3GPP Release 14 Enhanced NB-IoT. *IEEE Netw.* **2017**, *31*, 16–22. [CrossRef]
30. Ratasuk, R.; Mangalvedhe, N.; Xiong, Z.; Robert, M.; Bhatoolaul, D. Enhancements of narrowband IoT in 3GPP Rel-14 and Rel-15. In Proceedings of the IEEE Conference on Standards for Communications and Networking (CSCN), Helsinki, Finland, 18–20 September 2017.
31. Boisguene, R.; Tseng, S.; Huang, C.; Lin, P. A survey on NB-IoT downlink scheduling: Issues and potential solutions. In Proceedings of the 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 26–30 June 2017.
32. Xu, J.; Yao, J.; Wang, L.; Ming, Z.; Wu, K.; Chen, L. Narrowband Internet of Things: Evolutions, Technologies, and Open Issues. *IEEE Internet Things J.* **2018**, *5*, 1449–1462. [CrossRef]

33. Feltrin, L.; Tsoukaneri, G.; Condoluci, M.; Buratti, C.; Mahmoodi, T.; Dohler, M.; Verdone, R. Narrowband IoT: A Survey on Downlink and Uplink Perspectives. *IEEE Wirel. Commun.* **2019**, *26*, 78–86. [CrossRef]
34. Zayas, A.D.; Merino, P. The 3GPP NB-IoT system architecture for the Internet of Things. In Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops), Paris, France, 21–25 May 2017.
35. ERICSSON. Scope of Release 16-NB-IoT, RP-181187. 2018. Available online: http://www.3gpp.org/ftp/TSG_RAN/TSG_RAN/TSGR_80/Report/ (accessed on 23 December 2018).
36. Huawei. On NB-IoT Evolution in Rel-16, RP-180877. 2018. Available online: http://www.3gpp.org/ftp/TSG_RAN/TSG_RAN/TSGR_80/Report/ (accessed on 23 December 2018).
37. 3GPP. Revision of WI on Enhancements of NB-IoT, RP-161901. 2018. Available online: https://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view&contributionId=730352 (accessed on 23 December 2018).
38. 3GPP. TS 36.300 V13.8.0 (2017-06). 2018. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2430 (accessed on 23 December 2018).
39. Xu, T.; Darwazeh, I. Non-Orthogonal Narrowband Internet of Things: A Design for Saving Bandwidth and Doubling the Number of Connected Devices. *IEEE Internet Things J.* **2018**, *5*, 2120–2129. [CrossRef]
40. Beyene, Y.D.; Jantti, R.; Tirkkonen, O.; Ruttik, K.; Iraji, S.; Larmo, A.; Tirronen, T.; Torsner, J. NB-IoT Technology Overview and Experience from Cloud-RAN Implementation. *IEEE Wirel. Commun.* **2017**, *24*, 26–32. [CrossRef]
41. Ratasuk, R.; Mangalvedhe, N.; Kaikkonen, J.; Robert, M. Data Channel Design and Performance for LTE Narrowband IoT. In Proceedings of the IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, Canada, 18–21 September 2016 .
42. Ratasuk, R.; Mangalvedhe, N.; Zhang, Y.; Robert, M.; Koskinen, J. Overview of narrowband IoT in LTE Rel-13. In Proceedings of the IEEE Conference on Standards for Communications and Networking (CSCN), Berlin, Germany, 31 October–2 November 2016.
43. 3GPP. Specs. 2019. Available online: http://www.3gpp.org/ftp/Specs/archive/45_series/45.820/45820-d10.zip (accessed on 24 April 2019).
44. 3GPP. TS 36.213. 2018. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2427 (accessed on 29 December 2018).
45. Ericsson. Ground-Breaking Long-Range NB-IoT Connection. 2018. Available online: https://www.ericsson.com/en/press-releases/2018/9/ericsson-and-telstra-complete-ground-breaking-long-range-nb-iot-connection, (accessed on 25 December 2018).
46. Ericsson. Key Technology Choices for Optimal Massive IoT Devices. 2019. Available online: https://www.ericsson.com/en/ericsson-technology-review/archive/2019/key-technology-choices-for-optimal-massive-iot-devices (accessed on 20 March 2019).
47. Ericsson. Technology Overview. 2019. Available online: https://www.ericsson.com/en/ericsson-technology-review/archive/2019/ (accessed on 10 April 2019).
48. Wang, Y.E.; Lin, X.; Adhikary, A.; Grovlen, A.; Sui, Y.; Blankenship, Y.; Bergman, J.; Razaghi, H.S. A Primer on 3GPP Narrowband Internet of Things. *arXiv* **2016**, arXiv:1606.04171.
49. Kroll, H.; Korb, M.; Weber, B.; Willi, S.; Huang, Q. Maximum-Likelihood Detection for Energy-Efficient Timing Acquisition in NB-IoT. In Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW), San Francisco, CA, USA, 19–22 March 2017.
50. Ali, A.; Hamouda, W. On the Cell Search and Initial Synchronization for NB-IoT LTE Systems. *IEEE Commun. Lett.* **2017**, *21*, 1843–1846. [CrossRef]
51. Xu, T.; Darwazeh, I. Uplink Narrowband IoT Data Rate Improvement: Dense Modulation Formats or Non-Orthogonal Signal Waveforms? In Proceedings of the 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, Italy, 9–12 September 2018; pp. 142–146. [CrossRef]
52. Zou, J.; Xu, C. Frequency Offset Tolerant Synchronization Signal Design in NB-IoT. *Sensors* **2018**, *18*, 4077. [CrossRef]
53. Lin, X.; Adhikary, A.; Eric Wang, Y. Random Access Preamble Design and Detection for 3GPP Narrowband IoT Systems. *IEEE Wirel. Commun. Lett.* **2016**, *5*, 640–643. [CrossRef]

54. Cho, S.; Kim, H.; Jo, G. Determination of Optimum Threshold Values for NPRACH Preamble Detection in NB-IoT Systems. In Proceedings of the 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN), Prague, Czech Republic, 3–6 July 2018; pp. 616–618.
55. Feltrin, L.; Condoluci, M.; Mahmoodi, T.; Dohler, M.; Verdone, R. NB-IoT: Performance Estimation and Optimal Configuration. In Proceedings of the European Wireless 24th European Wireless Conference, Catania, Italy, 2–4 May 2018.
56. Shimura, A.; Sawahashi, M.; Nagata, S.; Kishiyama, Y. Physical Cell ID Detection Performance Applying Frequency Diversity Reception to NPSS and NSSS for NB-IoT. In Proceedings of the 24th Asia-Pacific Conference on Communications (APCC), Ningbo, China, 12–14 November 2018.
57. Martín, A.G.; Leal, R.P.; Armada, A.G.; Durán, A.F. NBIoT Random Access Procedure: System Simulation and Performance. In Proceedings of the Global Information Infrastructure and Networking Symposium (GIIS), Thessaloniki, Greece, 23–25 October 2018.
58. Baracat, G.H.; Brito, J.M.C. NB-IoT Random Access Procedure Analysis. In Proceedings of the 2018 IEEE 10th Latin-American Conference on Communications (LATINCOM), Guadalajara, Mexico, 14–16 November 2018; pp. 1–6. [CrossRef]
59. M, Y.; Cheng, Y.; Hossain, J.; Ahmed, M.S.G. RADB: Random Access with Differentiated Barring for Latency-Constrained Applications in NB-IoT Network. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 6210408.
60. Jeon, W.S.; Seo, S.B.; Jeong, D.G. Effective Frequency Hopping Pattern for ToA Estimation in NB-IoT Random Access. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10150–10154. [CrossRef]
61. Li, Y.; Chen, S.; Ye, W.; Lin, F. A Joint Low-Power Cell Search and Frequency Tracking Scheme in NB-IoT Systems for Green Internet of Things. *Sensors* **2018**, *18*, 3274. [CrossRef] [PubMed]
62. Ha, S.; Seo, H.; Moon, Y.; Lee, D.; Jeong, J. A Novel Solution for NB-IoT Cell Coverage Expansion. In Proceedings of the Global Internet of Things Summit (GIoTS), Bilbao, Spain, 4–7 June 2018.
63. Chung, H.; Lee, S.; Jeong, J. NB-IoT Optimization on Paging MCS and Coverage Level. In Proceedings of the 15th International Symposium on Wireless Communication Systems (ISWCS), Lisbon, Portugal, 28–31 August 2018; pp. 1–5.
64. Živic, N. Improved Up-Link Repetition Procedure for Narrow Band Internet of Things. In Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 13–15 December 2018.
65. Zhang, L.; Ijaz, A.; Xiao, P.; Tafazolli, R. Channel Equalization and Interference Analysis for Uplink Narrowband Internet of Things (NB-IoT). *IEEE Commun. Lett.* **2017**, *21*, 2206–2209. [CrossRef]
66. Yang, B.; Zhang, L.; Qiao, D.; zhao, G.; Imran, M. Narrowband Internet of Things (NB-IoT) and LTE Systems Co-existence Analysis. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, UAE, 9–13 December 2018.
67. Rusek, F.; Hu, S. Sequential channel estimation in the presence of random phase noise in NB-IoT systems. In Proceedings of the IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017.
68. Mangalvedhe, N.; Ratasuk, R.; Ghosh, A. NB-IoT deployment study for low power wide area cellular IoT. In Proceedings of the IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, 4–8 September 2016.
69. Malik, H.; Pervaiz, H.; Mahtab Alam, M.; Le Moullec, Y.; Kuusik, A.; Ali Imran, M. Radio Resource Management Scheme in NB-IoT Systems. *IEEE Access* **2018**, *6*, 15051–15064. [CrossRef]
70. Yu, Y.; Tseng, S. Downlink Scheduling for Narrowband Internet of Things (NB-IoT) Systems. In Proceedings of the IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal, 3–6 June 2018.
71. Liu, J.; Mu, Q.; Liu, L.; Chen, L. Investigation about the paging resource allocation in NB-IoT. In Proceedings of the 20th International Symposium on Wireless Personal Multimedia Communications (WPMC), Bali, Indonesia, 17–20 December 2017.
72. Kim, H.; Cho, S.C.; Lee, Y. Interference Analysis of Guardband NB-IoT System. In Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC), Sydney, NSW, Australia, 4–7 June 2017.
73. Hsieh, B.; Chao, Y.; Cheng, R.; Nikaein, N. Design of a UE-specific uplink scheduler for narrowband Internet-of-Things (NB-IoT) systems. In Proceedings of the 3rd International Conference on Intelligent Green Building and Smart Grid (IGBSG), Yi-Lan, Taiwan, 22–25 April 2018.

74. Andres-Maldonado, P.; Ameigeiras, P.; Prados, J.; Ramos, J.; Navarro-Ortiz, J.M.; Lopez-Soler, J. Analytic Analysis of Narrowband IoT Coverage Enhancement Approaches. In Proceedings of the 2018 Global Internet of Things Summit (GIoTS), Bilbao, Spain, 4–7 June 2018; pp. 1–6.
75. Yu, C.; Yu, L.; Wu, Y.; He, Y.; Lu, Q. Uplink Scheduling and Link Adaptation for Narrowband Internet of Things Systems. *IEEE Access* **2017**, *5*, 1724–1734. [CrossRef]
76. Jiang, Z.; Han, B.; Chen, P.; Yang, F.; Bi, Q. On Novel Access and Scheduling Schemes for IoT Communications. *Mob. Inf. Syst.* **2016**, *2016*, 1–9. [CrossRef]
77. Oh, S.; Shin, J. An Efficient Small Data Transmission Scheme in the 3GPP NB-IoT System. *IEEE Commun. Lett.* **2017**, *21*, 660–663. [CrossRef]
78. Cluzel, S.; Franck, L.; Radzik, J.; Cazalens, S.; Dervin, M.; Baudoin, C.; Dragomirescu, D. 3GPP NB-IOT Coverage Extension Using LEO Satellites. In Proceedings of the IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal, 3–6 June 2018.
79. Lauridsen, M.; Krigslund, R.; Rohr, M.; Madueno, G. An Empirical NB-IoT Power Consumption Model for Battery Lifetime Estimation. In Proceedings of the IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal, 3–6 June 2018.
80. Lee, J.; Lee, J. Prediction-Based Energy Saving Mechanism in 3GPP NB-IoT Networks. *Sensors* **2017**, *17*, 2008. [CrossRef] [PubMed]
81. Bello, H.; Jian, X.; Wei, Y.; Chen, M. Energy-Delay Evaluation and Optimization for NB-IoT PSM with Periodic Uplink Reporting. *IEEE Access* **2019**, *7*, 3074–3081. [CrossRef]
82. Khan, S.; Malik, H.; Alam, M.; Le Moullec, Y. DORM: Narrowband IoT Development Platform and Indoor Deployment Coverage Analysis. In Proceedings of the 2nd International Workshop on Recent Advances in Cellular Technologies and 5G for IoT Environments (RACT-5G-IoT 2019), Leuven, Belgium, 29 April–2 May 2019.
83. Kim, T.; Kim, D.M.; Pratas, N.; Popovski, P.; Sung, D.K. An Enhanced Access Reservation Protocol With a Partial Preamble Transmission Mechanism in NB-IoT Systems. *IEEE Commun. Lett.* **2017**, *21*, 2270–2273. [CrossRef]
84. 3GPP. Architecture Enhancements to Facilitate Communications with Packet Data Networks and Applications. 2017. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=862 (accessed on 20 April 2019).
85. 3GPP. TS 23.401 V14.3.0. 2017. Available online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=849 (accessed on 25 April 2019).
86. Andres-Maldonado, P.; Ameigeiras, P.; Prados-Garzon, J.; Navarro-Ortiz, J.; Lopez-Soler, J.M. Narrowband IoT Data Transmission Procedures for Massive Machine-Type Communications. *IEEE Netw.* **2017**, *31*, 8–15. [CrossRef]
87. Mohamed, M.O.; Abdelhamid, B.; El Ramly, S. Interference mitigation in heterogeneous networks using Fractional Frequency Reuse. In Proceedings of the International Conference on Wireless Networks and Mobile Communications (WINCOM), Fez, Morocco, 26–29 October 2016.
88. Monteiro, N.; Mihovska, A.; Rodrigues, A.; Prasad, N.; Prasad, R. Interference analysis in a LTE-A HetNet scenario: Coordination vs. uncoordination. In Proceedings of the Wireless VITAE 2013, Atlantic City, NJ, USA, 24–27 June 2013.
89. Palanisamy, P.; Nirmala, S. Downlink interference management in femtocell networks—A comprehensive study and survey. In Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 21–22 Feburary 2013.
90. Song, S.; Li, H.; Fan, Y.; Kong, W.; Zhang, W. Downlink Interference Rejection in Ultra Dense Network. In Proceedings of the 13th APCA International Conference on Control and Soft Computing (CONTROLO), Ponta Delgada, Portugal, 4–6 June 2018; pp. 361–364.
91. Chiumento, A.; Pollin, S.; Desset, C.; der Perre, L.V.; Lauwereins, R. Scalable HetNet interference management and the impact of limited channel state information. *EURASIP J. Wirel. Commun. Netw.* **2015**, *2015*, 74. [CrossRef]
92. Oo, T.; Tran, N.H.; Saad, W.; Niyato, D.; Han, Z.; Hong, C. Offloading in HetNet: A Coordination of Interference Mitigation, User Association, and Resource Allocation. *IEEE Trans. Mob. Comput.* **2017**, *16*, 2276–2291. [CrossRef]
93. Liou, R.H.; Lin, Y.B.; Tsai, S.C. An Investigation on LTE Mobility Management. *IEEE Trans. Mob. Comput.* **2013**, *12*, 166–176. [CrossRef]

94. Karandikar, A.; Akhtar, N.; Mehta, M. Mobility Management in LTE Networks. In *Mobility Management in LTE Heterogeneous Networks*; Springer: Berlin, Germany, 2017.
95. Francois, F.; Abdelrahman, O.H.; Gelenbe, E. Impact of Signaling Storms on Energy Consumption and Latency of LTE User Equipment. In Proceedings of the 2015 IEEE 17th International Conference on High Performance Computing and Communications, New York, NY, USA, 24–26 August 2015.
96. Koc, A.T.; Jha, S.C.; Vannithamby, R.; Torlak, M. Device Power Saving and Latency Optimization in LTE-A Networks Through DRX Configuration. *IEEE Trans. Wirel. Comm.* **2014**, *13*, 2614–2625.
97. Nikaein, N.; Krea, S. Latency for Real-Time Machine-to-Machine Communication in LTE-Based System Architecture. In Proceedings of the 17th European Wireless 2011—Sustainable Wireless Technologies, Vienna, Austria, 27–29 April 2011.
98. Laya, A.; Alonso, L.; Alonso-Zarate, J. Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives. *IEEE Commun. Surv. Tutor.* **2013**, *16*, 4–16. [CrossRef]
99. Cheng, J.; Lee, C.; Lin, T. Prioritized Random Access with dynamic access barring for RAN overload in 3GPP LTE-A networks. In Proceedings of the 2011 IEEE GLOBECOM Workshops (GC Wkshps), Houston, TX, USA, 5–9 December 2011.

Appendix 2

C.B. Mwakwata, M.M. Alam, Y. Le Moullec, H. Malik, S. Pärand. Cooperative Interference Avoidance Scheduler for Radio Resource Management in NB-IoT Systems. European Conference on Networks and Communications (EUCNC), IEEE, 2020.

Cooperative Interference Avoidance Scheduler for Radio Resource Management in NB-IoT Systems

Collins Burton Mwakwata, Muhammad Mahtab Alam , Yannick Le Moullec, Hassan Malik, Sven Parand*
Thomas Johann Seebeck Department of Electronics
Tallinn University of Technology
**Telia Estonia*
Estonia
Email: (collins.burton, muhammad.alam, yannick.lemoullec, hassan.malik)@taltech.ee
(sven.parand)@telia.ee

***Abstract*—The design changes on the physical (PHY) layer, i.e. the limited system bandwidth of one physical resource block (PRB), single antenna support, lower-order modulations, etc. inhibit the mapping of traditional long term evolution (LTE) radio resource management techniques to narrowband internet of things (NB-IoT) systems. Consequently, possible interference due to massive connectivity may severely degrade the expected system performance.**

In this regard, we propose an interference avoidance scheduling algorithm for NB-IoT systems. The algorithm entails a cooperative strategy in which the base stations share their respective scheduling tables which are then used to compute the interference for future transmitting user equipment (UEs). The computed interference values are then used as input to individual base station schedulers to perform scheduling. Each base station's scheduler then allocates the radio resources to the UEs with the lowest possible interference.

Extensive simulations are carried out to analyze the performance of our proposed algorithm and compare it to the conventional Round-Robin scheduling scheme. The results show that our proposed algorithm provides up to 36% throughput improvement to the NB-IoT UE as compared to Round-Robin. Similarly, for the same device's locations, the UEs are experiencing relatively better maximum coupling loss (MCL) which results in lower repetition numbers per coverage class.

***Index Terms*—NB-IoT, LPWAN, Interference Avoidance, Resource Scheduler, Radio Resource Management, mMTC**

I. Introduction

Narrow-Band Internet of Things (NB-IoT) is one of the enablers of the IoT use-cases such as environmental monitoring, smart gas metering, smart grids, smart water metering, smart waste management, etc. These use-cases are normally associated with different sensors installed in a fixed location to facilitate the corresponding expected services with low or no human intervention. The diversity of use-cases enforces different UE requirements including reporting circle, energy efficiency, payload size, nature of traffic, etc. For example, in a small-medium sized city, it is expected to have an average of 40 devices per household equipped with sensors to service different applications [1]. In this regard, such massive connections demand proactive radio resource management techniques to offer the required quality of service to the corresponding applications.

To cope with the growing demand for IoT use-cases, the 3rd Generation Partnership Project (3GPP) introduced the NB-IoT as a licensed IoT cellular technology to support the massive Machine-Type connections (mMTC). NB-IoT is a variant of the Long Term Evolution (LTE) with reduced complexity to enable low-cost devices. NB-IoT is classified as a Low-Power Wide-Area Networks (LPWAN) technology intended to enhance coverage for IoT use-cases especially for applications in hard-to-reach areas.

As per 3GPP, the design changes are as follows; NB-IoT's system bandwidth is a maximum of 200 kHz and can offer peak data rates of 250 $kbps$ and 226.7 $kbps$ in downlink and uplink, respectively. NB-IoT performs up to 128 repetitions to enable extended coverage, can tolerate a delay of up to 10 s during its transmissions, utilizes low order modulations i.e. Binary Phase Shift Keying (BPSK), support up to 52000 devices per cell and can transmit at maximum coupling loss (MCL) of 164 dB. More design changes are presented in [2].

Despite the advantages they bring, these changes hinder the effective implementation of traditional radio resource management techniques to NB-IoT systems because i) supporting massive devices with merely 200 kHz of radio resources is challenging even with low payload transmission, and additionally ii) when NB-IoT is deployed in in-band, guard-band or stand-alone mode, the limited system bandwidth hinders the traditional techniques such as carrier aggregation, inter-cell interference coordination (ICIC) with physical resource block (PRB) muting, etc. This is because PRB muting means shutting down the complete NB-IoT system bandwidth, and ICIC involves power and frequency partitioning between the competing/interfering base stations. Similarly, the Network Assisted Interference Cancellation and Suppression (NAICS) technique is no longer applicable. With NAICS, the network is required to provide the UEs with additional information on scheduled transmissions and hence enhance the performance of the receiver; however, NAICS is not a spectrum efficient technique and requires a complex receiver for interference mitigation [3]. Therefore, it is necessary to propose novel radio resource management techniques adapted to NB-IoT systems.

Previous studies such as [4] proposed an interference mitigation algorithm by considering the coexistence between NB-IoT

978-1-7281-4355-2/20/$31.00 ©2020 IEEE

and LTE; the proposed algorithm uses the channel frequency response (CFR) to mitigate the sampling mismatch between the NB-IoT UE and the base station / enhanced Node-B (eNB). The authors of [5] proposed an algorithm that leverages the average device delay and processing time to optimize NB-IoT resource management through basic scheduling. In [6], the authors presented a novel interference aware resource management for NB-IoT. Power control in a cooperative manner was proposed to minimize the interference impact. In [7], the authors proposed an algorithm to enhance the cell capacity; the optimal scheduling involves offset index selection and UE search spaces in NB-IoT networks. In [8], the authors proposed a link adaptation algorithm to enhance coverage; the algorithm uses a mathematical analysis of Shannon theorem. In [9], the authors proposed a link adaptation scheme that dynamically adjusts the maximum coupling loss (MCS) to optimize the block error rate (BLER) and Bit Error Rate performance.

However, the existing state of the art does not consider yet the inter-cell interference avoidance strategies to cope with the growing demand of IoT use-cases under stringent radio resources.

In this regard, we propose an inter-cell interference-avoidance scheduling algorithm to optimize the usage of limited NB-IoT radio resources and enhance the overall throughput. Our algorithm relies on the utilization of a cooperative strategy between NB-IoT's base stations by sharing the expected scheduling tables among themselves in advance. Each base station then computes the possible inter-cell interference that may arise between the prospective devices during their transmissions. The tables along with the computed interference information are then used as input to local schedulers. Proactive scheduling is then performed, avoiding to assign devices whose interference impact can lower the expected throughput. A Round-Robin algorithm is used as a benchmark and exhaustive simulations are performed. The results show that our proposed algorithm enhances the device achievable throughput up to 36% and reduces the number of repetitions per coverage class.

To the best of the authors' knowledge, this is the first work that entails the cooperative interference avoidance strategy to optimize the radio resources usage for NB-IoT systems.

The rest of the paper is organized as follows; section II presents our proposed cooperative interference-avoidance scheme, section III presents the simulation and performance evaluation, and section IV concludes the paper.

II. Proposed Cooperative Interference-Avoidance Strategy for NB-IoT System

In our proposed cooperative strategy the base stations share (over the X2-interface) the channel quality information (CQI) (i.e. SNR, location, path-loss, cell ID, expected payload, etc.), for the devices (UE) to be scheduled for the next radio frame. The proposed scheduler then uses this information to calculate the interference possibilities among shared UEs. The calculated interference values are hence used as input to the individual base station. Proactive scheduling is then performed

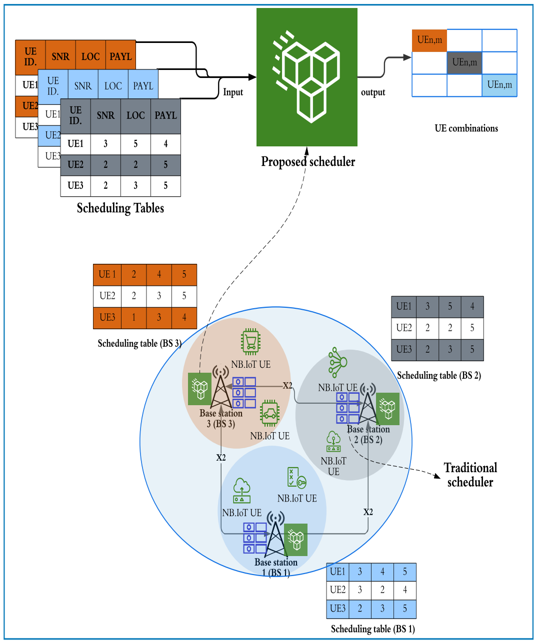

Fig. 1. Proposed cooperative interference-avoidance strategy for NB-IoT system

by providing the available resources to UEs whose impact in terms of inter-cell interference is the lowest.

The proposed scheduler comprises three main parts i.e. collection center, computing center, and scheduling center. The collection center receives and registers the scheduling tables from the individual base-stations. The computing center computes the inter-cell interference between UEs. The scheduling center makes the final decision about the UEs that have the best throughput performance when scheduled in the same slots. The system under study is considered to be a small and medium sized city based on the Okumura-Hata channel model whereby the UE path-loss model can then be expressed as in Equation (1), [10].

$$PL = A + B\ log(d) + C \tag{1}$$

where A, B and C depend on the antenna height and the frequency.

$$A = 69.55 + 26.16\ log(fc) - 13.82\ log(hb) - a(hm) \tag{2}$$

$$B = 44.9 - 6.55\ log(hb) \tag{3}$$

where fc and d are given in MHz and km, respectively, a and C depend on environmental factors, and hb and hm are heights for the base-station and UE, respectively.

The interference impact is based on the SINR which is calculated as shown in Equation (4), [11].

$$SINR_{k,n}^{DL} = \frac{p_{k,n}|h_{k,n}|^2}{\sum_{m=1}^{M} \omega_{n,m} p_{m,n} |h_{m,n}^C|^2 + N_0 B} \tag{4}$$

where $SINR_{k,n}^{DL}$, k, $p_{k,n}$, and $|h_{k,n}|^2$ are the down-link signal to interference-plus-noise ratio, the transmit power, and channel response of user n from base station k, respectively. $\omega_{n,m}$, $p_{m,n}$, and $|h_{m,n}^C|^2$ are the power classes, transmit power, and channel response of user m from base station n, respectively. $N_0 B$ is the channel noise which is considered constant.

The proposed scheduler functions as follows: it receives the scheduling tables from the individual base stations and then checks for inter-cell interference; if there is interference, it checks the interference weight with all the other UEs to be scheduled in the same radio frame. If the UEs have the best throughput performance, it forwards that combination (UE identities) to be used by the individual base stations. The flowchart of the proposed inter-cell interference-avoidance algorithm for the NB-IoT system is as depicted in Fig. 2.

III. Simulation and Performance Evaluation

TABLE I
MAIN SIMULATION PARAMETERS FOR THE PROPOSED COOPERATIVE STRATEGY FOR NB-IOT SYSTEM [12]

Simulation Parameters	
Name	Value
(a) Transmit power of base station, UE (dBm)	46 , 23
(b) Modulation scheme	BPSK
(c) Carrier frequency (MHz)	900
(d) Receiver Thermal Noise density (dBm/Hz)	-174
(e) No. cooperating base station	3
(f) Interference Margin (dB)	0
(g) Channel model	Okumura Hata
(h) Effective Noise Power (dBm)	d + q + f + 10log(r)
(i) Required / calculated SINR (dB)	
(j) Receiver sensitivity	h + i
(k) MCL (dB)	a - j
(l) Modulation scheme	BPSK
(m) No. of antenna support per UE	1
(o) Height of base station, UE (m)	100, 1
(p) Radius of a cell	1 km
(q) Noise figure of base station, UE	9, 5 dB
(r) Occupied System bandwidth (kHz)	180

1) Simulation Setup: Extensive system-level simulations are performed to analyze the proposed approach, as depicted in Fig.1. The simulation setup is considered close to the one presented in [12]; however, it is well adapted to fit the NB-IoT system. The NB-IoT UEs are considered fixed and hence the impact of Doppler spread on UE mobility is negligible. This suits well the use cases such as smart grid, smart water, smart gas metering, smart waste management, etc. [1].

The Round-Robin algorithm, as presented in [13], is used to compare the performance of the proposed algorithm. In Round-Robin, each eNB assigns the radio resources to UEs in a first-come-first-served way. That is, the first detected UE

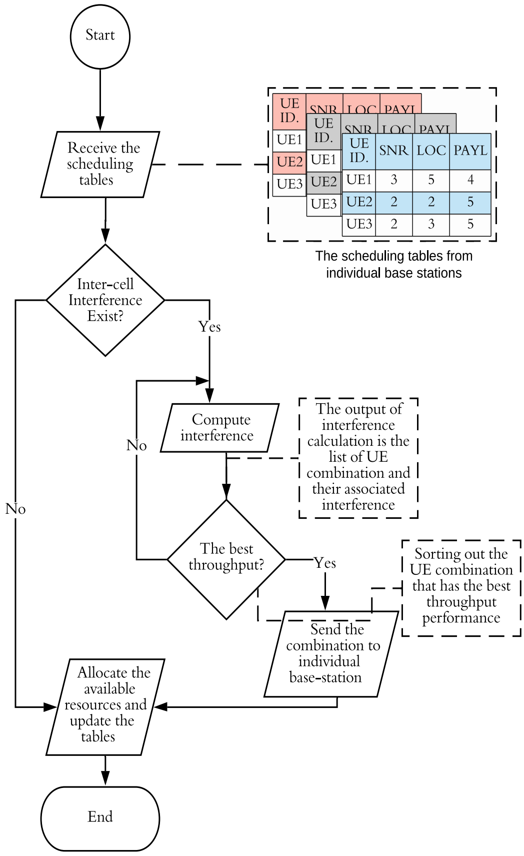

Fig. 2. Flowchart of our proposed NB-IoT inter-cell interference avoidance algorithm

is given the available resources regardless of the impact of interference it may cause/face.

For the interference-aware scheduling algorithm to better function, it is crucial to perform proper estimation of channel parameters to effectively sort out the UEs under possible interference. This is achieved by fixing the base stations and UEs while evaluating their corresponding channel conditions based on the defined model. The important simulation parameters are as presented in Table I [14].

2) Simulation Results: In Fig. 3, it can be observed that with the Round-Robin approach, the maximum achievable throughput per device is 60 $kbps$. The devices located at the center of the cell experience better throughput as compared to those towards the edge. This is due to the low path-loss which increases when moving from the center. It is also observed that the devices on the cell edge experience relatively high inter-cell interference as compared to those located at the center. Consequently, the achieved throughput is very low on the

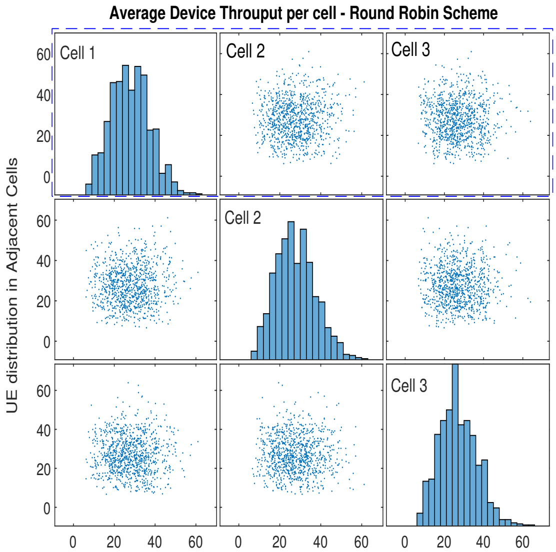

Fig. 3. Throughput distribution of NB-IoT UEs from the evaluation of the three cells when the Round-Robin scheduling algorithm is used. For example, when evaluating cell 1 the interference impact is considered from UEs from cell 2 and cell 3 and the same method is done for all cells. The dots represent the UEs distribution in adjacent cells

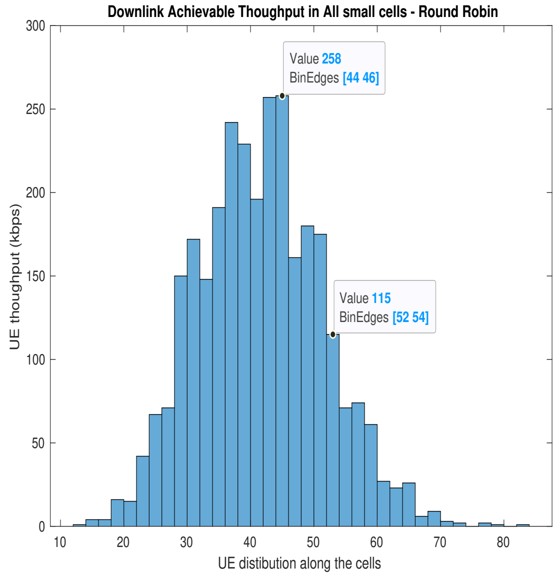

Fig. 5. Total achievable throughput for NB-IoT in the cells after using the Round-Robin scheduling algorithm

Fig. 4. Throughput distribution of NB-IoT UEs from the evaluation of the three cells when the proposed interference-aware scheduling algorithm is used. The dots represent the UEs distribution

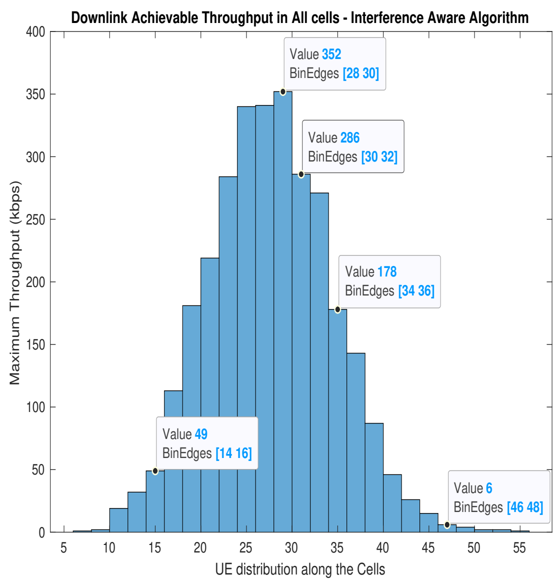

Fig. 6. Total achievable throughput for NB-IoT in the cells after using the proposed interference-avoidance scheduling algorithm

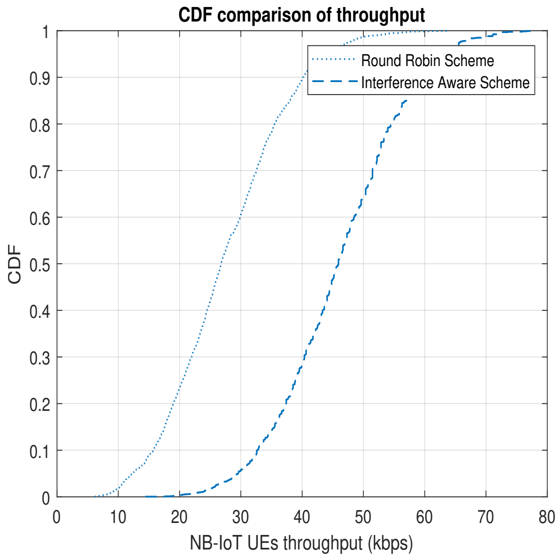

Fig. 7. Cumulative distribution functions (CDF) that represents the gain in throughput of the proposed intercell interference avoidance scheduling algorithm as compared to the Round Robin scheduling algorithm

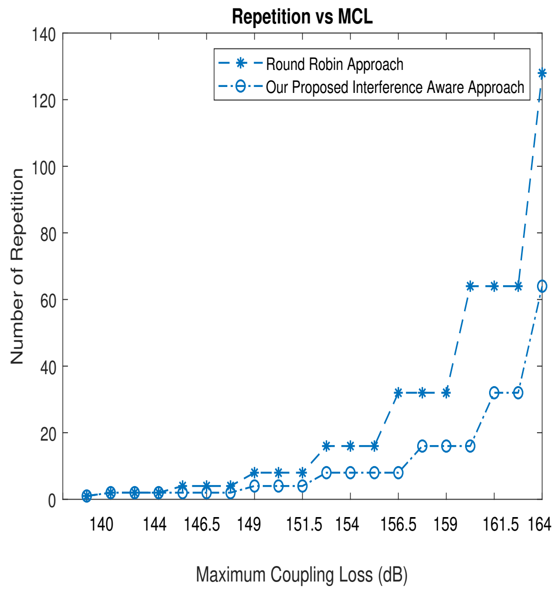

Fig. 9. Maximum number of repetitions experienced by the NB-IoT UEs under the two approaches

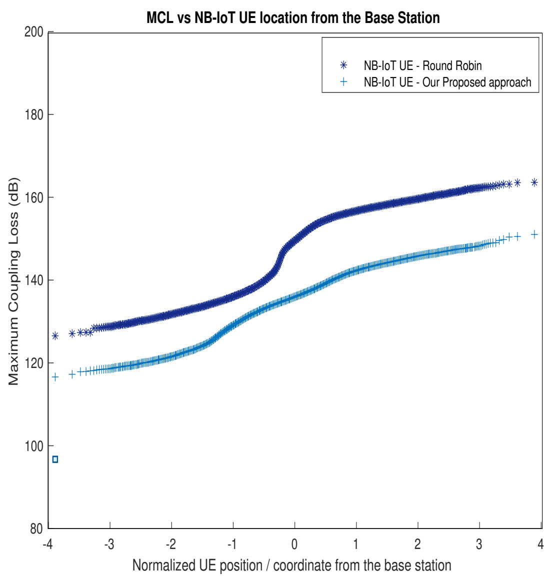

Fig. 8. Comparison of the maximum coupling loss between our proposed approach and Round-Robin for the UEs served in the same fixed locations.

edges of the cell.

In Fig. 4, with our proposed approach, it is observed that the overall UE throughput has relatively raised and the devices located at the center of the cell can experience a throughput of up to 80 $kbps$. Similarly, as the devices get farther from the center, the throughput drops; this is due to the increase in the experienced path-loss. The overall relative increase in throughput as compared to the Round-Robin algorithm is due to avoidance of interference and hence the NB-IoT UEs could guarantee better transmission with minimum errors.

Figure 5 presents the NB-IoT UE's maximum achievable throughput over the three cells. It is observed that the UEs that are exactly at the center of the cell can achieve an average throughput of about 250 $kbps$. This peak value corresponds to the 3GPP proposed framework as presented in [14] which is used during system modeling. However, some UEs that are exactly at the center of the cell can experience relatively lower throughput compared to UEs at a distant location. This is due to the inter-cell interference that these UEs may experience when scheduled at the same time.

Figure 6 presents the NB-IoT UE's maximum achievable throughput over the three cells after implementing the interference-aware scheduling algorithm. It is observed that the maximum achievable total NB-IoT UE throughput over the three cells is 350 $kbps$, while gradually decreasing as UEs are located far from the center. (Please note that 3GPP has given 250 $kbps$, but in our simulation, we considered that the impact of control channel overhead is negligible). However, this peak value is only possible when the impact of interference is considered negligible. As moving farther from the cell center, the throughput uniformly drops due to shadowing, fading and increase in path-loss.

Figure 7 presents the comparison of the cumulative distribution function (CDF) of achievable throughput for the two approaches under study. For the same number of devices per cell, it can be observed that the proposed approach outerperforms the Round-Robin approach and about 50 percentile of devices are experiencing almost twice the minimum achievable throughput.

Figure 8 presents the comparisons of MCL between our proposed approach and Round-Robin, it is observed that for the same device locations under the three coverage classes (i.e. MCL = 144 dB, MCL = 158 dB and MCL = 164 dB) the devices are served with relatively lower MCL as compared to the Round Robin approach. This is due to reduced interference impact and hence improving the receiver sensitivity. As a result, our approach makes the devices experience a relatively lower number of repetitions in all coverage classes, as shown in Fig. 9.

In general, it is observed that significant system gains in throughput and repetition reduction per coverage class are achieved when our proposed algorithm is used.

IV. Conclusion

This paper has presented a novel inter-cell interference aware scheduling algorithm for NB-IoT systems. Our proposed scheme utilizes the information shared in advance between the cooperating base stations to compute the interference weight, and hence schedules UEs with the minimum possible interference for the next transmissions. Unlike previous studies, our scheme considers the NB-IoT devices reduced complexity which may lead to poor channel estimation. Exhaustive simulations are carried to analyze the performance improvements of our approach. It is observed that our approach could guarantee a throughput improvement up to 36% as compared to the conventional Round-Robin approach. Furthermore, maximum achievable total NB-IoT UE throughput over the three cells is around 350 $kbps$ for devices at the cell center; however, as the UE moves far from the center, the throughput relatively drops due to channel variations and inter-cell interference impact on cell edge UEs. It is also observed that our proposed approach serves UEs at the same positions with better MCL as compared to the conventional scheme hence can reduce the number of repetitions per coverage class. The future outlook involves further studies on dynamic traffic models, energy efficiency, latency and mobility analysis for NB-IoT systems.

Acknowledgment

This work was supported in part from the European Unions Horizon 2020 research and innovation program under grant agreement No 668995, European Union Regional Development Fund in the framework of the Tallinn University of Technology Development Program 2016-2022. Furthermore, the authors would like to thank Telia Estonia for their cooperation.

References

[1] C. B. Mwakwata, H. Malik, M. Mahtab Alam, Y. Le Moullec, S. Parand, and S. Mumtaz, “Narrowband internet of things (nb-iot): From physical (phy) and media access control (mac) layers perspectives,” *Sensors*, vol. 19, no. 11, 2019.

[2] Y. . E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, “A primer on 3gpp narrowband internet of things,” *IEEE Communications Magazine*, March 2017.

[3] J.-Y. Hwang and S.-W. Lee, “Advanced receiver for interference cancellation in 3gpp lte-a,” *The Journal of Korea Information and Communications Society*, vol. 39A, pp. 389–397, 07 2014.

[4] L. Zhang, A. Ijaz, P. Xiao, and R. Tafazolli, “Channel equalization and interference analysis for uplink narrowband internet of things (nb-iot),” *IEEE Communications Letters*, vol. PP, 05 2017.

[5] B. Hsieh, Y. Chao, R. Cheng, and N. Nikaein, “Design of a ue-specific uplink scheduler for narrowband internet-of-things (nb-iot) systems,” in *3rd International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, April 2018.

[6] H. Malik, H. Pervaiz, M. Mahtab Alam, Y. Le Moullec, A. Kuusik, and M. Ali Imran, “Radio resource management scheme in nb-iot systems,” *IEEE Access*, vol. 6, pp. 15051–15064, 2018.

[7] R. Boisguene, S. Tseng, C. Huang, and P. Lin, “A survey on nb-iot downlink scheduling: Issues and potential solutions,” in *13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, June 2017.

[8] P. Andres-Maldonado, P. Ameigeiras, J. Prados, J. Ramos, J. Navarro-Ortiz, and J. M. Lopez-Soler, “Analytic analysis of narrowband iot coverage enhancement approaches,” pp. 1–6, 06 2018.

[9] C. Yu, L. Yu, Y. Wu, Y. He, and Q. Lu, “Uplink scheduling and link adaptation for narrowband internet of things systems,” *IEEE Access*, 2017.

[10] Z. Nadir and M. Ahmad, “Pathloss determination using okumura-hata model and cubic regression for missing data for oman,” *Lecture Notes in Engineering and Computer Science*, vol. 2181, 03 2010.

[11] N. Tabia, A. Gondran, O. Baala, and A. Caminada, “Interference model and evaluation in lte networks,” in *2011 4th Joint IFIP Wireless and Mobile Networking Conference (WMNC 2011)*, pp. 1–6, Oct 2011.

[12] A. Müller and P. Frank, “Cooperative interference prediction for enhanced link adaptation in the 3gpp lte uplink,” in *2010 IEEE 71st Vehicular Technology Conference*, pp. 1–5, May 2010.

[13] N. Bechir, M. Nasreddine, A. Mahmoud, H. Walid, and M. Sofien, “Novel scheduling algorithm for 3gpp downlink lte cellular network,” *Procedia Computer Science*, vol. 40, pp. 116 – 122, 2014. Fourth International Conference on Selected Topics in Mobile and Wireless Networking.

[14] 3GPP, “3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) (Release 13) ,” Technical Specification (TS) 36.331, 3rd Generation Partnership Project (3GPP), 08 2016. V13.1.0 (2015-11).

Appendix 3

C.B. Mwakwata, O. Elgarhy, Y. Le Moullec, M.M. Alam, S. Pärand, I. Annus. Inter-cell Interference Reduction Scheme for Uplink Transmission in NB-IoT Systems. 2021 International Wireless Communications and Mobile Computing (IWCMC), 400-405. IEEE, 2021.

2021 International Wireless Communications and Mobile Computing (IWCMC) | 978-1-7281-8616-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/IWCMC51323.2021.9498970

Inter-cell Interference Reduction Scheme for Uplink Transmission in NB-IoT Systems

Collins Burton Mwakwata
Thomas Johann Seebeck of Electronics
Tallinn University of Technology
Tallinn, Estonia
collins.burton@taltech.ee

Osama Elgarhy
Thomas Johann Seebeck of Electronics
Tallinn University of Technology
Tallinn, Estonia
osama.elgarhy@taltech.ee

Yannick Le Moullec
Thomas Johann Seebeck of Electronics
Tallinn University of Technology
Tallinn, Estonia
yannick.lemoullec@taltech.ee

Muhammad Mahtab Alam
Thomas Johann Seebeck of Electronics
Tallinn University of Technology
Tallinn, Estonia
muhammad.alam@taltech.ee

Sven Pärand
Telia Estonia Ltd.
Tallinn, Estonia
sven.parand@telia.ee

Ivar Annus
Department of Civil Engineering and Architecture
Tallinn University of Technology
Tallinn, Estonia
ivar.annus@taltech.ee

***Abstract*—In this paper, we propose an inter-cell interference (ICI) minimization scheme for uplink transmission in the narrow-band internet of things (NB-IoT) systems. We first establish the theoretical ICI problem formulation and propose its corresponding solution for the orthogonal multiple access (OMA) NB-IoT system. Based on the theoretical formulation, we design a cooperative radio resource scheduler that reduces the impact of ICI and allocates transmit powers to reduce the energy consumption to the scheduled users in a multi-cell scenario. We compare the performance of the proposed scheme with that of some benchmark OMA schedulers. The results show that the proposed technique significantly reduces the impact of ICI and hence is more suitable for the massive connectivity of the NB-IoT system. For example, the users operating under the proposed approach experience up to 50% reduced energy consumption when compared to the best channel quality indicator (CQI) scheme. Furthermore, 30% and 35% improvements in terms of achieved user's data rates are obtained as compared to the MaxMin and round-robin schemes, respectively.**

***Index Terms*—NB-IoT, LPWAN, Inter-cell interference minimization, mMTC, Cellular IoT, ICI.**

I. Introduction

Massive machine-type communications (mMTC) is one of the fifth-generation (5G) service verticals which is designed to support high-density internet of things (IoT) connections [1], [2]. Even-though mMTC is enabled by licensed and unlicensed IoT technologies, in this work, our focus is on licensed technologies, specifically, Narrow-Band IoT (NB-IoT).

NB-IoT is derived from the long term evolution (LTE) technology; however, its system bandwidth is of a maximum of $200\ kHz$, with $15\ kHz$ or $3.75\ kHz$ sub-carrier spacing on the uplink. More details about NB-IoT system-level specifications, design changes, and standards are presented in [3].

Work supported by Tallinn University of Technology Development Program 2016-2022 (2014-2020.4.01.16-0032), Estonian Research Council (PRG667), European Union's Horizon 2020 Research and Innovation programme (951867 and 668995). The authors also thank Telia Estonia for their cooperation.

Due to its narrow system bandwidth and overall reduced complexity, it is very hard to map the traditional radio resources management (RRM) techniques to the NB-IoT system. For example, NB-IoT is expected to accommodate up to 52000 simultaneously transmitting devices per cell [4]. Therefore, the question is "how to effectively use this stringent available bandwidth to serve such dense networks and concurrently reduce inter-cell interference (ICI)?" Indeed, if inter-cell interference is not managed, severe degradation of the quality of service (QoS) for such dense deployments becomes inevitable [5], [6].

Simultaneously, the number of IoT use-cases grow rapidly, and the ongoing massive IoT deployment needs to operate for a very long time with as low power consumption as possible. In this regard, energy consumption becomes a critical issue. The NB-IoT standard provides discontinuous reception (eDRX) and power-saving mode (PSM) operations to save devices' energy. But, how to further reduce the transmission energy while ensuring the required quality of service is still an open issue. It should be noted that the focus of this work is on cooperative scheduling to minimize ICI; The interested reader can refer to [7]–[9] for different power control schemes that can be used in 5G systems.

Recent studies have addressed various related key issues; e.g. in [5] the authors studied the factors that affect cell data rate and proposed a radio resource allocation algorithm that takes into consideration the repetition factor for each user, time offset, and quality of service (QoS) constraints. In [10], the authors studied the impact of interference between NB-IoT and LTE in a coexisting network scenario; the presented analysis shows that the reduced complexity of NB-IoT user equipment (UE) makes them prone to carrier frequency offset, which significantly increases interference caused by radio frequency (RF) impairments.

In [11], the authors derived an uplink system model for the NB-IoT IoT system. Their results reveal that the actual channel frequency response (CFR) is not a simple Fourier transform of the channel impulse response, due to sampling rate mismatch between the NB-IoT user and LTE base station. Consequently, they proposed a new channel equalization algorithm by deriving the effective CFR. In addition, they analytically derived interference to facilitate the co-existence of NB-IoT and LTE signals.

The reduced complexity of the NB-IoT UEs necessitates the implementation of novel techniques to guarantee the required QoS. The above-proposed techniques are examples of improvements in the NB-IoT system. However, the impact of ICI is not well studied although it has proved to be the main cause of performance degradation in legacy and recent cellular technologies.

Therefore, in contrast to the above studies, the new contributions presented in this paper are:

- First, the theoretical ICI minimization problem is presented, and its corresponding solution is proposed.
- Second, a novel cooperative scheduler is proposed to reduce the ICI impact and hence improve the QoS.
- Third, power allocation is implemented for the scheduled users to reduce unnecessary energy consumption while guaranteeing the required QoS.

Finally, the proposed performance enhancements are evaluated and compared to proportional fair (PF), max-min, best channel quality indicator (Best CQI), and round-robin (RR) schedulers.

To the best of the authors' knowledge, this is the first work that proposes the ICI minimization scheme for the NB-IoT uplink system.

The rest of the paper is organized as follows: Section II presents the problem formulation and proposed solution. Section III presents the proposed cooperative scheduler and simulation setup. Section IV presents the numerical results and discussion. Finally, Section V concludes the paper.

II. Problem formulation and proposed solution

A. Problem formulation

The NB-IoT uplink system has four possible resource unit configurations to choose from, here we employ the single tone resource unit configuration mode of deployment (i.e. one tone per user); however, the analysis can be replicated for other resource configurations. The tone bandwidth is given by $B_0 = B/X$, where B is the available system bandwidth and X is the resource unit spacing (i.e. for NB-IoT uplink, $B = 180\ kHz$, and $X = 15\ kHz$ or $X = 3.75\ kHz$). The index set for the available resource units is denoted as $z = \{1, 2, \ldots, Z\}$. Let K_c be the set of users belonging to cell c, where the number of cells is C, i.e. $c = \{1, 2, \ldots, C\}$, and k is the user index; thus, a user k in cell c will be denoted as k_c. The achievable rate of user k belonging to cell c on a given resource unit z is denoted by $R^z_{k_c} = B_0 \log_2(1 + SINR^z_{k_c})$ where $SINR^z_{k_c}$ is the signal to interference plus noise ratio experienced by user k belonging to cell c on a given resource unit z, and is given as:

$$SINR^z_{k_c} = a^z_{k_c}\Big(\frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\sum_{l\neq c, l\in C}\sum_{j\in K_l}|h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} + P_n}\Big) \quad (1)$$

where $|h^z_{k_c,c}|$ denotes user k_c's channel gain on resource unit z to its own base station in cell c, and $P^z_{k_c}$ denotes user k_c's transmission power on resource unit z. In the denominator, the interference term comes from other cells l, with a group of users K_l within the cell. We use j as the interference user index, thus a user j in cell l will be denoted as j_l. Moreover, $|h^z_{j_l,c}|$ denotes the channel gain of an interfering user, belonging to cell l, j_l on resource unit z to the base station of cell c, and $P^z_{j_l}$ denotes the interfering user j_l's transmission power on resource unit z. Binary variables are used for scheduling: variable $a^z_{k_c}$ denotes the resource unit occupancy coefficient such that $a^z_{k_c} = 1$ if the resource unit z is used by user k_c, and $a^z_{k_c} = 0$ otherwise. P_n is the noise power at the receiver.

The optimization goal is to minimize the inter-cell interference experienced by user k from adjacent cell users. The problem can either be modeled as the interference experienced by user k on a given resource unit, or, in order to avoid using the two binary variables, the problem can be modeled as minimizing the interference on the resource units of cell c, which is a realistic assumption since we adopted the full buffer model, where I^z_c is the interference on resource unit z in cell c. The objective function can be expressed as:

$$\min \sum_{c\in C}\sum_{z\in Z} I^z_c \quad (2)$$

Substituting the interference I^z_c, the objective function becomes:

$$\min \sum_{c\in C}\sum_{z\in Z}\sum_{l\neq c, l\in C}\sum_{j\in K_l}|h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} \quad (3)$$

subject to:

$$SINR^z_{k_c} \geq \vartheta_{k_c,min} \quad (4)$$

which is a constraint in order to satisfy the required quality of service (QoS), where $\vartheta_{k_c,min}$ is the minimum acceptable $SINR$ that user k_c can have to satisfy the QoS,

$$a^z_{k_c}\Big(\frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\sum_{l\neq c, l\in C}\sum_{j\in K_l}|h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} + P_n}\Big) \geq \vartheta_{k_c,min} \quad (5)$$

$$0 \leq P^z_{k_c} a^z_{k_c} \leq P_{max}, \forall c \in C, \forall k \in K_c, \forall z \in Z \quad (6)$$

where P_{max} is the maximum allowed transmit power per device.

$$\sum_{k\in K_c} a^z_{k_c} \leq 1, \forall z \in Z, c \in C. \quad (7)$$

The above constraint guarantees that, within the same cell, a resource unit can only be occupied by one user at a given time.

*B. **Proposed solution***

It can be noted that the proposed optimization problem in this work is a mixed binary integer non-linear programming (MBINP) problem. The variables to be optimized are $a^z_{k_c}$ and $P^z_{k_c} a^z_{k_c}$ which are very difficult to solve. In this regard, we apply a step-wise algorithm [12] for resource unit and power allocation. The algorithm follows three main steps: the first step initializes the transmit power, the second step performs the resource unit allocation, and the third step optimizes the power allocation.

- **Step One: transmit power initialization**

To perform the resource unit allocation, we need to establish an initial transmit power. For our analysis we choose one of the following two methods:

1) setting the initial transmit power equal to the maximum allowed power per device, i.e. P_{max}, or
2) we set the transmit power equal to a required power, calculated from the relation given by the minimum required SINR and average interference power, or maximum tolerated interference level if either is available for the use case or scenario.

Regarding the first option, the transmit power of the user is set to the maximum allowed power for the device.

For the second one, we have a known interference level, e.g average, tolerable, threshold, etc. We define this level at $In^z_{k_c}$. We can compute the transmit power by performing the following procedure:

The SINR for a generic user k_c with a known noise and interference level, $In^z_{k_c}$, is given by:

$$SINR^z_{k_c} = \frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{In^z_{k_c}} \tag{8}$$

The transmit power can be calculated as:

$$P^z_{k_c} = \frac{In^z_{k_c} SINR^z_{k_c}}{|h^z_{k_c,c}|^2} \tag{9}$$

By using the minimum acceptable $SINR$ per user, we have an inequality:

$$P^z_{k_c} \geq \frac{In^z_{k_c} \vartheta_{k_c,min}}{|h^z_{k_c,c}|^2} \tag{10}$$

Thus, we take the lowest acceptable transmit power, i.e. the equality case:

$$P^z_{k_c} = \frac{In^z_{k_c} \vartheta_{k_c,min}}{|h^z_{k_c,c}|^2} \tag{11}$$

- **Step two: resource unit allocation**

After step one, the optimization problem can be written as a function of the resource unit allocation binary variable only, for a fixed transmit power, thus no power constraint:

$$\min \sum_{c \in C} \sum_{z \in Z} \sum_{l \neq c, l \in C} \sum_{j \in K_l} |h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} \tag{12}$$

subject to:

$$SINR^z_{k_c} \geq \vartheta_{k_c,min} \tag{13}$$

$$\sum_{k \in K_c} a^z_{k_c} \leq 1, \forall z \in Z, c \in C. \tag{14}$$

It can further be seen that the optimization problem is a 0-1 assignment integer linear programming problem about $a^z_{k_c}$. To obtain the solution, we use the cooperative scheduling scheme which is discussed later on in Section III.

- **Step three: transmit power allocation**

After performing the resource unit allocation, the result will have several interfering users on each resource unit. Moreover, these interfering users on a specific resource unit are not interfering with other users on any other resource unit because of the OMA scheme; i.e., intra-cell interference can be ignored. Thus, the transmit power allocation for the optimization problem, to minimize interference, can be solved for each resource unit separately because the solutions for each resource unit are independent from each other [12]. Thus, the problem will be solved per resource unit having one interfering user per cell, then this can be repeated for all the other resource units which are occupied by interfering users.

In order to follow the above approach, we reformulate the problem. The new problem has one user per cell interfering with each other. The optimization problem is to find the power allocation among interference users as in [13], [14].

The optimization problem, after the resource unit allocation step, can be written as:

$$\min \sum_{c \in C} \sum_{l \neq c, l \in C} |h_{l,c}|^2 P_l \tag{15}$$

subject to:

$$SINR_c \geq \vartheta_{c,min} \tag{16}$$

$$0 \leq P_c \leq P_{max}, \forall c \in C. \tag{17}$$

where $|h_{c,c}|^2$ is the channel gain of the user from cell c to its base station c, P_c is the transmit power of this user belonging to cell c, $|h_{l,c}|^2$ is the channel gain of the interfering user belonging to cell l to the base station c, and P_l is this interfering user's transmit power.

The $SINR$ for the user belonging to cell c on a given resource unit is given as:

$$SINR_c = \frac{|h_{c,c}|^2 P_c}{\sum_{l \neq c} |h_{l,c}|^2 P_l + P_n} \tag{18}$$

The constraint (16) is not linear. In the following steps, we linearize this constraint as per [15].

We start by substituting 18 in 16.

$$\frac{|h_{c,c}|^2 P_c}{\sum_{l \neq c} |h_{l,c}|^2 P_l + P_n} \geq \vartheta_{c,min} \tag{19}$$

where $\vartheta_{c,min}$ is the minimum required $SINR$ to satisfy the required quality of service for the user belonging to cell c.

By doing cross multiplication

$$|h_{c,c}|^2 P_c \geq \vartheta_{c,min} \Big(\sum_{l \neq c} |h_{l,c}|^2 P_l + P_n \Big) \tag{20}$$

and equivalently,

$$-|h_{c,c}|^2 P_c + \vartheta_{c,min}\Big(\sum_{l\neq c}|h_{l,c}|^2 P_l\Big) \leq -\vartheta_{c,min} P_n \qquad (21)$$

By expanding for $c = 1, 2, \ldots, C$ the inequalities become:

$$c = 1 : -|h_{1,1}|^2 p_1 + \vartheta_1 |h_{2,1}|^2 p_2 + \vartheta_1 |h_{3,1}|^2 p_3 +, \\ \ldots, +\vartheta_1 |h_{C,1}|^2 p_C \leq \vartheta_1 P_n$$

$$c = 2 : -|h_{2,2}|^2 p_2 + \vartheta_2 |h_{1,2}|^2 p_1 + \vartheta_2 |h_{3,2}|^2 p_3 +, \\ \ldots, +\vartheta_2 |h_{C,2}|^2 p_C \leq \vartheta_2 P_n$$

$$c = 3 : -|h_{3,3}|^2 p_3 + \vartheta_3 |h_{1,3}|^2 p_1 + \vartheta_3 |h_{2,3}|^2 p_2 +, \\ \ldots, +\vartheta_3 |h_{C,3}|^2 p_C \leq \vartheta_3 P_n$$

$\vdots$

etc, which can be written in matrix form such as

$$\tilde{A}\tilde{p} \leq \tilde{c} \qquad (22)$$

This can be solved by linear programming solutions in Matlab.

III. Proposed cooperative scheduler and Simulation setup

Algorithm 1 Proposed scheme

procedure GENERATE UE PARAMETERS ▷ using Okumura Hata channel model
  $k \leftarrow |h^z_{k_c,c}|^2$
  while $P^z_{k_c} \neq 0, j \neq i$ **do**
    $SINR^z_{k_c} \leftarrow a^z_{k_c}\left(\frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\sum_{l\neq c, l\in C}\sum_{j\in K_l}|h^z_{j_l,c}|^2 P^z_{j_l} a^z_{j_l} + P_n}\right)$
  return $SINR^z_{k_c}$ ▷ along with other channel parameters
procedure SHARE TO THE SCHEDULER ▷ to compute interference weights
  while $In^z_{k_c} = \frac{|h^z_{k_c,c}|^2 P^z_{k_c}}{\vartheta_{i,min}} - P_n$ **do**
    $select\_the\_UEs\_from\_each\_cell$
  return k ▷ UE IDs for available resources
procedure POWER ALLOCATION
  $\frac{In^z_{k_c}\vartheta_{k_c,min}}{|h^z_{k_c,c}|^2} \leftarrow p$
  while constraint 10 is satisfied **do**
    $calculate\_Rate\_R_k$
  return R_k

A cooperative strategy is considered in which three base stations are connected and communicate before the final resource allocation decision. The centralized cooperative scheduler is considered as the unit that receives the scheduling tables from cooperating base stations. At each base station, the UE channel parameters are observed, and for the cell edge users, their channel parameters are shared together with scheduling tables to the cooperative scheduler. The scheduler then computes the interference weights by taking into consideration i) one transmitting user and ii) one interfering user using the same radio resource from each base station. The users that have the minimum impact of interference are then selected and shared with the base stations to be scheduled at a given frame. When the base stations receive the list of these users and the available resources, power allocation is performed to reduce the unnecessary energy consumption due to excessive transmit power allocation for the cell edge users. We adopt simulation parameters as presented in [16], unless specified otherwise. The overview of the proposed scheme is shown in Algorithm 1. We also selected and implemented additional scheduling schemes i.e. proportional fair, max-min, best CQI, and round-robin as benchmarks. We fixed 10 UEs from each of the cooperating base stations and compare the performance of each scheduler.

For channel quality (CQI) estimation, our proposed scheme implements the Okumura-Hata channel model for small-medium cities. For power allocation, each base station assigns different transmit powers to their corresponding UEs such that good channel condition UEs are allowed to use a maximum transmit power of 14 dBm, UEs with moderate CQI are allowed to use a maximum transmit power of 20 dBm, and UEs with bad CQI are allowed to use the full maximum transmit power of 23 dBm. Compensating the reduced NB-IoT TX power (i.e. 14 dBm) is achieved by increasing the NB-IoT transmission time to maintain the same energy per bit like that of the UE with the maximum TX power (i.e. 23 dBm). Finally, power allocation is performed by considering the UE minimum and maximum power constraints as in Equation 11.

For fairness analysis, we use the Jain's fairness index [17], which is given as:

$$f(R_1, R_2, \ldots, R_k) = \frac{\left[\sum_{k=1}^{K} R_k\right]^2}{K\sum_{k=1}^{K}(R_k)^2} \qquad (23)$$

where K is the total number of UEs under analysis and R_k is the instantaneous UE data rate of user k. With this metric, the fairness is highest when the index value is equal to 1 and the lowest when it is equal to 0.

IV. Numerical results and discussion

This section presents the obtained numerical results which are used for evaluating the performance metrics of our proposed scheme.

A. Achieved data rates

Fig. 1 presents the UE achieved data rates. It can be noted that with the proposed scheme, the UEs achieve better data rates than when using the RR and MaxMin scheduling schemes. This is because the MaxMin scheduling scheme tries to maximize the minimum average data rates and hence allocates more resources to UEs experiencing bad channel conditions, even though the channel conditions are not favorable for transmission. In this regard, UEs with better channel conditions are left without resources, hence the negative impact on the overall system throughput. This is not the case for the RR scheduling scheme where the UEs receive the available

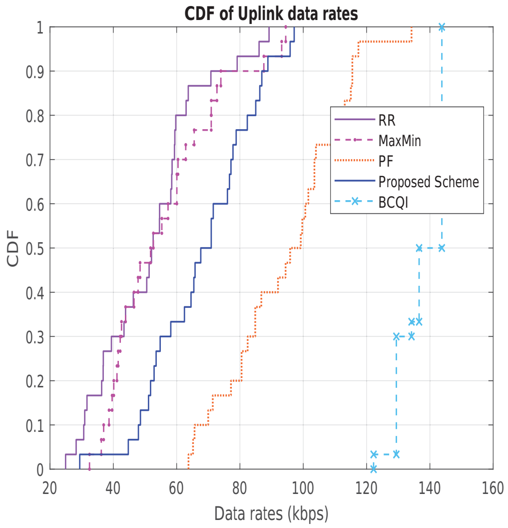

Figure 1. Comparison of UE achieved data rates between our proposed scheme and other scheduling schemes as benchmarks

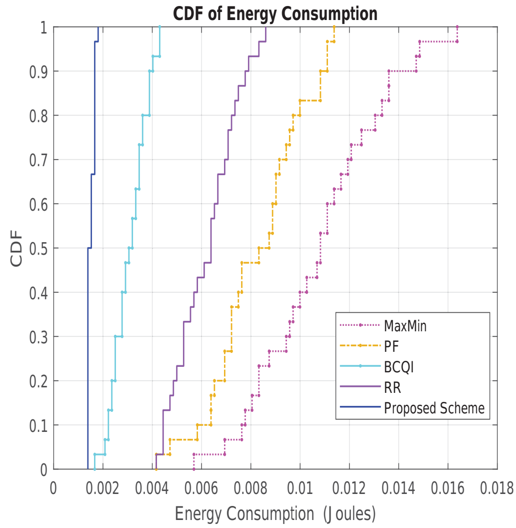

Figure 2. Comparison of UE energy consumption between our proposed scheme and other scheduling schemes as benchmark

resources despite their channel conditions. The downside of RR is when the assigned UE does not have the capacity to transmit due to bad channel conditions; this yields to a drop in overall system data rates.

For example, 50% of UEs under our proposed scheme experience up to 70 $kbps$ while those under the RR and the MaxMin scheduling schemes achieve 50 $kbps$. On the other hand, the Best CQI scheduling scheme performs better than the other scheduling schemes because it only allocates the available resources to UEs that are experiencing relatively best channel conditions. The downside of Best CQI is that it does not include UEs that experience bad channel conditions; as a consequence, UEs from cell edge fall into outage. PR follows after Best CQI in terms of throughput performance, this is because it pursues the maximum rate by assigning resources to UEs with the highest priority. For cell edge users, our proposed scheme outperforms the benchmark schemes because it takes into consideration the impact of ICI which is the bottleneck of performance for these users.

B. UE energy consumption

Fig 2 presents the UE energy consumption. With the proposed scheme, the average experienced energy consumption per UE is the lowest. This is due to the reduced impact of ICI and power allocation to the UE to be scheduled. In this regard, the proposed scheme maximizes SINR, and with better SINR, the UE is considered to experience better channel conditions and therefore gets the available resources for the uplink transmissions with minimum possible repetitions. For example, 50% of UEs under our proposed scheme consumes 60% less in terms of energy consumption as compared to BCQI.

Contrary to the MaxMin scheduling scheme which pursues to maximize the minimum rate (mainly experienced by cell edge UE) by allocating more resources the these UEs which generally use maximum transmit power to counteract the impact of the bad channel conditions. In this context, the UE under the Best CQI achieves relatively lower energy consumption than RR and PF scheduling schemes. BCQI allocates resources to UEs with best channel conditions and hence the needed transmit power becomes lower than for UEs at the cell edge. Since PF aims to maximize the rate of priority UEs regardless of their channel conditions, the energy consumption increases due to the utilization of maximum power for the uplink transmission.

C. UE fairness

Fig. 3 presents the UE fairness for different scheduling schemes. It can be noticed that since most of the schedulers take into account the channel conditions of UEs in order to assign the radio resources, their corresponding fairness increase with the increase in SINR values. This is not the case for the RR scheduling scheme. Since Best CQI pursues to maximize the overall system throughput by considering the UEs with best channel conditions, it automatically excludes all other UEs with relatively bad channel conditions; as a result, it has the lowest fairness index. On the other hand, the RR scheme performs better than the other benchmark

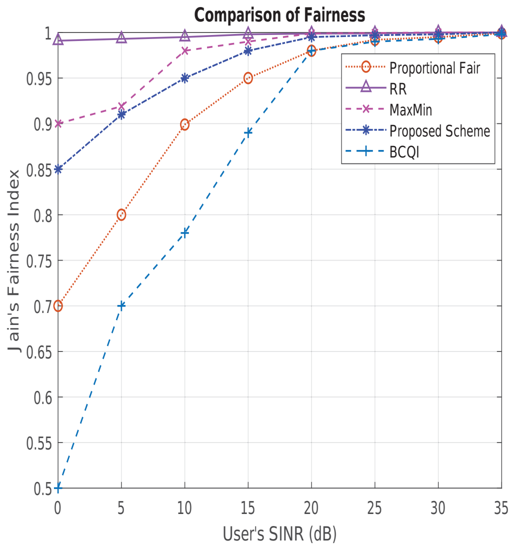

Figure 3. Comparison of Fairness Index between our proposed Scheduler and other existing schedulers

scheduling schemes since it allocates the UE regardless of its channel condition. Our proposed scheme appears in between other scheduling schemes; this is because it considers both the UEs with bad channel conditions and the UEs with good channel conditions, and allocates the resources to these UEs by selecting the pair that will yield minimum ICI impact and allocates powers to these UEs regardless their channel conditions.

Overall, it can be observed that the users operating under the proposed approach experience up to 50% of reduction of energy consumption when compared to the best CQI scheme, thanks to the reduced ICI impact and power allocation. Similarly, 30% and 35% improvements are achieved in terms of user achieved data rates as compared to MaxMin and round-robin schemes, respectively. Additionally, the proposed approach achieves a higher fairness index as compared to Best CQI and PF especially when users are experiencing lower signal to interference plus noise ratio (SINR)

V. Conclusion

In this paper, we have presented an inter-cell interference minimization scheme for massive NB-IoT connectivity. The cooperative scheduling and power allocation are proposed to minimize the impact of ICI and reduce the unnecessary energy consumption for the allocated UE. Moreover, we analyzed and compared the performance of our proposed scheme with respect to other scheduling schemes i.e. PR, RR, Best CQI, and MaxMin benchmarks. Overall, the reduced impact of ICI significantly improves the performance of both cell edge users and cell center users. Additionally, the transmit power allocation minimizes the overall energy consumption for both cell center and cell-edge users.

Our future outlook involves studying the complexity of the proposed scheme and its implementation complexity when using non-orthogonal multiple access (NOMA)

References

[1] C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. De Carvalho, Y. Ji, C. Stefanović, P. Popovski, Q. Wang, M. Schellmann, E. Kosmatos, P. Demestichas, M. Raceala-Motoc, P. Jung, S. Stanczak, and A. Dekorsy, "Towards massive connectivity support for scalable mmtc communications in 5g networks," *IEEE Access*, vol. 6, pp. 28969–28992, 2018.

[2] A. Ghasempour and T. K. Moon, "Optimizing the number of collectors in machine-to-machine advanced metering infrastructure architecture for internet of things-based smart grid," in *2016 IEEE Green Technologies Conference (GreenTech)*, pp. 51–55, 2016.

[3] C. B. Mwakwata, H. Malik, M. Mahtab Alam, Y. Le Moullec, S. Parand, and S. Mumtaz, "Narrowband internet of things (nb-iot): From physical (phy) and media access control (mac) layers perspectives," *Sensors*, vol. 19, no. 11, 2019.

[4] 3GPP, "3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) (Release 13)," Technical Specification (TS) 36.331, 3rd Generation Partnership Project (3GPP), 08 2016. V13.1.0 (2015-11).

[5] H. Malik, H. Pervaiz, M. Mahtab Alam, Y. Le Moullec, A. Kuusik, and M. Ali Imran, "Radio resource management scheme in nb-iot systems," *IEEE Access*, 2018.

[6] F. Qamar, M. H. D. N. Hindia, K. Dimyati, K. A. Noordin, and I. S. Amiri, "Interference management issues for the future 5g network: a review," *Telecommunication Systems*, vol. 71, pp. 627–643, Aug 2019.

[7] B. Maaz, K. Khawam, S. Tohmé, S. Lahoud, and J. Nasreddine, "Joint user association, power control and scheduling in multi-cell 5g networks," pp. 1–6, 03 2017.

[8] S. Saeidian, S. Tayamon, and E. Ghadimi, "Downlink power control in dense 5g radio access networks through deep reinforcement learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2020.

[9] A. Sodhro, S. Pirbhulal, A. Sangaiah, S. Lohano, G. Sodhro, and Z. Luo, "5g-based transmission power control mechanism in fog computing for internet of things devices," *Sustainability*, vol. 10, p. 1258, Apr 2018.

[10] G. J. González, F. H. Gregorio, and J. Cousseau, "Interference analysis in the lte and nb-iot uplink multiple access with rf impairments," in *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pp. 1–4, 2018.

[11] L. Zhang, A. Ijaz, P. Xiao, and R. Tafazolli, "Channel equalization and interference analysis for uplink narrowband internet of things (nb-iot)," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2206–2209, 2017.

[12] O. Elgarhy, L. Reggiani, H. Malik, M. M. Alam, and M. A. Imran, "Rate-latency optimization for nb-iot with adaptive resource unit configuration in uplink transmission," *IEEE Systems Journal*, pp. 1–13, 2020.

[13] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 1, pp. 57–62, 1992.

[14] L. P. Qian, Y. J. Zhang, and J. Huang, "Mapel: Achieving global optimality for a non-convex wireless power control problem," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1553–1563, 2009.

[15] O. Elgarhy and L. Reggiani, "Application of the water filling algorithm to the sum rate problem with minimum rate and power constraint," in *2018 Advances in Wireless and Optical Communications (RTUWO)*, pp. 12–16, 2018.

[16] C. B. Mwakwata, M. M. Alam, Y. Le Moullec, H. Malik, and S. Parand, "Cooperative interference avoidance scheduler for radio resource management in nb-iot systems," in *2020 European Conference on Networks and Communications (EuCNC)*, pp. 154–159, 2020.

[17] S. A. AlQahtani and M. Alhassany, "Comparing different lte scheduling schemes," in *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 264–269, 2013.

Appendix 4

C.B. Mwakwata, O. Elgarhy, M.M. Alam, Y. Le Moullec, S. Pärand, K. Trichias, K. Ramantas. Cooperative Scheduler to Enhance Massive Connectivity in 5G and Beyond by Minimizing Interference in OMA and NOMA. IEEE Systems Journal. IEEE, 2021.

Cooperative Scheduler to Enhance Massive Connectivity in 5G and Beyond by Minimizing Interference in OMA and NOMA

Collins Burton Mwakwata, Osama Elgarhy, *Member, IEEE*, Muhammad Mahtab Alam, *Senior Member, IEEE*, Yannick Le Moullec, *Senior Member, IEEE*, Sven Pärand, Konstantinos Trichias, and Kostas Ramantas

***Abstract*—The fifth-generation (5G) and beyond 5G (B5G) wireless networks introduced massive machine-type communications (mMTC) to cope with the growing demand of massive Internet of Things (IoT) applications. However, the heterogeneous characteristics of massive IoT and diverse quality of service (QoS) requirements may lead to severe interference that could degrade the expected QoS of the cellular ecosystem. Therefore, this article studies the impact of interference caused by mMTC connections. We theoretically model the intercell interference (ICI) minimization problem for the existing orthogonal multiple access (OMA) technique and propose its corresponding solution. Furthermore, we jointly solve the ICI and the cochannel interference minimization problem for the IoT users when the nonorthogonal multiple access (NOMA) technique is used. For the proposed OMA and NOMA schemes, we design a cooperative scheduler to reduce the impact of such interference. The results show that our proposed schemes provide up to 58%, 75%, and 100% more improvements in terms of user's data rates, energy consumption, and connection density, respectively.**

***Index Terms*—Intercell interference (ICI), massive machine-type communication (MMTC), narrow-band IoT (NB-IoT), nonorthogonal multiple access (NOMA), orthogonal multiple access (OMA).**

Nomenclature

Mathematical Symbols

σ_N	Receiver's noise power.
$\vartheta_{x_c,lim}$	$SINR$ threshold for user k_c to satisfy its QoS.
$\vartheta_{c,lim}$	$SINR$ threshold to satisfy the QoS for the UE in cell c.
$a^z_{i_c}$	Allocation matrix of user i at cell c occupying subcarrier z.
$a^z_{q_l}$	Channel occupancy matrix of user q_l at subcarrier z
$a^z_{x_c}$	Allocation matrix of user x_c at subcarrier z
$h^j_{l,c}$	Channel gain from user j, belonging to cell l and the NOMA group M_l within the cell, on cell c.
$h^j_{l,c}$	Channel gain from user j, belonging to cell l and the NOMA group M_l within the cell, on cell c
$h^y_{c,c}$	Channel gain of NOMA user y belonging to the same group M_c
$h^z_{i,c}$	Channel gain of the coallocated interfering user i at cell c on subcarrier z
$h^z_{q_l,c}$	Channel gain of user q_l at cell c
$h^z_{x_c,c}$	Channel gain of user q_l to the base station c at subcarrier z
$h_{c,c}$	Channel gain of the transmitting user c at base station c
$h_{q_l,c}$	Channel gain user q_l at cell c
I^z_c	Interference on resource unit z in cell c
P^j_l	Power of the NOMA user y belonging to the same group M_c
P^z_i	Power of the coallocated interfering user i at subchannel z
$P^z_{q_l}$	Interference power caused by user q_l at subcarrier z
$P^z_{q_l}$	Transmission power of the interfering user q_l at resource unit z
$P^z_{x_c}$	Transmission power of user k_c's at resource unit z
P_{maxm}	Maximum power that can be used by the user for its transmissions.
$SINR^z_{x_c,NOMA}$	User x_c's SINR at subcarrier z under NOMA approach.
$SINR^z_{x_c}$	SINR of user x_c attached to cell c at subcarrier z.

Other Acronyms

$3GPP$	3rd generation partnership project.
5G	5th generation.

Manuscript received April 8, 2021; revised July 13, 2021 and September 1, 2021; accepted September 15, 2021. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 668995 and Grant 951867, in part by European Union Regional Development Fund in the framework of the Tallinn University of Technology Development Program under Grant 2016–2022, and in part by Estonian Research Council under Grant PRG667. *(Corresponding author: Collins Burton Mwakwata.)*

Collins Burton Mwakwata, Osama Elgarhy, Muhammad Mahtab Alam, and Yannick Le Moullec are with the Thomas Johann Seebeck of Electronics, Tallinn University of Technology, 19086 Tallinn, Estonia (e-mail: collins.burton@taltech.ee; osama.elgarhy@taltech.ee; muhammad.alam@taltech.ee; yannick.lemoullec@taltech.ee).

Sven Pärand is with the Telia Estonia Ltd, 10616 Tallinn, Estonia (e-mail: sven.parand@telia.ee).

Konstantinos Trichias is with the WINGS ICT Solutions, 17121 Athens, Greece (e-mail: ktrichias@wings-ict-solutions.eu).

Kostas Ramantas is with the Iquadrat Informatica, 08006 Barcelona, Spain (e-mail: kramantas@iquadrat.com).

Digital Object Identifier 10.1109/JSYST.2021.3114338

AP	Access point.
AVs	Autonomous vehicles.
$B5G$	Beyond 5th generation.
CCI	Cochannel interference.
$CoMP$	Coordinated multi point.
CQI	Channel quality indicator.
$eDRX$	Extended discontinuous reception.
$eMBB$	Enhanced mobile broadband.
FDR	Full duplex relaying
HDR	Half duplex relaying.
ICI	Intercell interference.
IoT	Internet of things.
KKT	Karush–Kuhn Tucker.
$LPWAN$	Low-power wide area network.
$LTE-M$	Long term evolution MTC.
$M2M$	Machine to machine.
$mMTC$	Massive machine-type communications.
$NB-IoT$	Narrowband Internet of things.
$NOMA$	Nonorthogonal multiple access.
NR	New radio.
OMA	Orthogonal multiple access.
PD	Power domain.
QoS	Quality of service.
RF	Radio frequency.
SC	Superposition coding.
SIC	Successive interference cancellation.
UE	User equipment.

I. Introduction

UNLIKE the previous mobile technology generations where the primary focus was to enable human-to-human communications, the fifth-generation (5G) focuses equally on enabling industrial communications by means of service verticals such as massive Internet of Things (IoT), mission-critical communications, and enhanced mobile broadband (eMBB) communications.

It is predicted that by the end of 2023, the number of connected devices needed for supporting the massive IoT deployment will reach 15 billion [1]. Such growth in connectivity will also address the requirements of use cases such as utility monitoring, health care IoT applications, autonomous vehicles (AVs) controlling, and mission-critical applications [2], [3]. In this regard, this article focuses on solving the interference challenges that are brought by the dense deployment of wireless IoT devices in order to enhance the IoT connectivity.

Legacy noncellular commercial technologies such as Wi-Fi, and Bluetooth low-energy (BLE) have limited coverage ranges, which hinders the massive deployment of IoT use cases. This is because these technologies only support short-range wireless access for a few hundred devices [4]. Therefore, to cope with the growing demand for massive connectivity for wide-area coverage, the 3rd generation partnership project (3GPP) introduced massive machine-type communications (mMTC). mMTC is enabled by licensed IoT technologies (e.g., narrow-band IoT (NB-IoT) [5], and unlicensed technologies (e.g., LoRa) [6]. Both of these technologies are categorized as low power wide area networks (LPWAN), aiming at servicing devices located in hard-to-reach areas, with minimum human intervention. However, in contrast to unlicensed technologies, licensed technologies reuse the existing cellular infrastructure and are, therefore more economical and advantageous for cellular telecommunication operators.

The current 5G deployments implement orthogonal multiple access (OMA) schemes which provide orthogonality in terms of frequency resources. However, for massive IoT technologies (i.e., NB-IoT and LTE-M), these OMA schemes are not able to reach the capacity demand for supporting 52,000 devices per cell. Additionally, the 5G broadband and 5G new radio (NR) capabilities bring the possibility of massive connectivity support of up to 1 000 000 devices per square kilometer [7], [8]. In this regard, proactive scheduling and advanced multiple access techniques to support such dense deployment become of great significance.

The nonorthogonal multiple access (NOMA) scheme is considered to be the promising technique to provide capacity enhancement of above 100 000 devices per cell [9]. Contrary to the OMA approach, the NOMA approach gives the possibility to simultaneously superpose multiple devices in a given available radio resource by allocating different power coefficients or codes to enable the successive interference cancellation (SIC) at the receiver [10]. In this regard, NOMA brings an exponential increase in device support as compared to OMA, but at the cost of increased receiver complexity [11].

Despite the advantages that NOMA brings to 5G and B5G networks, it is still unclear if it can be implemented in low-power IoT devices. This is because NOMA involves superposition coding (SC) and SIC at the transmitter and receiver, respectively, which are highly computationally complex for mMTC applications [12].

Furthermore, for both OMA and NOMA approaches, if the radio resources are not well managed, the massive connectivity will lead to massive interference, which will severely degrade the performance of legacy, 5G, and B5G network systems. That is why our work proposes an interference mitigation framework to enhance the cell performance of the OMA and NOMA schemes in a multicell scenario. The main contributions presented in this article are as follows.

1) First, we explicitly formulate the massive interference problem for the OMA and NOMA schemes and propose the corresponding solutions to optimally schedule the radio resources and hence reduce the massive interference.
2) Second, we propose a cooperative scheduling strategy to minimize massive interference for the OMA and NOMA schemes by sharing the scheduling tables between the base stations to increase the overall network capacity.
3) Third, we present the performance enhancements obtained with our proposed approaches and compare the results with existing OMA and NOMA techniques.

To the best of the authors' knowledge, this is the first work that presents a framework to mitigate massive interference caused by massive connectivity of IoT deployment for both OMA and NOMA techniques in 5G and B5G networks.

The rest of the article is organized as follows. Section II presents the related works. Section III presents the analysis of OMA in mMTC systems. For system modeling, we use the NB-IoT system to represent 5G mMTC technology [13]. Section IV presents the analysis of NOMA in mMTC systems. Section V presents the design of the proposed scheduler to mitigate the impact of intercell interference for both OMA and NOMA schemes. Section VI presents the simulation setup, the performance evaluation and achieved enhancements for the OMA and NOMA schemes. The concluding remarks of the article are given in Section VII.

II. Related Works

Several works have studied the OMA/NOMA schemes and their suitability in 5G and B5G systems. For example, in [14], the authors intended to minimize the total energy consumption subject to the computation capacity and execution latency limits. They obtained an optimal transmit power and computation resource allocation based on the Karush–Kuhn Tucker (KKT) conditions. Their results showed that the total energy consumption for both NOMA and OMA schemes increases with the number of NB-IoT user equipment (UEs). However, when compared to OMA, NOMA reduces the total energy consumption by 53.23%. Critically, it should be noted that the authors neglected the impact of intercell interference (ICI).

In [15], the authors investigated the downlink performance of NOMA with randomly deployed cellular users. From the presented analytical formulations, it is shown that the NOMA scheme leads to significant performance gains in terms of ergodic sum-rate. However, the allocated power and the targeted data rate could directly influence the outage performance, i.e., if the allocated power is lower than the required power for successful transmission, the UE will suffer from the outage.

In [16], the authors dealt with the connection density maximization problem in NB-IoT networks by using NOMA. The authors used the bottom–up power filling algorithm and proposed item clustering heuristic approach which allows any number of devices to be multiplexed per subcarrier. It should be noted that the authors suggested multiplexing any number per subcarrier without considering the impact of ICI, which is a potential threat to meeting the performance requirements of NB-IoT massive connectivity.

In [17], the authors proposed two cooperative relaying schemes, i.e., ON/OFF—full-duplex relaying (ON/OFF—FDR), and ON/OFF—half-duplex relaying (ON/OFF—HDR) schemes. Either of the proposed schemes is applied to the cell center user (with good channel conditions) to help relaying the direct NOMA transmissions on the downlink of cell edge users. In this regard, the ON/OFF relaying decision depends upon the quality of direct and relay links from the base station to the cell edge user. From the results, it is shown that the proposed cooperative scheme significantly improves the outage performance and the sum rate of both cell-center and cell-edge users. However, for mMTC devices such as in the LPWAN category, relaying of information leads to an increase in device complexity and cost, which is the limitation for most massive IoT use-cases.

TABLE I
Summarized Comparison of Contributions Between This Work and the Existing Literature

Article	Covered aspects in the contributions				
[Ref]	OMA	NOMA	CCI-aware	ICI-aware	UE Scheduler
[16]		✓			
[17]		✓	✓		
[18]		✓	✓		
[19]		✓	✓		
[20]		✓		✓	
This work	✓	✓	✓	✓	✓

In [18], the authors proposed a novel resource allocation technique for NOMA, based on cooperative cellular networks. In their proposed framework, the NOMA users with good channel conditions act as group heads, hence can relay information to NOMA users with bad channel conditions. Despite the gains of the proposed scheme for high complexity devices, it should be noted that the reduced complexity of NB-IoT devices, power-saving mode, and extended discontinuous reception (eDRx) make relaying of information (i.e., at the low complexity device) unfeasible.

Additionally, new advancements have been made in order to realize the goal of massive IoT under cellular technologies. For example, proactive techniques such as intelligent reflecting surfaces, that enhance the IoT links to the corresponding access point (AP) by counteracting the high pathloss, are introduced in [19]; the improved links are then exploited to better optimize the offloading of computations from the AP to the mobile edge computing (MEC) server. Similarly, proactive radio resource scheduling by means of machine learning techniques [20], and modern link-level adaptation by means of novel interference management approaches are being explored [20]. However, these techniques are not in the scope of this article. Table I presents the comparisons between contributions of this work and the existing literature.

The next section explores the OMA approach and presents the proposed solution to mitigate the massive interference that is caused by the dense deployments of IoT devices in a multicell scenario.

III. Analysis of OMA in mMTC Systems

A. System Model and Problem Formulation for OMA Scheme

For the analysis, we use NB-IoT since it is a long-range promising technique for 5G massive connectivity that currently uses OMA techniques for resource unit scheduling.

Before delving into the details, observe that the notations and abbreviations used in the mathematical analyzes throughout the article are summarized in Appendix A.

We assume that $z = \{1, 2, \ldots, Z\}$ represents the index of the resource units. x_c represents the cell c's UEs, and C, i.e., $c = \{1, 2, \ldots, C\}$, represents the number of cells used in simulation. Therefore, the signal to interference plus noise ratio of user x_c in cell c at unit z is given as

$$SINR^z_{x_c} = a^z_{x_c}\left(\frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l\neq c, l\in C}\sum_{q\in Q_l}|h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N}\right) \tag{1}$$

where $|h^z_{x_c,c}|$ is the channel gain of user x_c at resource z to the base station in cell c, $P^z_{x_c}$ is the transmission power of user x_c at resource z. l represent the interfering cells, with the group of users Q_l and q represents the index of that user. $|h^z_{q_l,c}|$ represents the channel gain of user q_l on unit z attached at cell c, and $P^z_{q_l}$ represents the transmission power of user q_l at unit z. $a^z_{x_c}$ represents the channel allocation matrix, i.e., $a^z_{x_c} = 1$ when the resource is in use, and $a^z_{x_c} = 0$ otherwise. σ_N denotes the receiver's noise power.

In this regard, we aim to minimize the ICI at user k in order to improve the detected SINR to satisfy the expected quality of service. Hence, the optimization problem becomes

$$\min \sum_{c\in C}\sum_{z\in Z}\sum_{l\neq c, l\in C}\sum_{q\in Q_l}|h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} \tag{2}$$

subject to

$$SINR^z_{x_c} \geq \vartheta_{x_c,lim} \tag{3}$$

where $\vartheta_{x_c,lim}$ is the user x_c's $SINR$ threshold to satisfy its QoS

$$a^z_{x_c}\left(\frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l\neq c, l\in C}\sum_{q\in Q_l}|h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N}\right) \geq \vartheta_{x_c,lim} \tag{4}$$

$$0 \leq P^z_{x_c} a^z_{x_c} \leq P_{maxm}, \forall c \in C, \forall x_c \in X_c, \forall z \in Z \tag{5}$$

where P_{maxm} is the maximum power that a given device can use for its transmission

$$\sum_{x_c\in X_c} a^z_{x_c} \leq 1, \forall z \in Z, c \in C. \tag{6}$$

B. Proposed Solution for the OMA Scheme

From the above analysis, the formulation represents a mixed binary integer nonlinear programming (MBINP) problem, with $a^z_{x_c}$ and $P^z_{x_c} a^z_{x_c}$ which are very difficult to solve. Therefore, a step-wise algorithm is used as presented in [21], in order to perform the resource unit and power allocation. The proposed algorithm will implement three main steps as follows.

1) First: initializing transmit power

We aim to set the initial transmit power equal to a required power, which is a function of the $SINR$ threshold to satisfy the required QoS. In this regard, interference level, e.g., average, tolerable, threshold, is already known from the statistics of the channel conditions. Hence, we denote this level as $In^z_{x_c}$ and compute the initial transmit power as follows:

$$SINR^z_{x_c} = \frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{In^z_{x_c}}. \tag{7}$$

The transmit power becomes

$$P^z_{x_c} = \frac{In^z_{x_c} SINR^z_{x_c}}{|h^z_{x_c,c}|^2}. \tag{8}$$

Considering the $SINR$ threshold, the inequality becomes

$$P^z_{x_c} \geq \frac{In^z_{x_c}\vartheta_{x_c,lim}}{|h^z_{x_c,c}|^2}. \tag{9}$$

Therefore, the lowest acceptable transmit power to satisfy the QoS can be presented as

$$P^z_{x_c} = \frac{In^z_{x_c}\vartheta_{x_c,lim}}{|h^z_{x_c,c}|^2}. \tag{10}$$

1) Second: resource allocation

Since the power is already initialized, the optimization problem becomes

$$\min \sum_{c\in C}\sum_{z\in Z}\sum_{l\neq c, l\in C}\sum_{q\in Q_l}|h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} \tag{11}$$

subject to

$$SINR^z_{x_c} \geq \vartheta_{x_c,lim} \tag{12}$$

$$\sum_{x_c\in X_c} a^z_{x_c} \leq 1, \forall z \in Z, c \in C. \tag{13}$$

Now the equation represent a 0-1 assignment problem, hence, we implement the cooperative scheduling scheme as presented in Section V.

1) Third: power allocation

We ignore the impact of intracell interference, thanks to the use of OMA scheduling scheme. However, the intercell interference from adjacent cells' users is experienced at each resource units, therefore we solve the interference problem for each resource unit. In this regard, we assume that, implementing optimal transmit power will reduce unnecessary energy consumption per user.

Therefore, the optimization goal becomes

$$\min \sum_{c\in C}\sum_{l\neq c, l\in C}|h_{l,c}|^2 P_l \tag{14}$$

subject to

$$SINR_c = \left(\frac{|h_{c,c}|^2 P^t_c}{\sum_{l\neq c}|h_{l,c}|^2 P_l + \sigma_N}\right) \geq \vartheta_{c,lim} \tag{15}$$

$$0 \leq P^t_c \leq P_{maxm}, \forall c \in C \tag{16}$$

where $|h_{c,c}|^2$ and $|h_{l,c}|^2$ are the channel gains of transmitting user, and interfering user, respectively. P^t_c and P_l are the transmit powers of of transmitting user and interfering user, respectively.

Since constraint (15) is nonlinear, we therefore make it linear as follows:

$$|h_{c,c}|^2 P^t_c \geq \vartheta_{c,lim}\left(\sum_{l\neq c}|h_{l,c}|^2 P_l + \sigma_N\right) \tag{17}$$

equivalently

$$-|h_{c,c}|^2 P^t_c + \vartheta_{c,lim}\left(\sum_{l\neq c}|h_{l,c}|^2 P_l\right) \leq -\vartheta_{c,lim}\sigma_N. \tag{18}$$

Algorithm 1: Proposed OMA Scheme.

procedure User equipment creation ▷ applying Okumura-Hata model
  $k \leftarrow |h^{z}_{x_c,c}|^2$
  while $P^{z}_{x_c} \neq 0, j \neq i$ **do**
    $Equation(1)$
  return $SINR^{z}_{x_c}$
procedure SHARE THE SCHEDULING TABLES
  while $Equation(10)$ **do**
    $compute\_the\_best\_combination\_of\_UEs$
  return x_c
procedure OPTIMAL POWER ALLOCATION
  $\frac{In^{z}_{x_c}\vartheta_{x_c,lim}}{|h^{z}_{x_c,c}|^2} \leftarrow p$
  while $\frac{In^{z}_{x_c}\vartheta_{x_c,lim}}{|h^{z}_{x_c,c}|^2} \leftarrow p$ **do**
    $calculate\_Rate\_R_k$
  return R_k

Performing the inequality expansion for $c = 1, 2, \ldots, C$

$$c = 1 : -|h_{1,1}|^2 p_1 + \vartheta_1 |h_{2,1}|^2 p_2 + \vartheta_1 |h_{3,1}|^2 p_3 +, \ldots, +\vartheta_1 |h_{C,1}|^2 P^t_c \leq \vartheta_1 \sigma_N$$

$$c = 2 : -|h_{2,2}|^2 p_2 + \vartheta_2 |h_{1,2}|^2 p_1 + \vartheta_2 |h_{3,2}|^2 p_3 +, \ldots, +\vartheta_2 |h_{C,2}|^2 P^t_c \leq \vartheta_2 \sigma_N$$

$$c = 3 : -|h_{3,3}|^2 p_3 + \vartheta_3 |h_{1,3}|^2 p_1 + \vartheta_3 |h_{2,3}|^2 p_2 +, \ldots, +\vartheta_3 |h_{C,3}|^2 P^t_c \leq \vartheta_3 \sigma_N.$$

$\vdots$

The above expansion follows a matrix form which can be shortened as

$$\tilde{A}\tilde{p} \leq \tilde{c}. \tag{19}$$

In this article, Algorithm 1 presents the simulation implementation with additional procedures as discussed in Section V.

Since the OMA approaches employ orthogonality when allocating the available resources, most of the 5G mMTC systems fail to reach the cell capacity target as specified in the standard due to the limited available spectrum.

To overcome this limitation, the NOMA scheme presents significant advantages regarding spectrum efficiency, hence it is a promising technique to accommodate massive IoT applications for beyond 5G networks [22].

A generic architecture presenting the principles of the OMA and NOMA schemes in 5G networks is shown in Fig. 1. As can be seen, the OMA scheme allocates orthogonal physical resource blocks (PRB) to different user equipment (UE) transmitting at a given time slot. The NOMA scheme allocates a given PRB to multiple UEs, with different power coefficients or codes in order to guarantee the successful decoding of data at the receiver.

The next section studies the NOMA approach and proposes the corresponding solution in order to mitigate the cochannel and intercell interference in a multicell scenario.

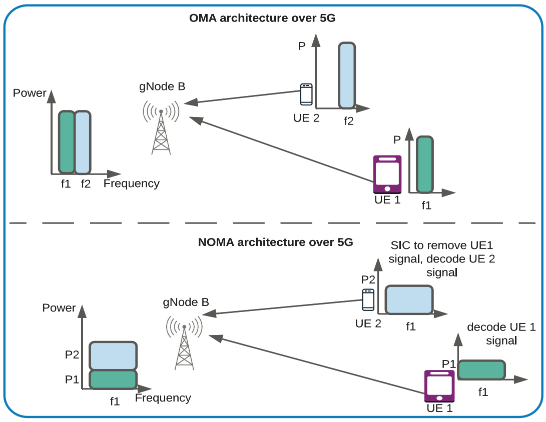

Fig. 1. Generic architecture representing the OMA and NOMA schemes over 5G networks. In the OMA scheme, every UE is provided with a unique physical resource block at a given time. In the NOMA scheme, multiple UEs are superposed in a given resource block but with different power coefficients or codes to enable the superposition coding and the SIC at the transmitter and receiver, respectively.

IV. Analysis of NOMA in mMTC Systems

A. System Model and Problem Formulation for NOMA Scheme

We consider a system of x transmitting users served by cooperating base stations, and x={1,2,..., X} be its index set of users. We consider M to be a positive, maximum number of devices that can be supported per subcarrier. $z = \{1, 2, \ldots, Z\}$ represents the index of the resource units. x_c represents the cell c's UEs, and C, i.e., $c = \{1, 2, \ldots, C\}$, represents the number of cells used in simulation. Therefore, the signal to interference plus noise ratio of the NOMA user x_c at unit z is given as

$$SINR^{z}_{x_c,NOMA} = a^{z}_{x_c}\left(\frac{|h^{z}_{x_c,c}|^2 P^{z}_{x_c}}{I^{z}_{c} + \sigma_N}\right) \tag{20}$$

where I^{z}_{c} is the total interference experienced by user x_c from the coallocated interfering users i and users l from adjacent cells, which is given as

$$I^{z}_{c} = \sum_{i \neq k, i \in M} |h^{z}_{i,c}|^2 P^{z}_{i} a^{z}_{i_c} + \sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^{z}_{q_l,c}|^2 P^{z}_{q_l} a^{z}_{q_l}. \tag{21}$$

As was also the case for OMA, we aim to minimize the ICI at user x_c from users q_l, and interference from the NOMA users i of the same cell assigned to the same resource unit. The objective function can therefore be expressed as (22)

$$\min \sum_{c \in C} \sum_{z \in Z} \left(\sum_{i \neq k, i \in M} |h^{z}_{i,c}|^2 P^{z}_{i} a^{z}_{i_c} + \sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^{z}_{q_l,c}|^2 P^{z}_{q_l} a^{z}_{q_l} \right) \tag{22}$$

subject to

$$a_{x_c}^z\left(\frac{|h_{x_c,c}^z|^2P_{x_c}^z}{I_c^z+\sigma_N}\right)\geq\vartheta_{x_c,lim} \quad (23)$$

$$0\leq P_{x_c}^z a_{x_c}^z\leq P_{maxm},\forall c\in C,\forall x_c\in X,\forall z\in Z \quad (24)$$

where P_{maxm} is the maximum allowed power per device.

$$\sum_{x_c\in X} a_{x_c}^z\leq 1,\forall z\in Z,c\in C \quad (25)$$

$$\sum_{x_c\in X} a_{x_c}^z\leq M,\forall i\in M\forall z\in Z,c\in C. \quad (26)$$

It can be seen that the objective function is a combinatorial optimization problem and is hence difficult to solve. In this regard, the proposed solution is presented as follows.

B. Proposed Solution for NOMA Scheme in NB-IoT System

To solve the NOMA problem we follow the same steps as in OMA. First, we set an initial interference power for all the users. Second, we perform the scheduling for all the users. Finally, we implement the power allocation to further reduce the interfering powers at the desired receiver. The initial interference power will be allocated as we did for OMA. However, the channel allocation problem in (22) will have the following two assumptions:

1) the power is not a variable;
2) there are no power constraints.

Therefore, we perform power allocation after the channel assignment. We rewrite the optimization problem in a similar way to that of OMA, i.e., by working per resource unit since there is no interference from adjacent resource units; however, we have to add the NOMA interference users in a given resource unit. Since in the OMA we had one user per resource unit per cell, there was no need to add a subscript for the resource unit. However, because of NOMA, we have more than one user, thus, we define M_c as the group of NOMA users per cell per resource unit, and x_c is a user in cell c that belongs to M_c, and we omit the resource unit index. In this regard, the optimization goal becomes (27)

$$\min\sum_{c\in C}\sum_{x_c\in M_l}\left(\sum_{l\neq c,l\in C}\sum_{q\in M}|h_{l,c}^j|^2P_l^j+\sum_{y\neq x_c,y\in M_c}|h_{c,c}^z|^2P_c^y\right) \quad (27)$$

where $h_{l,c}^j$ is the channel gain from user j, belonging to cell l and the NOMA group M_l within the cell, on cell c. P_l^j is the power of this user. These two terms represent the intercell interference from all the NOMA users of other cells. As for the NOMA part; $h_{c,c}^z$ is the channel gain of NOMA user y belonging to the same group M_c. subject to

$$SINR_{x_c,NOMA}^z\geq\vartheta_{c,lim} \quad (28)$$

$$0\leq P_{x_c}^z\leq P_{maxm},\forall c\in C. \quad (29)$$

The $SINR_{x_c,NOMA}^z$ is then given as (31) shown at bottom of the next page, which can be solved in the same way as for OMA. However, the number of inequalities will be larger. Moreover, this equation does not take into account the SIC effect on removing interference from other NOMA users within the same cell in the same resource unit. The effect of the SIC can be included in the inequalities by simply putting zero for the NOMA interference users within the same cell as the main user after passing through the SIC. The constraint (28) is not linear; in this regard, we start by substituting (30) into (28), hence linearize as follows:

$$\mathrm{SINR}_{x_c,NOMA}^z=\left(\frac{|h_{x_c,c}^z|^2P_{x_c}^z}{\frac{\sum_{l\neq c}\sum_{j\in M_l}|h_{l,c}^j|^2P_l^j+\sigma_N}{+\sum_{y\neq x_c,y\in M_c}|h_{c,c}^z|^2P_c^y}}\right) \quad (30)$$

$$\left(\frac{|h_{c,c}^k|^2P_c^k}{\frac{\sum_{l\neq c}\sum_{j\in M_l}|h_{l,c}^j|^2P_l^j+\sigma_N}{+\sum_{y\neq x_c,y\in M_c}|h_{c,c}^z|^2P_c^y}}\right)\geq\vartheta_{c,lim} \quad (31)$$

$$|h_{c,c}^k|^2P_c^k\geq\vartheta_{c,lim}\left(\sum_{l\neq c}\sum_{j\in M_l}|h_{l,c}^j|^2P_l^j+\sum_{y\neq x_c,y\in M_c}|h_{c,c}^z|^2P_c^y+\sigma_N\right) \quad (32)$$

$$|h_{c,c}^k|^2P_c^k-\vartheta_{c,lim}\left(\sum_{l\neq c}\sum_{j\in M_l}|h_{l,c}^j|^2P_l^j\right) -\vartheta_{c,lim}\left(\sum_{y\neq x_c,y\in M_c}|h_{c,c}^z|^2P_c^y\right)\geq\vartheta_{c,lim}\sigma_N \quad (33)$$

equivalently

$$-|h_{c,c}^k|^2P_c^k+\vartheta_{c,lim}\left(\sum_{l\neq c}\sum_{j\in M_l}|h_{l,c}^j|^2P_l^j\right) +\vartheta_{c,lim}\left(\sum_{y\neq x_c,y\in M_c}|h_{c,c}^z|^2P_c^y\right)\leq\vartheta_{c,lim}\sigma_N. \quad (34)$$

Substituting $c=1,2,\ldots,C$ equation becomes

$c=1,k=1:$

$$-|h_{1,1}^1|^2p_1^1+\vartheta_{1,min}\Big(|h_{2,1}^1|^2p_2^1+|h_{2,1}^2|^2p_2^2 +\cdots+|h_{3,1}^1|^2p_3^1+|h_{3,1}^2|^2p_3^2 +\cdots,\ldots,+|h_{C,1}^1|^2p_C^1+|h_{C,1}^2|^2p_C^2+\cdots\Big) +\vartheta_{1,min}\Big(|h_{1,1}^2|^2p_1^2+|h_{1,1}^3|^2p_1^3+, +,\ldots,+|h_{1,1}^{M_1}|^2p_1^{M_1}\Big)\leq\vartheta_{1,lim}\sigma_N$$

$c=1,k=2:$

$$-|h_{1,1}^2|^2p_1^2+\vartheta_{1,min}\Big(|h_{2,1}^1|^2p_2^1+|h_{2,1}^2|^2p_2^2$$

TABLE II
MAIN SIMULATION PARAMETERS FOR THE PROPOSED COOPERATIVE SCHEDULING STRATEGY [25]

Simulation Parameters	
Name	Value
(a) Transmit power of base station, {UE} (dBm)	46 , $\{14, 20, 23\}$
(b) Modulation scheme	BPSK
(c) Carrier frequency (MHz)	900
(d) Receiver Thermal Noise density (dBm/Hz)	-174
(e) No. cooperating base station	3
(f) No. of UE per cell	10
(g) Interference Margin (dB)	0
(h) Channel model	Okumura Hata
(i) Effective Noise Power (dBm)	d + q + f + 10log(r)
(j) Required / calculated SINR (dB)	
(k) Receiver sensitivity	h + i
(l) MCL (dB)	a - j
(m) Modulation scheme	BPSK
(n) No. of antenna support per UE	1
(o) Height of base station, UE (m)	100, 1
(p) Radius of a cell (km)	1
(q) Noise figure of base station, UE (dB)	9, 5
(r) Occupied System bandwidth (kHz)	180

$$
\begin{aligned}
&+\cdots+|h_{3,1}^{1}|^{2}p_{3}^{1}+|h_{3,1}^{2}|^{2}p_{3}^{2}\\
&+\cdots,\ldots,+|h_{C,1}^{1}|^{2}p_{C}^{1}+|h_{C,1}^{2}|^{2}p_{C}^{2}+\cdots\Big)\\
&+\vartheta_{1,min}\Big(|h_{1,1}^{1}|^{2}p_{1}^{1}+|h_{1,1}^{3}|^{2}p_{1}^{3}\\
&+,\ldots,+|h_{1,1}^{M_1}|^{2}p_{1}^{M_1}\Big)\leq\vartheta_{2,lim}\sigma_N\\
c=2,k=1:\ &-|h_{2,2}^{1}|^{2}p_{2}^{1}+\vartheta_{2,min}\Big(|h_{1,2}^{1}|^{2}p_{1}^{1}+|h_{1,2}^{2}|^{2}p_{1}^{2}\\
&+\cdots+|h_{3,2}^{1}|^{2}p_{3}^{1}+|h_{3,2}^{2}|^{2}p_{3}^{2}\\
&+\cdots,\ldots,+|h_{C,2}^{1}|^{2}p_{C}^{1}+|h_{C,2}^{2}|^{2}p_{C}^{2}+\cdots\Big)+\vartheta_{2,min}\\
&\Big(|h_{2,2}^{2}|^{2}p_{2}^{2}+|h_{2,2}^{3}|^{2}p_{2}^{3}+,\ldots,+|h_{2,2}^{M_2}|^{2}p_{2}^{M_2}\Big)\leq\vartheta_{3,lim}\sigma_N
\end{aligned}
$$

$\vdots$ etc.

The above expansion can be shorten as a matrix of the following form:

$$\tilde{B}\tilde{q}\leq\tilde{v}. \tag{35}$$

In this regard, (34) can be solved by linear programming solutions in MATLAB. Algorithm 2 presents the proposed implementation of the NOMA approach; simulation parameters are presented in Table II, unless specified otherwise.

C. Complexity Analysis

As seen in Algorithm 2, from line 1 to line 5 the algorithm computes the channel parameters for all users attached to the corresponding base stations. This operation has a computation cost of $O(n)$. Then from line 6 to line 14, there is the nested while or for-loop such that in the first loop, the interference weight is analyzed, and users (i.e., which have lower interference impact on each other) are superposed at a given subcarrier. In the second loop, the transmit power is allocated to users in order to reduce unnecessary energy consumption. This operation has the computation cost of $O(n^2)$. From line 16 to the end of the algorithm, we evaluate the achieved user performance and the computation cost is $O(n)$. In this regard, the computation complexity becomes

$$O(n+n^2+n). \tag{36}$$

Thus, the computational complexity of our proposed algorithm is $O(n^2)$, i.e., quadratic complexity.

If we analyze the computation complexity in the single form (i.e., without considering the interference impact), from line 1 to line 5 the algorithm computes the channel parameters for all users attached to the corresponding base stations. The operation still has a computation cost of $O(n)$. However, from line 6 to line 14, we will have only one whole or for-loop to allocate different power coefficients to NOMA users to enable the SIC at the receiver. This operation has a computation cost of $O(n)$. From line 16 to the end of the algorithm, we evaluate the achieved user performance and the computation cost remains $O(n)$. In this regard, if we do not consider interference reduction, then the computation cost becomes $O(n+n+n)=O(n)$.

Therefore, the complexity overhead of our proposed scheme ($O(n^2)$ versus $O(n)$) is an acceptable tradeoff, given the performance enhancements brought by the interference reduction.

In the next section, we present the proposed cooperative scheduler that is used to minimize the previously studied impact of massive interference for both OMA and NOMA schemes.

V. PROPOSED COOPERATIVE SCHEDULER

We consider cooperation between base stations in order to enhance the interference minimization by sharing the scheduling tables, which contain the channel parameters of UEs to be scheduled. For example, during OMA scheduling, the scheduler computes the interference possibilities for each UE by considering the inter-cell interference that is caused by UEs that are allocated at the same resource at a given time. Furthermore, we assume that the impact of cochannel interference is negligible, i.e., by orthogonality, hence the main impact of interference is ICI. To reduce such impact of ICI, we utilize the shared scheduling tables to compute the best combination of UEs that have the minimum impact of interference. From the retrieved best combination, each base station allocates the available resource unit at a given time slot for its corresponding UE. For the NOMA approach, each base station classifies the UEs into three groups based on their channel parameters, i.e., good, moderate, and bad UEs. We assume that we have two main sources of interference, i.e., NOMA interference from users that are simultaneously

Algorithm 2: Proposed NOMA Scheme.

procedure User Equipment Creation ▷
  $x_c \leftarrow |h^z_{x_c,c}|^2$
  while $P^z_{x_c} \neq 0, j \neq i$ **do**
    $Equation(30)$
  return $SINR^z_{x_c}$
procedure SHARE THE SCHEDULING TABLES
  while $In^z_{x_c} = \frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\vartheta_{i,min}} - \sigma_N$ **do**
    $compute\_the\_best\_combination\_of\_UEs$
    $Divide\_the\_UEs\_in\_three\_groups$
    $Superpose\_One\_UE\_from\_each$
    $group\_in\_a\_given\_subcarrier$
  while $P^z_{x_c} \neq 0$ **do**
    $allocate\_power\_according\_to\_constraint$:
    $0 < P^z_{x_c} \leq P_{maxm}, \forall c \in C$
  return x_c
procedure EVALUATE
  while $\frac{I^z_c \vartheta_{c,lim}}{|h^z_{x_c,c}|^2} \leftarrow p$ **do**
    $calculate\_Rate\_R_k$
    $calculate\_Energy\_Consumption$
  return $R_{x_c}, energy$

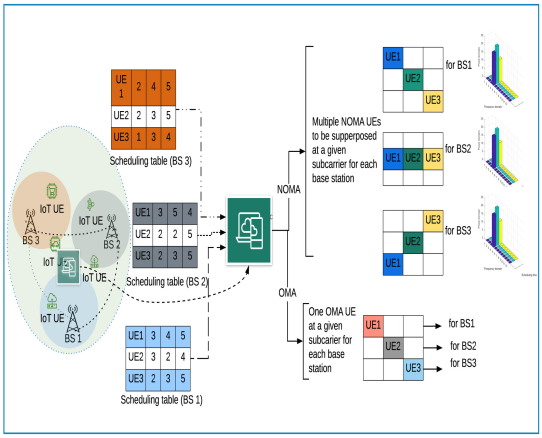

Fig. 2. Proposed radio resource management scheme exploiting the NOMA scheme in NB-IoT systems. Each cooperating base station (BS1 to BS3) share their respective scheduling tables for their future transmission. Then ICI is avoided by allocating resources to UEs whose impact in terms of interference is the lowest among the UEs. Then the base stations implement the OMA or NOMA scheme for their corresponding choice of strategy.

allocated at a given resource unit at a given time slot, and the ICI from other users transmitting at the same resource unit but from adjacent cells. Then the scheduling tables from each base station are shared with the cooperative scheduler. After receiving the tables, the scheduler selects one UE from each group of users to be simultaneously superposed at a given resource unit.

In this regard, a maximum number of three UEs can simultaneously occupy a given resource unit at a given time slot. The scheduler computes the best combination of UEs for all the available resource units before sharing the respective allocation of slots within a frame to the base stations. Additionally, the scheduler performs the power allocation to reduce the impact of cochannel interference as well as ICI. During power allocation, we assume the power constraints for each group as follows: good channel users $P_{const} = 14\ dBm$, moderate channel users $P_{const} = 20\ dBm$, and bad channel users $P_{const} = 23\ dBm$. Different power coefficients are assigned to users to successively perform SIC at the receiving base station. We assume that the good channel users are close to their serving base station and hence can be given lower power constraints, and vice-versa is true for bad channel users. An overview of the proposed cooperative strategy for OMA and NOMA is presented in Fig. 2.

In general, unlike the joint processing in coordinated multipoint (CoMP) in LTE systems where a UE at the cell-edge is served by two or more base stations to improve signals quality and increase throughput [23], in our proposed cooperative approach each base station serves its own users. The simulation parameters are similar as presented in [24], with some modifications adapted for the NOMA approach. The overview of the followed scheme is highlighted in Algorithms 1 and 2. We also selected additional scheduling schemes, i.e., proportional fair (PF), max–min, and round-robin as benchmarks for comparison purposes.

Furthermore, we adapt the Okumura–Hata channel model for small-medium cities as presented in [26]. And we use Jain's fairness index to analyze the fairness of the studied schemes.

It should be noted that, even though the measure of fairness is generally subjective, we assume that if a system reaches fairness, then all the connected devices should achieve individual fairness. In this regard, the Jains' fairness index provides quantitative insight into the overall system fairness; however, it can not identify the UEs that are unfairly treated. Entropy could also be used to categorize the fairness performance among the studied schemes; however, its effectiveness regarding fairness measurements is not clear yet [27].

Furthermore, in an adequate fairness model, especially for low complexity IoT devices such as NB-IoT in massive connectivity scenarios, long-term fairness is more important due to the scarcity of the radio resource. In this regard, throughout this article, the fairness analysis is performed at the end of the allocation life cycle.

For performance evaluation, the parameters used in the simulation are presented in Table II. We consider three cooperating base stations, as shown in Fig. 2. We perform scheduling for each slot in a given total scheduling frame consisting of 10 time-slots, and 12 resource units (subcarriers). For the OMA approach, only 1 UE can occupy a given resource unit, at a given time slot in a given base station. This yields a capacity limit of 10 UEs per base station for the total scheduling frame. However, for the NOMA approach, up to 3 UEs can occupy a given resource unit at a given time slot in a given base station. This yields a capacity limit of 30 UEs per subcarrier within a total scheduling frame. It should be noted that with increased system bandwidth the number of connected devices exponentially increases.

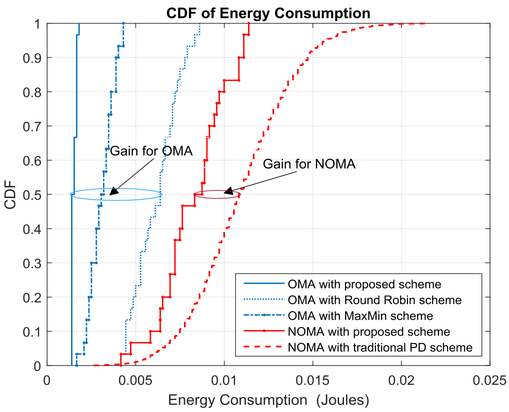

Fig. 3. Comparison of UE energy consumption between OMA and NOMA schemes.

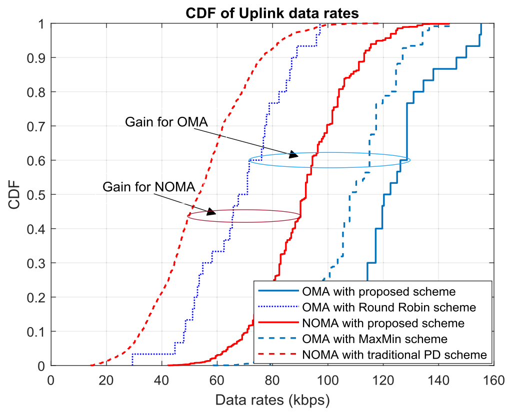

Fig. 4. Comparison of UE achieved data rates for OMA and NOMA schemes.

VI. Performance Evaluation

We perform the analysis for 1000 iterations; for each iteration, the UEs are randomly distributed across each cell in order to calculate the channel parameters at different positions. We select a set of transmitting UEs from all base stations and another set of interfering UEs from adjacent cells at a given time slot. For the OMA scheme, we consider the set of interfering UEs as the UEs having the same time slot but from adjacent cells. For NOMA, however, we consider the interfering set as the UEs from adjacent cells transmitting at the same time slot, and UEs transmitting at the same time slot but from the same cell. We compute different performance metrics and the results are as presented in the next section.

Some of the simulation parameters may impact the results significantly. For example, increasing the number of users per cell increases the number of interfering users and hence can lower the actual performance as compared to the expected one. In this regard, it is advisable to set the expected quality of service requirement for the serving base station and for the served users. Similarly, as the radius of the cell increases, the experienced path-loss at the user increases; in this regard, it is advisable to follow the base station settings from the telecommunication operator as the benchmark for the scenario under study.

Fig. 3 presents the UE energy consumption for OMA and NOMA with the proposed scheme against the benchmark schemes. It can be noted that the OMA scheme experiences relatively lower energy consumption as compared to the NOMA scheme. For example, 50% of UEs under the proposed OMA experience about 40% and 75% lower energy consumption than MaxMin and Round Robin, respectively. Similarly, for NOMA, our proposed scheme achieves lower energy consumption as compared to the traditional power domain NOMA (PD NOMA). The energy consumption enhancements are enabled by the reduced impact of intercell interference and the optimal power allocation that reduces the excessive transmission power while guaranteeing the expected QoS at the transmitting users. Furthermore, the reduced interference impact maximizes the SINR, hence relatively reduces the number of repetitions at the transmitting UEs.

For example, the MaxMin scheme maximizes the minimum achieved QoS by allocating more resources to cell-edge users, this approach causes UEs to use maximum transmit power which yields more energy consumption. On the other hand, Round Robin implements a first-come first-saved strategy while allocating resources to UEs; while doing so, cell edge UEs suffer from uncontrolled massive interference from adjacent cells hence increases the transmit power to counteract the ICI.

On the other hand, traditional PD NOMA simultaneously allocates the same available resources to a given number of UEs, when the ICI is not well managed these UEs suffer from both the co-channel interference as well as interference from adjacent cell UEs. In this regard, the impact of interference at a given subcarrier is more significant, hence the UEs are forced to use the maximum allowed power to transmit which leads to excessive energy consumption.

Our proposed NOMA scheme takes into account both the co-channel interference and ICI; in this regard, UEs are better scheduled and their transmit power is optimized. In doing so, the overall energy consumption is reduced.

Fig. 4 presents the achieved UE data rates. From the analysis of the results, our proposed scheme outperforms both the Round Robin and Maxmin schemes. Contrary to the MaxMin scheme that favors the UEs under bad channel conditions, and Round Robin that operates under a first-come-first-served strategy, our proposed scheme considers the UEs in both good and bad channel conditions by allocating different power coefficients to avoid interference. By doing so, the UEs under our proposed scheme, especially under the OMA approach achieve relatively higher throughput. It should be noted that the OMA approach has fewer UEs per resource unit, therefore experience a low impact of interference which leads to higher achieved throughput.

It is observed that, with our proposed approach, more than 50% of the UEs achieve above 120 $kbps$, however, Round Robin and the MaxMin scheduling schemes achieve only 70 and 105 $kbps$, respectively. On the other hand, the UEs under the

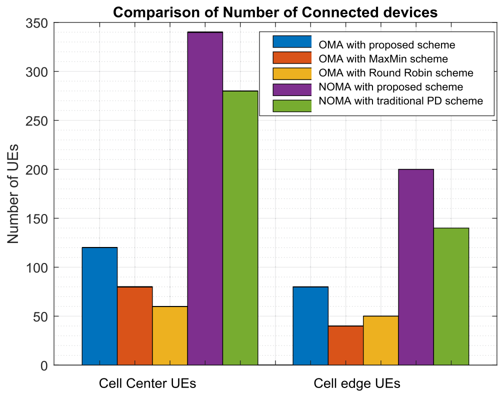

Fig. 5. Comparison of the total number of devices that can be connected in a given scheduling frame for OMA and NOMA schemes.

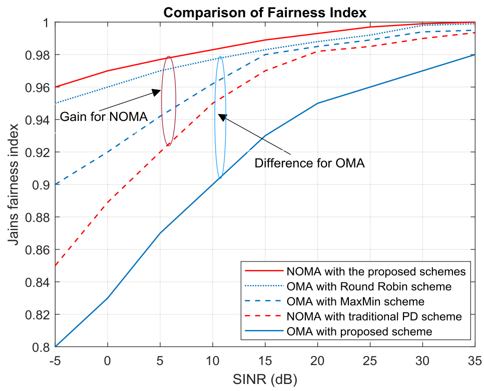

Fig. 6. Comparison of the degree of fairness for OMA and NOMA schemes.

proposed NOMA scheme achieve an average of 40 $kbps$ higher than the UEs under traditional PD NOMA. The enhancements are due to the controlled ICI impact, as well as limiting the number of UEs that can be superposed at a given subcarrier.

Fig. 5 presents the number of connected devices for OMA and NOMA schemes, at the cell center and the cell edge. It can be noticed, our proposed NOMA scheme outperforms the OMA schemes by more than double for both cell-center users as well as for cell-edge users. Similarly, our proposed NOMA scheme outperforms the traditional NOMA scheme in terms of connected UEs by up to 21%. These enhancements are achieved thanks to the minimized ICI impact, enhanced by the cooperative scheduling between the adjacent base stations. Contrary to the OMA schemes, the overall connectivity enhancements for the NOMA schemes are due to the possibility of superposing multiple UEs in the same tone. For example, in the OMA scheme, only one UE can occupy one or multiple tones at a given uplink scheduling frame.

Fig. 6 presents the UE fairness when our proposed scheme is compared to other scheduling schemes from literature. During the fairness analysis, the simulated SINR range was based on real-time SINR values from IoT sensors; the SINR values range from -5 dB (i.e., lowest SINR) to 35 dB (highest SINR). Since most of the schedulers consider the channel parameters before attributing the radio resources to UEs, the higher $SINR$ values trigger the increase in scheduling fairness. Our proposed NOMA scheme outperforms the traditional PD scheme and OMA schemes hence is more suitable for massive connectivity in dense networks. With proactive scheduling (avoiding ICI) and optimal power allocation, the same available resources can be used for devices at the cell edge and devices at the cell center.

On the other hand, the Round Robin scheme outperforms the benchmark OMA schemes by allocating resources to UEs regardless of their channel condition. For example, our proposed OMA scheme lags behind both Round Robin and MaxMin schemes; this is due to its selection process which incurs prioritization, and as a consequence, results in lower fairness. If these schemes are implemented in practical systems, the fairness measurements shown above can help to compensate the devices that are treated unfairly (low fairness index) in the previous allocation step and improve the targeted QoS in the current allocation step.

The potential drawback of our proposed strategy is that as the number of cooperating base stations increase, the back-haul delay increases. Similarly, with the increased number of devices per cell, the sharing of scheduling tables may generate potential delays. Additionally, our proposed approach necessitates high synchronization between cooperating base stations in order to ensure real-time end-to-end performance. In this regard, it may increase the computational complexity at the base stations. In this regard, it is necessary to implement strong computing machines with real-time synchronization clocks at the base stations and utilize high-speed links between the base stations in order to ensure the feasibility of our proposed schemes in real systems.

VII. Conclusion

In this article, we analyzed the impact of massive interference due to massive connectivity in 5G and B5G networks. We proposed the corresponding solutions for the OMA and NOMA approaches to enhance the users' and cell performance. To assess our proposed approach, we compared it with benchmark schemes from the literature. The simulation results show that the proposed NOMA scheme is more spectrum efficient than OMA as it supports more than twice the number of connected devices for the same number of available resources. Furthermore, other network performance metrics such as throughput, user's energy consumption, and fairness are analyzed, discussed, and compared for both the OMA and NOMA schemes. In general, the reduced impact of interference and the proposed power allocation techniques reduce the average energy consumption per device hence are more suitable for massive IoT deployments as it enhances the device's battery life longevity. Our future outlook involves analyzing the complexity tradeoff that our proposed scheme will influence at the base station. Similarly, we aim to study the flexible duplexing technique in order to efficiently use the available spectrum to further enhance the massive connectivity of IoT devices for 5G and beyond networks.

References

[1] Cisco, "Cisco annual Internet report (2018–2023) white paper," Cisco, San Jose, CA, USA, White Paper, Dec. 2020.

[2] R. Abozariba, M. K. Naeem, M. Patwary, M. Seyedebrahimi, P. Bull, and A. Aneiba, "Noma-based resource allocation and mobility enhancement framework for IoT in next generation cellular networks," *IEEE Access*, vol. 7, pp. 29158–29172, 2019.

[3] M. M. Alam, H. Malik, M. I. Khan, T. Pardy, A. Kuusik, and Y. Le Moullec, "A survey on the roles of communication technologies in IoT-based personalized healthcare applications," *IEEE Access*, vol. 6, pp. 36611–36631, 2018.

[4] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[5] C. B. Mwakwata, H. Malik, M. Mahtab Alam, Y. Le Moullec, S. Parand, and S. Mumtaz, "Narrowband Internet of Things (NB-IoT): From physical (PHY) and media access control (MAC) layers perspectives," *Sensors*, vol. 19, no. 11, 2019, Art. no. 2613.

[6] M. Lauridsen, H. Nguyen, B. Vejlgaard, I. Z. Kovacs, P. Mogensen, and M. Sorensen, "Coverage comparison of GPRS, NB-IoT, LoRa, and sigfox in a 7800 km^2 area," in *Proc. IEEE 85th Veh. Technol. Conf.*, 2017, pp. 1–5.

[7] S. Böcker, C. Arendt, P. Jörke, and C. Wietfeld, "Lpwan in the context of 5G: Capability of lorawan to contribute to MMTC," in *Proc. IEEE 5th World Forum Internet Things*, 2019, pp. 737–742.

[8] D. Loghin *et al.*, "The disruptions of 5G on data-driven technologies and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1179–1198, Jun. 2020.

[9] A. S. Marcano and H. L. Christiansen, "Impact of NOMA on network capacity dimensioning for 5G hetnets," *IEEE Access*, vol. 6, pp. 13587–13603, 2018.

[10] A. Shahini and N. Ansari, "NOMA aided narrowband IoT for machine type communications with user clustering," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7183–7191, Aug. 2019.

[11] A. Benjebbour, K. Saito, A. Li, Y. Kishiyama, and T. Nakamura, "Non-orthogonal multiple access (NOMA): Concept, performance evaluation and experimental trials," in *Proc. Int. Conf. Wireless Netw. Mobile Commun.*, 2015, pp. 1–6.

[12] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surv. Tut.*, vol. 19, no. 2, pp. 721–742, Apr.–Jun. 2017.

[13] A. Hoglund *et al.*, "3GPP release-16 preconfigured uplink resources for LTE-M and NB-IoT," *IEEE Commun. Standards Mag.*, vol. 4, no. 2, pp. 50–56, Jun. 2020.

[14] L. P. Qian, Z. Zhu, N. Yu, and Y. Wu, "Joint minimization of transmission energy and computation energy for mec-aware noma NB-IoT networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2019, pp. 1–7.

[15] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[16] S. Mishra, L. Salaün, and C. S. Chen, "Maximizing connection density in NB-IoT networks with NOMA," in *Proc. IEEE 91st Veh. Technol. Conf.*, May 2020, pp. 1–6.

[17] T. N. Do, D. B. da Costa, T. Q. Duong, and B. An, "Improving the performance of cell-edge users in NOMA systems using cooperative relaying," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 1883–1901, May 2018.

[18] A. Nazari, M. R. Javan, and S. S. Hosseini, "Resource allocation in power domain noma-based cooperative multicell networks," *IET Commun.*, vol. 14, no. 7, pp. 1162–1168, 2020.

[19] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for Internet of Things," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 619–623, Mar. 2021.

[20] F. AL-Tam, A. Mazayev, N. Correia, and J. Rodriguez, "Radio resource scheduling with deep pointer networks and reinforcement learning," in *Proc. IEEE 25th Int. Workshop Comput. Aided Model. Des. Commun. Links Netw.*, 2020, pp. 1–6.

[21] H. T. Cheng, A. Callard, G. Senarath, H. Zhang, and P. Zhu, "Step-wise optimal low power node deployment in LTE heterogeneous networks," in *Proc. IEEE Veh. Technol. Conf.*, 2012, pp. 1–4.

[22] P. Ameigeiras, Y. Wang, J. Navarro-Ortiz, P. E. Mogensen, and J. M. Lopez-Soler, "Traffic models impact on OFDMA scheduling design," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, pp. 1–13, Feb. 2012.

[23] D. Lee *et al.*, "Coordinated multipoint transmission and reception in LTE-advanced: Deployment scenarios and operational challenges," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[24] C. B. Mwakwata, O. Elgarhy, Y. L. Moullec, M. M. Alam, S. Pärand, and I. Annus, "Inter-cell interference reduction scheme for uplink transmission in NB-IoT systems," in *Proc. Int. Wireless Commun. Mobile Comput.*, 2021, pp. 400–405.

[25] C. B. Mwakwata, M. M. Alam, Y. Le Moullec, H. Malik, and S. Parand, "Cooperative interference avoidance scheduler for radio resource management in NB-IoT systems," in *Proc. Eur. Conf. Netw. Commun.*, 2020, pp. 154–159.

[26] K. Yang, T. Roste, F. Bekkadal, K. Husby, and O. Trandem, "Long-distance propagation measurements of mobile radio channel over sea at 2 GHz," in *Proc. IEEE Veh. Technol. Conf.*, 2011, pp. 1–5.

[27] H. Shi, R. V. Prasad, E. Onur, and I. Niemegeers, "Fairness in wireless networks: Issues, measures and challenges," *IEEE Commun. Surv. Tut.*, vol. 16, no. 1, pp. 5–24, Jan.–Mar. 2014.

Collins Burton Mwakwata received the M.Sc. degree in telecommunication and networks engineering from Université de Blida 1, Blida, Algerie, in 2017. He is currently working toward the Ph.D. degree in ICT—telecommunication with the Tallinn University of Technology, Tallinn, Estonia.

In 2018, he joined Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology. His research interests include radio resource management (interference management) in 5G and beyond-5G cellular systems, mMTC, NB-IoT, and RAN slicing.

Osama Elgarhy received the B.Sc. degree in electrical and communication engineering from Akhbar ElYom Academy, Giza Governorate, Egypt, in 2009, and the M.Sc. and Ph.D. degrees in telecommunication engineering from Politecnico Di Milano, Milan, Italy, in 2016 and 2020, respectively.

From 2009 until 2013, he was a Teaching Assistant with Pyramids Higher Institute of Engineering, Cairo, Egypt. He spent a research period in Azcom Technology, Milan, Italy, during 2016. He is currently a Postdoctoral Researcher with the Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, Tallinn, Estonia. His current research interests include resource allocation for cellular systems, NB-IoT, and device to device communications.

Muhammad Mahtab Alam (Member, IEEE) received the M.Sc. degree in electrical engineering from Aalborg University, Denmark, in 2007, and the Ph.D. degree in signal processing and telecommunications from the University of Rennes1 (INRIA Research Center), Rennes, France, in 2013.

He conducted Postdoctoral Research with the Qatar Mobility Innovations Center, Doha, Qatar, from 2014 to 2016. In 2016, he was elected as the European Research Area Chair in the cognitive electronics project and an Associate Professor with the Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, Tallinn, Estonia. In 2018, he obtained tenure professorship to chair Telia Professorship under the cooperation framework between Telia and the Tallinn University of Technology. He has authored or coauthored over 60 research publications. His research interests include self-organized and self-adaptive wireless sensor and body area networks specific to energy-efficient communication protocols and accurate energy modeling, the Internet of Things, public safety and critical networks, embedded systems, digital signal processing, and software-defined radio.

Yannick Le Moullec received the M.Sc. degree in electrical engineering from Université de Rennes I, Rennes, France, in 1999 and the Ph.D. degree in Electrical Engineering from Université de Bretagne Sud, Lorient, France, in 2003.

From 2003 to 2013, he successively held Postdoc, Assistant, and Associate Professor positions with Aalborg University, Aalborg, Denmark. He then joined Tallinn University of Technology, Tallinn, Estonia as a Senior Researcher and then on a professorship of Cognitive Electronics. He has supervised ten Ph.D. theses and 50+ M.Sc. theses. He was Co-PI for the H2020 COEL ERA-Chair project. His research interests include HW/SW codesign, embedded systems, reconfigurable systems, and IoT.

Sven Pärand received the M.Sc. degree in telecommunications in 2006 and the Ph.D. degree in electronics and telecommunication from the Tallinn University of Technology, Tallinn, Estonia, in 2018.

He has been an Engineer with Telia Estonia Ltd., Tallinn, Estonia, since 2012, initially working on IMS and migration toward the Next Generation Network. Starting form 2018, he moved on to work as a 5G development Manager with the aim of deploying the 5G network with Telia Estonia. He is currently the mobile services owner with the same company responsible for the management and development of all mobile services across all of the mobile generations.

Konstantinos Trichias received the Dipl.-Ing degree in electrical and computer engineering from the University of Patras, Greece, in 2008 and the M.Sc. degree in electrical/telecommunications engineering from the University of Twente, The Netherlands, in 2011.

He specializes on next-generation heterogeneous wireless and mobile networks, as well as the integration and smooth interoperability of the aforementioned technologies with novel networking paradigms such as SDN, ITS/V2X, and IoT, targeting the successful integration of multiple vertical industries (smart cities/industry 4.0, automotive, etc.) into the 5G ecosystem. He has participated in several (inter)national research and industry consultancy projects from multiple positions (PM, TM, QM), and is currently serving as the Technical Coordinator of the H2020-ICT-18-2018 5G MOBIX and as Project Coordinator of the H2020-ICT-41-2020 VITAL-5G projects. He has served as a 3GPP RAN1 & RAN2 delegate on behalf of KPN/TNO and has numerous patent applications in the area of radio access systems.

Kostas Ramantas received the Diploma degree in computer engineering, the M.Sc. degree in computer science and engineering, and the Ph.D. degree in computer engineering from the University of Patras, Patras, Greece, in 2006, 2008, and 2012, respectively.

In June 2013, he joined IQUADRAT Informatica, Barcelona, Spain, as a Senior Researcher, where he is actively involved in EU-funded research projects. His research interests include modeling and simulation of network protocols, and scheduling algorithms for QoS provisioning.

Dr. Ramantas has been a recipient of two national scholarships and has participated in the EC funded ICT-BONE and ePhoton/One+ Networks of Excellence, conducting joint research with many European research groups.

Appendix 5

C.B. Mwakwata, M.M. Alam, Y. Le Moullec, . Performance enhancement of RAN Slice scheduling by using Machine Learning Techniques adapted for beyond 5G wireless Networks. IEEE Systems Journal, IEEE, Under Review, 2022.

mMTC Users Classification Empowering a Predictive Cooperative Scheduler in RAN Slicing for 5G and Beyond Networks

Collins Burton Mwakwata, Muhammad Mahtab Alam, *Senior Member, IEEE*, Yannick Le Moullec, *Senior Member, IEEE*, and Sven Pärand

***Abstract*— Radio resource management in the frames of radio access network (RAN) slicing is an emerging research topic. The advancements in RAN scheduling taking into account the users' expected quality of service (QoS) and interference management both in orthogonal and non-orthogonal multiple access (OMA, NOMA) techniques lead to challenging open research problems for massive machine-type communications (mMTC) in 5G and beyond networks. In this paper, we introduce machine learning (ML)-based mMTC users classification and prediction to minimize the impact of interference within a RAN slice. The applied ML algorithms enhance the scheduling performance by utilizing the user equipment (UE) periodicity and buffer size requirement for their next transmissions.**

The predictive algorithm results show that the allocation of radio resources for the future scheduling frame can reduce the unnecessary packet transmissions and enhance the management of slices by providing the best slice configuration for the network at a given scheduling frame. This yields reduced energy consumption with better data rates according to the service level agreement (SLA) of a given slice. As compared to benchmark traditional OMA and cooperative OMA, our ML-enabled scheduler increases throughput by 100% and 50%, decreases energy consumption by two to three times, and increases the number of supported devices by 60% and 23% and 160% and 50% for cell center and cell edge, respectively. It attains comparable throughput and number of supported devices than cooperative NOMA, but with three times lower energy consumption and without the cooperative communication overheads.

***Index Terms*— RAN Slicing, Machine Learning, mMTC, 5G Scheduling, beyond 5G networks**

I. Introduction

The fifth-generation (5G) and beyond 5G mobile networks are designed to cope with the increasing demand for connectivity to support use cases requiring human-to-human and machine-to-machine communications [1]. However, unlike legacy technologies, 5G and beyond 5G networks bring into play specific service verticals, namely massive machine-type communications (mMTC), enhanced mobile broadband (eMBB), and ultra-reliable low latency communications (uRLLC) [2]. These verticals have different quality of service (QoS) requirements but utilize the same physical infrastructure [3].

For example, one of the main goals of mMTC is to support a higher connection density of up to billions of devices per square kilometer [4], [5] for beyond 5G networks. eMBB aims to deliver peak data rates of up to 20 Gbps, peak spectral efficiency of up to 30 bps/Hz and 15 bps/Hz on the downlink and uplink, respectively, maximum tolerable latency of 4 ms, and high energy efficiency with seamless mobility support [6]. Finally, uRLLC provides up to only 1 ms user-plane latency, reliability of $1 - 10^{-5}$ success probability for transmitting 32 bytes in 1 ms with 0 ms mobility interruption time [7]. In this regard, it is very challenging to support this diversity of service requirements under the same physical infrastructure without interrupting the services that are simultaneously running on the same network.

To cope with such a challenge, RAN slicing has been proposed as one of the enablers of network orchestration by flexibly customizing and managing the base stations utilizing softwarization and virtualization to support a multi-service-multi-tenant architecture [8]. In this regard, a RAN slice can be characterized by specific QoS requirements that necessitate a particular system behavior to support a given application. For example, user equipment (UE) with reduced capabilities (REDCAP) operating under 5G NR Light [9], or narrow-band IoT (NB-IoT) can be served by a slice with radio access that is up to 10 seconds delay-tolerant with very limited or no mobility. On the other hand, Cat-M UEs can be served by a slice that guarantees mobility support for applications such as

This project has received funding partly from The European Union's Horizon 2020 Research and Innovation Program under Grant 668995 and Grant 951867 and European Union Regional Development Fund in the framework of the Tallinn University of Technology Development Program 2016-2022 and Estonian research Council grant PRG667 and TAR16013 Center of Excellence 'EXCITE IT'. The authors express their gratitude to Priit Roosipuu for collecting the 5G data.

C. B. Mwakwata is with Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, 12616, Tallinn, Estonia (e-mail: collins.burton@taltech.ee).

M. M. Alam is with Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, 12616, Tallinn, Estonia (e-mail: muhammad.alam@taltech.ee).

Y. Le Moullec is with Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, 12616, Tallinn, Estonia (e-mail: yannick.lemoullec@taltech.ee).

S. Pärand is with Telia Estonia Ltd, 10616, Tallinn, Estonia (e-mail: sven.parand@telia.ee).

smart logistics [10], [11].

The recent literature discusses the advancement of RAN slicing to support massive connectivity utilizing different techniques as follows. In [12], the authors studied RAN slicing for massive connectivity of IoT applications by optimizing the random access (RA) procedure to maximize the success probabilities and improve service multiplexing of mMTC and uRLLC traffic to save unnecessary energy consumption.

In [13], the authors adopted reinforcement learning to dynamically tune the discontinuous reception parameters to enhance the radio resource control (RRC) layer in the RAN environment by implementing their proposed architecture which is built on an open-source software platform (OAI) to create, modify and delete slices in the RRC layer to satisfy the diverse service requirements needed for IoT devices.

In [14], the authors investigated network slicing in virtualized wireless networks to solve the spectral efficiency problem by proposing a resource allocation algorithm to enhance uRLLC reliability. Even though the authors focused on eMBB and uRLLC, their work provides a framework suitable for allocation problems as compared to adaptive particle swarm optimization (APSO), equal power allocation (EPA), and equal sub-carrier allocation (ESA).

In [15], the authors revisited their previously proposed functional framework for the next generation RAN slice management to incorporate the recently specified principles and features of the new service-based management architecture in 3GPP Release 15 specifications. Furthermore, the authors presented the specific management object classes and attribute to enhance the provisioning of RAN slices and the applicability of the overall functional framework and information models in an illustrative next-generation RAN architecture. Specifically, the models are used to represent the manageable aspects of a sliced next-generation RAN operated by a neutral host provider and how the proposed functional framework operates through two examples: one illustrating the provisioning of a new RAN slice and another describing how the configuration of already activated RAN slices is modified in response to traffic demand variations.

In [16], the authors investigated the feasibility of the non-orthogonal RAN resource allocation on the uplink transmission of mMTC, eMBB, and URLLC from a common base station. Their study shows that the proposed heterogeneous non-orthogonal multiple access (H-NOMA) that involves UEs with heterogeneous service requirements can lead to significant performance improvement as compared to traditional orthogonal slicing. The enhancements are enabled by the capability of H-NOMA to provide service isolation, hence ensuring required performance thresholds for all services by leveraging their heterogeneous reliability requirements.

Moreover, several topics related to service level agreement (SLA) and the corresponding radio resource management techniques have been discussed in [17], [18]. The possibility of using machine learning techniques to enhance the RAN slicing is presented in [19]–[21].

Radio resource scheduling is one of the most proposed approaches to enable optimal resource usage and to enhance the massive connectivity of mMTC UEs. However, if the radio resources are not well managed, the massive connectivity causes massive interference between UEs that are competing for the available radio resources in the heterogeneous network. One of the promising techniques to enable better scheduling by minimizing the inter-cell interference and guaranteeing the required QoS is the use of a cooperative scheduler where the adjacent base stations cooperate to schedule their UEs for their respective transmissions as presented in [22], [23].

However, sharing of the scheduling tables increases the overhead in the X2 interface, and the shared data need the brute-force computation to select the best pair to be scheduled at given radio resources. Therefore, it is necessary to study the proactive data-driven approaches that can reduce the computational complexity as well as proactively classify and predicts the next transmissions to enable proactive scheduling.

It should be noted that less attention is given to how the machine learning techniques can be used to classify and predict the users' transmission characteristics, hence enhancing the scheduling of RAN slices. In this regard, our work studies in detail the applicability of machine learning algorithms adapted to slice scheduling to increase the number of connected devices per slice, while providing the expected QoS requirements according to shared SLA.

Therefore, in contrast to the previous studies, this work presents the following contributions.

- First, we study the feasibility of different machine techniques to enhance the intra-slice scheduling of massive IoT RAN slices by using real-time data that is collected from a live $5G$ network , and select the most suitable one for our data set.
- Second, we implement the selected machine learning algorithm on a given set of network parameters to classify and predict the common patterns that can enhance the slice scheduling.
- Third, we perform intra-slice UE scheduling for the massive IoT RAN slice to enhance the management and performance by predicting the transmission periodicity and the required buffer size based on the collected real-time data.
- Finally, we evaluate the performance enhancements by comparing our proposed ML-enabled scheduler with benchmark traditional OMA as well as cooperative OMA and NOMA schemes.

To the best of the authors' knowledge, this is the first work that studies the applicability of available machine learning algorithms on collected real-time 5G networks parameters to enhance the scheduling of RAN slices for 5G and beyond 5G networks.

The rest of the paper is organized as follows: Section II presents the system modeling. Section III presents the proposed machine learning-enabled scheduler. Section IV presents the numerical results and discussion. Finally, concluding remarks are drawn in Section V.

II. System modeling

1) Problem Formulation: We consider a multi-cell network structure where several UEs are transmitting to their corresponding base stations. In this scenario, adjacent base stations simultaneously receive the unwanted signals transmitted by adjacent cell UEs. In this regard, inter-cell interference in terms of transmit power is experienced. We assume the scheduling is performed for a given slice; however, the analysis can be replicated for several slices. Let $z = \{1, 2, \ldots, Z\}$ be the index of the resource units. x_c represents the cell c's UEs, and C, i.e $c = \{1, 2, \ldots, C\}$, is the number of cells used in the simulation. The achieved data rate of a given slice is derived from the Shannon formula of cell capacity given by

$$R_{SLC} = B_{SLC} log(1 + SINR_{SLC}) \tag{1}$$

where R_{SLC} is the achieved rate of a given slice based on the SLA. B_{SLC} is the allocated bandwidth for a given slice in order to satisfy the expected QoS, given by

$$B_{SLC} = \frac{B}{\mu} \tag{2}$$

where B is the overall system bandwidth, and μ is the bandwidth splitting coefficient that depends on the SLA. Finally, $SINR_{SLC}$ is the achieved signal to noise plus interference ratio of the allocated UE, given by

$$SINR_{SLC} = \frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N} \tag{3}$$

where $|h^z_{x_c,c}|^2$ represents the channel gain of the transmitting UE, and $P^z_{x_c}$ is the transmitting power of the allocated UE, which is subject to maximum allowed power constraint per each transmitting UE given by

$$0 \leq P^z_{x_c} \leq P_{maxm}, \forall c \in C. \tag{4}$$

Therefore, we aim to maximize the sum rate of the system by maximizing the rate of each allocated slice while guaranteeing the expected QoS of each slice. In this regard, the sum rate maximization problem can be represented as:

$$\max \sum_{c \in C} \sum_{z \in Z} a^z_{x_c} \log\left(1 + \frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N}\right) \tag{5}$$

Subject to:

$$SINR_{SLC} \geq \vartheta_{x_c,SLC} \tag{6}$$

i.e.,

$$a^z_{x_c}\left(\frac{|h^z_{x_c,c}|^2 P^z_{x_c}}{\sum_{l \neq c, l \in C} \sum_{q \in Q_l} |h^z_{q_l,c}|^2 P^z_{q_l} a^z_{q_l} + \sigma_N}\right) \geq \vartheta_{x_c,SLC} \tag{7}$$

$$0 \leq P^z_{x_c} a^z_{x_c} \leq P_{maxm}, \forall c \in C, \forall x_c \in X_c, \forall z \in Z. \tag{8}$$

$$\sum_{x_c \in X_c} a^z_{x_c} \leq 1, \forall z \in Z, c \in C \tag{9}$$

where $\vartheta_{x_c,SLC}$ is the $SINR$ constraint to satisfy the required QoS of a given slice. It is considered that only the UEs above this threshold can guarantee successful transmission.

2) Proposed Solution: As it is seen above, the problem is of a mixed binary integer non-linear programming nature, i.e. it is very challenging to maximize the sum-rate while minimizing the level of acceptable interference for the scheduled UE to satisfy the expected QoS requirement. In this regard, we apply the solutions to minimize the interference between allocated UEs, as proposed in [22]. Then to optimize the allocation matrix, we implement the machine learning schemes, hence allocating the resources based on classification and prediction. Finally, we perform power allocation to further minimize the unnecessary energy consumption of the transmitting UEs. The corresponding discussion of machine learning and the enhanced scheduler setup are presented in the following section.

III. Proposed Approach for Users Classification and Prediction

To begin with, the IoT UEs' channel parameters data are collected from a live 5G network; in our case in Haapsalu, Estonia, for which the heat map is shown in Fig. 1. Then data processing is performed to eliminate the coverage holes in areas where no actual communication parameters data was collected. Next, we run different machine learning algorithms to classify the UEs according to their corresponding channel parameters (i.e., time stamp, latitude, longitude, EARFCN, NRSSI, NRSRP, NRSRQ, NSINR, Tx Power, etc.). Depending on the parameters that are needed for the system design, one can choose what parameters to be used as input parameters and which ones as output parameters. For our proposed scheduler, we used time stamp, latitude, longitude, EARFCN, NRSSI, NRSRP, Tx Power, NRSRQ, as input parameters and NSINR, as output parameter (i.e., we later use NSINR values for setting the performance threshold during MATLAB simulations). We deploy several machine learning algorithms on the processed data sets to classify the users in different clusters to enhance the UE scheduling; several output parameters such as minimum classification errors, true response vs. prediction response, etc. are used to judge the quality of the classification performance. For the above-specified data set, we present the best-performing algorithms as compared to all possible lightweight ML algorithms. From the analysis presented later in Section IV-A, it can be noted that rational quadratic Gaussian Process Regression (GPR) performs better than the fine Gaussian Support Vector Machine (SVM) when compared to the perfect prediction on the collected samples. Then, based on the collected insights from both classification and regression algorithms, the intra-slice scheduler was designed to predict the periodicity and buffer size for the next scheduling frames.

A. Intra-Slice Scheduling

Since our objective is to optimize the performance of massive IoT slices, the simulation is performed to map the collected real-time channel parameters to enhance the number of connected devices by predicting the UE periodicity and expected buffer size for the next frame. In this regard, unnecessary radio resources are released to support other slices that require higher bandwidth and/or transmission time slots.

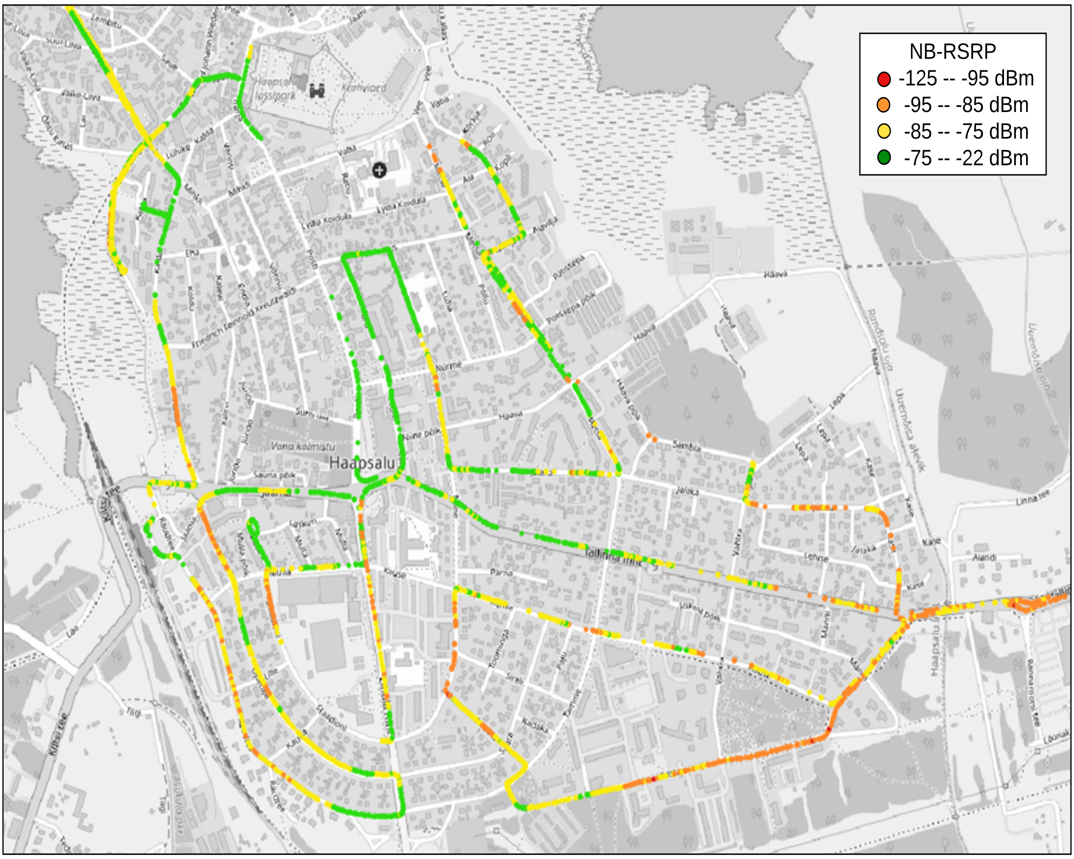

Fig. 1. The heat map representing the data collection campaign in Haapsalu town, Estonia. The channel parameters that were collected from NB-IoT UE include: time stamp, latitude, longitude, EARFCN, NRSSI, NRSRP, NRSRQ, NSINR, Tx Power, etc.

From the allocated and released resources, the performance of the network is analyzed to evaluate the effectiveness of the proposed algorithm in comparison to traditional scheduling algorithms that utilize fixed radio resources for a given expected QoS requirement. It should be noted that inter-slice scheduling is out of scope for this work, but will be included in our future outlook.

The overall proposed framework is presented in Fig 2 and summarized below.

- The optimization parameter is selected in the Network Slice Sub-net Management Function (NSSMF) which acts as the brain of the slice, where slice function selection, configuration, and coordination are originated.
- The NSSMF decides to instantiate, scale, terminate or move the slice based on the commands it receives from the Network Slice Management Function (NSMF), which receives the translation of related service requirements from the Communication Service Management Function (CSMF).
- Then the machine learning algorithm is applied to the collected data from the 5G network. Based on the nature of the data, UE or network parameters cleaning is performed because some might be missing due to coverage holes or temporary UE disconnection from the network.
- The classification and prediction are performed on the data to give the knowledge on the behavior of UE hence the controller and the RAN scheduler coordinate and cooperate to allocate the optimal radio resources to the massive IoT slice to satisfy the required quality of service.
- When the UEs are allocated to transmit on the network, either the UE or the network parameters are monitored and the performance is evaluated; in case of a completely new parameter or a significant change in current collected values, the changed parameter is injected back to be used in the machine learning to decide whether it can bring more enhancements to the slice performance in

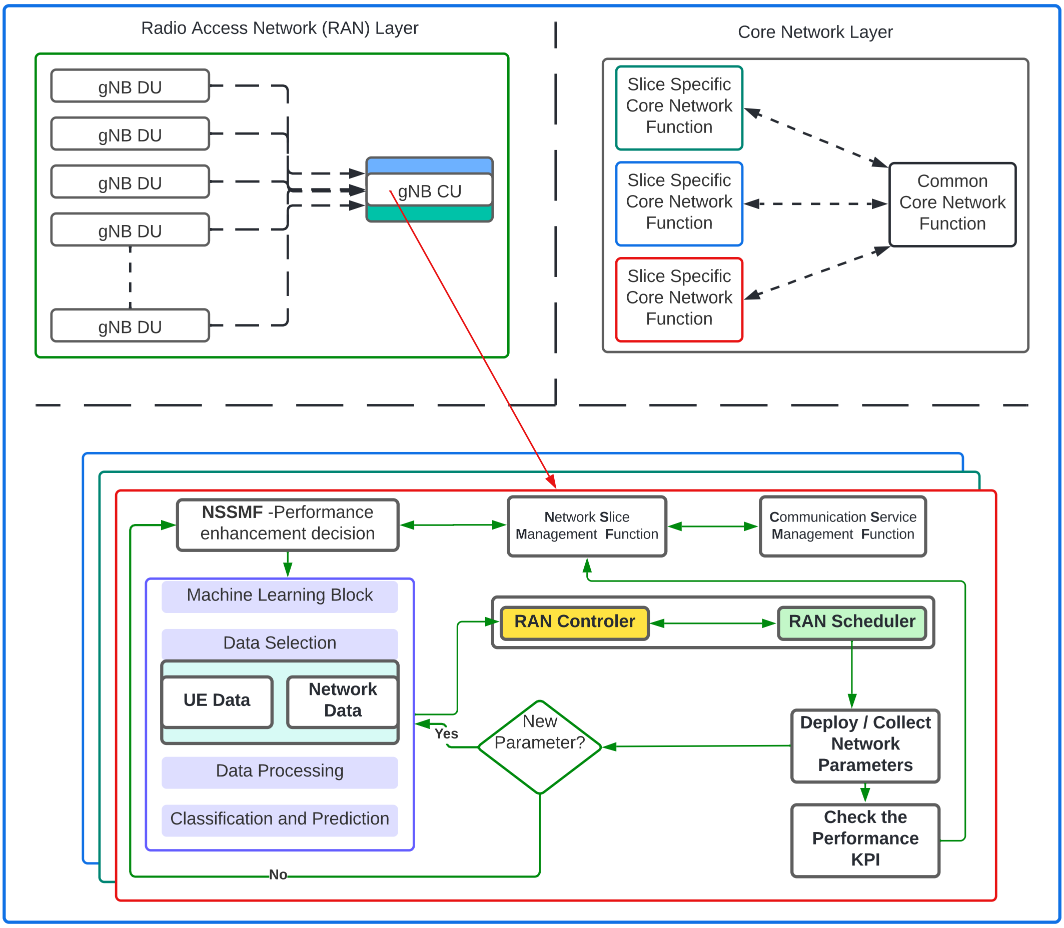

Fig. 2. The proposed framework that utilizes machine learning to perform the user clustering and prediction to enhance the RAN slice scheduling for 5G and beyond networks

terms of UE energy consumption, throughput, number of connected devices, etc.

B. Machine Learning Enabled Cooperative Scheduler

Our starting point is a cooperative scheduler designed to mitigate the impact of inter-cell interference caused by transmitting users from adjacent cells. The overall scheduling framework and its settings are presented in [22]. This scheduler is comprised of sharing the scheduling tables between the base stations to mitigate the inter-cell interference by proactively allocating the radio resources to the users' combinations that result in minimum possible interference to maximize the expected quality of service requirements. However, to reduce the overhead in the X2 interface due to the need for information sharing between the base stations, we deploy the machine learning framework presented in Fig. 2 to perform the prediction of the base stations' next transmission capabilities within a given slice. In doing so, the machine learning scheme is utilized in all the cooperating base stations for clustering not only the current transmissions but also the prediction of the corresponding upcoming transmissions of a given set of UEs in a given slice, hence reducing unnecessary exchange of UEs which can not successfully transmit in next frame. The overall scheduler can be visualized as in Fig. 3.

For example, given the learned transmission pattern and the SINR distribution, a certain number of UEs can tolerate a 10 ms delay for a small packet size transmission that can occupy less bandwidth as compared to a set of UEs that require a 1 ms delay for the same packet size. In this regard, latency-sensitive UEs can occupy and release a given radio resource faster and

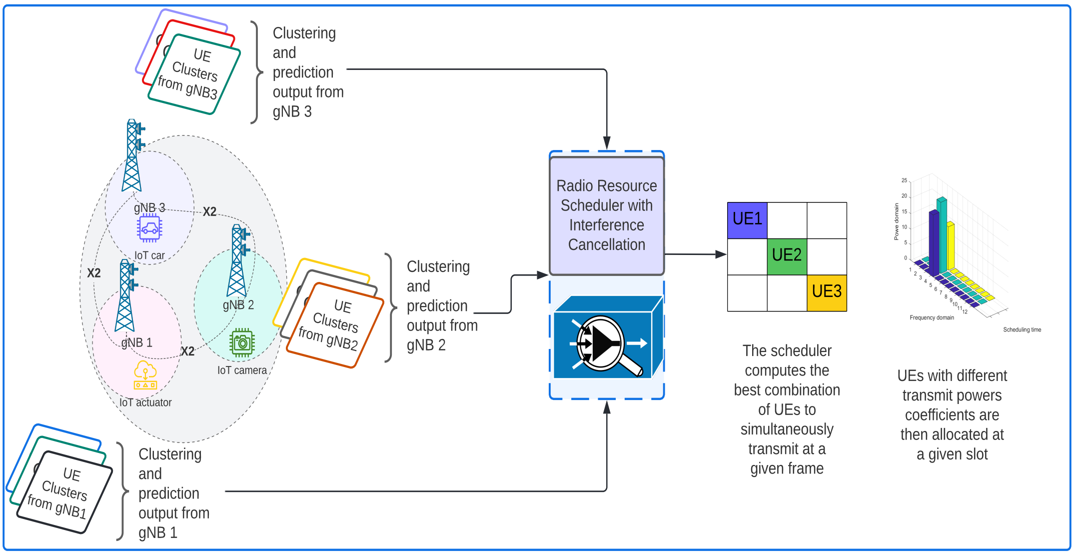

Fig. 3. The proposed ML-enabled scheduler that utilizes the machine learning output (i.e. users classification and transmission prediction) from the cooperating base stations to allocate the radio resources to the UEs with minimum possible interference

let the slice be used by the latency insensitive UEs while the transmission periodicity is being monitored.

The simulation is performed in MATLAB to adapt different scheduling frameworks based on SINR, and the simulation parameters as used in [24]. We adapt the scheduling on the traditional and cooperative OMA schemes; however, for comparison purposes, the cooperative NOMA scheme is also considered to further compare the performance gains. In the current work, the machine learning framework is not adapted to the NOMA scheme due to its already relatively high computational complexity when performing the successive interference cancellation for the UE allocated at the same radio resources.

The next section presents the performance enhancements when ML is used in RAN scheduling. Furthermore, the scheduling schemes are compared and the improvements are displayed, and additional discussion is given on the complexity analysis when machine learning is used.

IV. Performance evaluation

A. Evaluation of Classification and Prediction Accuracy

We implement supervised learning algorithms (i.e., the algorithms analyze the input parameters and provide the function which represents the relationship between input and output parameters). For assessing classification performance, we use minimum classification error (MCE) since it is the most commonly used criterion for pattern classification which tries to reduce the overall classification error when classifying a given group of network parameters [25]. For example, from the analysis, it is seen that rational quadratic Gaussian Process Regression (GPR) performs better than the fine Gaussian support vector machine (SVM) when performing regression. This is due to the nature of the collected data. It should be noted that, with different data set, different machine learning algorithms might perform better; however, for our analysis, the focus is on how the machine learning algorithms could give better insights to different data types to enhance the UEs' scheduling.

Furthermore, the Fine Gaussian SVM experiences a Root Mean Square Error (RMSE) of 0.453; however, the Rational Quadratic GPR experiences an RMSE of 0.28. In this regard, it is evident that the Rational Quadratic GPR is the candidate machine learning algorithm for the data we collected from the $5G$ network. However, it should be noted that, depending on the application, the training time becomes a critical constraint because providing higher accuracy means increasing the training time needed to fine-tune the selected parameters.

Another criterion used to assess the performance of the machine learning is prediction error performance as seen in Fig. 4 and Fig. 5. It can be noted that the predictions are evenly distributed around the perfect prediction line; in this regard, the Rational Quadratic GPR better predicts the UE parameters as compared to SVM, hence can be used for their next scheduling enhancements by allocating the resources to UE based on the predicted transmission patterns.

In general, the choice of the machine learning algorithm for either classification or prediction depends on the nature of the data set, the computation power of the system, the overall acceptable training time by the system, accuracy of the given output, number of features, the possibility to interpret, etc.

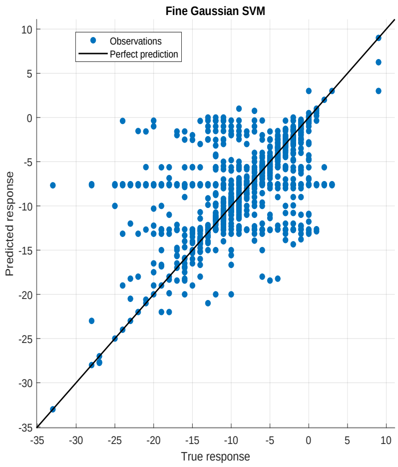

Fig. 4. Performance of Fine Gaussian SVM in terms of predicted response vs. true response

For example, from our study, the size of the collected data is small due to the limited number of deployed 5G networks, with limited IoT devices operating under cellular networks for the specific trial use-case. In this regard, it is imperative to choose the algorithms with low variance and higher bias. For the studies which include big data sets with a higher number of observations, it is imperative to apply algorithms with low bias and with high variance such as Kernel SVM, K-Nearest Neighbors (KNN), etc.

B. Complexity analysis of Proposed Scheme

In our study, we use the prediction accuracy to decide which machine learning algorithm is to be used for which scenario for the massive IoT slice. Additionally, the total training time is used as a very critical scheduling constraint because longer training times are acceptable for delay-tolerant applications, and shorter training times are essential for delay-sensitive applications. In this regard, we analyze the experienced trade-off between prediction accuracy and total training time for massive IoT slices to enhance intra-slice scheduling. Therefore, from the presented machine learning algorithms in our study, we chose the Rational Quadratic GPR as the best candidate to cluster and predict according to the collected data set.

When analyzing the computational complexity of the proposed schemes, it should be noted that in machine learning, this is somewhat subjective because machine learning complexity mainly refers to the number of features implemented in a given predictive model and whether that model is linear, nonlinear, logarithmic, exponential, etc. However, the overall computation complexity of the scheduler is presented in [22].

C. Evaluation of the ML-enabled Scheduler Performance

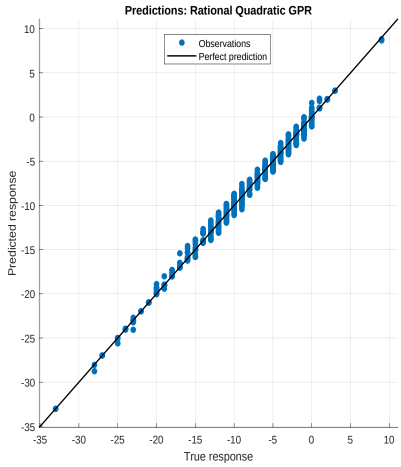

Fig. 5. Performance of Rational Quadratic GPR in terms of predicted response vs. true response

We consider that the radio resources are allocated to the UE for their corresponding future transmissions. Based on the UE clusters, the radio resources are allocated accordingly, i.e., latency-sensitive UEs are given the radio resources within a sub slice that accommodate higher bandwidth but with shorter transmission times, and fewer supported repetitions, and latency-insensitive UEs are given less bandwidth but with longer transmission times, and more repetitions. In this regard, contrary to the orthogonal approach of fixed radio resources to accommodate all UE types, in our proposed framework, the total available bandwidth for massive IoT slice is dynamically sub-divided to support diverse UE requirements into different sub-slices.

The overall description and simulation parameters of the benchmark schedulers and their thorough analysis are presented in [22].

In Fig. 6, it can be seen that the ML-enabled scheduler attains near identical throughput performance than the cooperative NOMA scheduler. This is due to the possibility of classifying the UEs based on their channel parameters, hence proactively scheduling to mitigate interference and optimize the throughput performance. It should be noted that this comparable throughput performance is achieved without the communication overheads (for the exchange of scheduling tables) needed in cooperative NOMA. On the other hand,

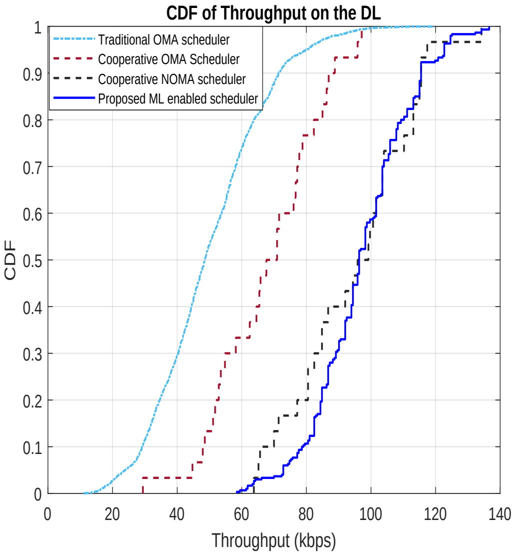

Fig. 6. Comparison of UE achieved data rates between the ML-enabled scheduler and benchmark OMA and NOMA scheduling schemes

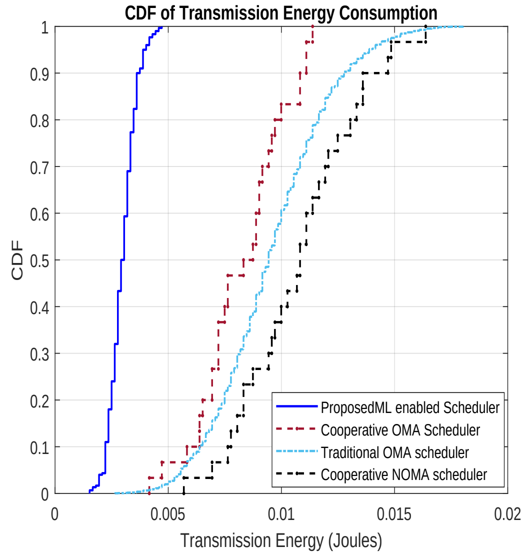

Fig. 7. Comparison of the average UE energy consumption per transmission between the ML-enabled scheduler and benchmark traditional OMA scheme and cooperative OMA and NOMA scheduling schemes

the ML-enabled scheduler provides approximately $100\ \%$ improvement in terms of throughput as compared to the traditional OMA scheduler presented in the literature, and up to $49.2\ \%$ throughput improvement as compared to the cooperative OMA scheduler that utilizes the cooperative scheduling between the base stations to mitigate the impact of the inter-cell interference.

In Fig. 7, it can be seen that the ML-enabled scheduler yields lower energy consumption as compared to all other benchmark schedulers thanks to the classification and prediction that enhances the scheduling by allocating the resources to UE only when needed, hence preventing unnecessary energy consumption especially when the UE channel conditions do not permit successful transmissions hence triggering more re-transmissions. It can be noted that the UEs under the ML-enabled scheduling experience up to 3 times less energy consumption as compared to the cooperative NOMA scheduler, and 2 to 3 times less energy consumption as compared to the cooperative OMA and traditional OMA schedulers, respectively.

Finally, Fig. 8 presents the number of connected devices when the ML-enabled scheduler is used. It can be noted that our proposed ML-enabled scheduler outperforms both the traditional OMA and cooperative OMA schemes by 60% and 23%, in terms of the number of connected UE for cell-center users and by almost 165% and 50% respectively, for the cell-edge users.

Additionally, it can be noted that our proposed ML-enabled scheduler performs slightly below (-7%) the achieved capacity of the cooperative NOMA scheduler for the cell center users and performs slightly better (10%) than the proposed NOMA scheme for the cell edge users. It should again be kept in mind that those slightly lower and higher results are achieved without the communication overheads for the exchange of scheduling tables used in cooperative NOMA. These overall enhancements are achieved thanks to the classification and prediction that is provided by machine learning to enhance the overall scheduling by proactively allocating the radio resources according to the predicted transmission patterns to minimize the risks of UE falling into the outage.

V. Conclusion

This work studied the feasibility of a machine-learning algorithm to enhance the RAN slice scheduling for beyond 5G networks. From the analysis, it can be noted that, depending on the nature of the data, possibility to interpret, speed of training, etc., different machine learning algorithms can be applied on either UE or network data to enhance the RAN scheduling for massive connectivity support of massive IoT slices. It is observed that the ML-enabled scheduler outperforms the benchmark scheduling schemes significantly. When compared to traditional OMA and cooperative OMA, our ML-enabled scheduler increases throughput by 100% and 50%, decreases energy consumption by 2 times to 3 times, and increases the number of supported devices by 60% and 23% and 160% and 50% for cell center and cell edge, respectively. Our ML-enabled scheduler attains comparable throughput and number of supported devices than cooperative NOMA, but with 3

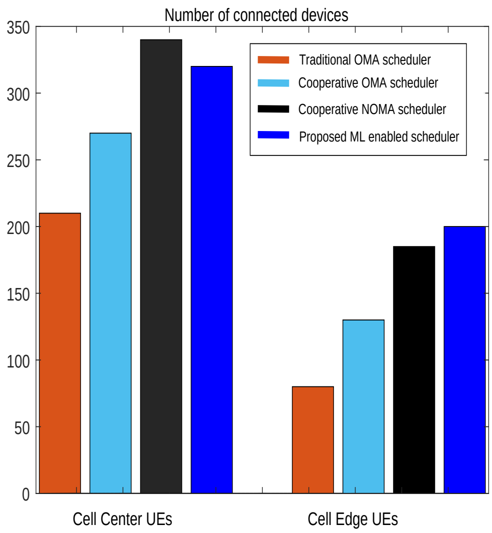

Fig. 8. Comparison of number of connected devices when different scheduling schemes are used. The ML-enabled scheduler outperforms all the OMA schedulers for cell center and cell edge UEs

times lower energy consumption and without the cooperative communication overheads.

Our future outlook aims to implement advanced techniques such as intelligent reflecting surfaces (IRS) and directional beamforming and deep learning approaches to increase the cell capacity for the massive IoT devices in 5G and beyond networks. Additionally, our future study includes inter-slice scheduling to maximize the spectrum efficiency for massive IoT applications that require different QoS under heterogeneous slice configurations.

REFERENCES

[1] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[2] S. Li, L. D. Xu, and S. Zhao, "5g internet of things: A survey," *Journal of Industrial Information Integration*, vol. 10, pp. 1–9, 2018.

[3] L. Chettri and R. Bera, "A comprehensive survey on internet of things (iot) toward 5g wireless systems," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 16–32, 2020.

[4] S. Alraih, I. Shayea, M. Behjati, R. Nordin, N. F. Abdullah, A. Abu-Samah, and D. Nandi, "Revolution or evolution? technical requirements and considerations towards 6g mobile communications," *Sensors*, vol. 22, no. 3, 2022.

[5] O. Liberg, M. Sundberg, Y.-P. E. Wang, J. Bergman, J. Sachs, and G. Wikström, "Chapter 16 - choice of iot technology," in *Cellular Internet of Things (Second Edition)* (O. Liberg, M. Sundberg, Y.-P. E. Wang, J. Bergman, J. Sachs, and G. Wikström, eds.), pp. 687–707, Academic Press, second edition ed., 2020.

[6] Q. Chen, J. Wang, and H. Jiang, "Urllc and embb coexistence in mimo non-orthogonal multiple access systems," 2021.

[7] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, "5g urllc: Design challenges and system concepts," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–6, 2018.

[8] A. Ksentini and N. Nikaein, "Toward enforcing network slicing on ran: Flexibility and resources abstraction," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 102–108, 2017.

[9] M. Pagin, T. Zugno, M. Giordani, L.-A. Dufrene, Q. Lampin, and M. Zorzi, "Can 5g nr-light operate at millimeter waves? design guidelines for mid-market iot use cases," 2021.

[10] X. Foukas, M. K. Marina, and K. Kontovasilis, "Orion: Ran slicing for a flexible and cost-effective multi-service mobile network architecture," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, MobiCom '17, (New York, NY, USA), p. 127–140, Association for Computing Machinery, 2017.

[11] S. E. Elayoubi, S. B. Jemaa, Z. Altman, and A. Galindo-Serrano, "5g ran slicing for verticals: Enablers and challenges," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 28–34, 2019.

[12] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55765–55779, 2018.

[13] L. Cao, R. Li, J. Crowcroft, Z. Zhao, and H. Zhang, "Intelligent slicing of radio resource control layer for cellular iot: Design and implementation," 2020.

[14] T. Ma, Y. Zhang, F. Wang, D. Wang, and D. Guo, "Slicing resource allocation for embb and urllc in 5g ran," *Wireless Communications and Mobile Computing*, vol. 2020, p. 6290375, Jan 2020.

[15] R. Ferrús, O. Sallent, J. Pérez-Romero, and R. Agustí, "On the automation of ran slicing provisioning: solution framework and applicability examples," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, p. 167, Jun 2019.

[16] P. Yang, X. Xi, T. Q. S. Quek, J. Chen, X. Cao, and D. Wu, "Ran slicing for massive iot and bursty urllc service multiplexing: Analysis and optimization," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14258–14275, 2021.

[17] P. Caballero, A. Banchs, G. de Veciana, and X. Costa-Pérez, "Multi-tenant radio access network slicing: Statistical multiplexing of spatial loads," *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 3044–3058, 2017.

[18] B. Khodapanah, A. Awada, I. Viering, D. Oehmann, M. Simsek, and G. P. Fettweis, "Fulfillment of service level agreements via slice-aware radio resource management in 5g networks," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pp. 1–6, 2018.

[19] B. Han and H. D. Schotten, "Machine learning for network slicing resource management: A comprehensive survey," 2020.

[20] M. R. Raza, C. Natalino, P. Öhlen, L. Wosinska, and P. Monti, "Reinforcement learning for slicing in a 5g flexible ran," *J. Lightwave Technol.*, vol. 37, pp. 5161–5169, Oct 2019.

[21] Y.-J. Liu, G. Feng, Y. Sun, S. Qin, and Y.-C. Liang, "Device association for ran slicing based on hybrid federated deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15731–15745, 2020.

[22] C. B. Mwakwata, O. Elgarhy, M. M. Alam, Y. Le Moullec, S. Pärand, K. Trichias, and K. Ramantas, "Cooperative scheduler to enhance massive connectivity in 5g and beyond by minimizing interference in oma and noma," *IEEE Systems Journal*, pp. 1–12, 2021.

[23] H. Malik, H. Pervaiz, M. Mahtab Alam, Y. Le Moullec, A. Kuusik, and M. Ali Imran, "Radio resource management scheme in nb-iot systems," *IEEE Access*, vol. 6, pp. 15051–15064, 2018.

[24] C. B. Mwakwata, M. M. Alam, Y. Le Moullec, H. Malik, and S. Parand, "Cooperative interference avoidance scheduler for radio resource management in nb-iot systems," in *2020 European Conference on Networks and Communications (EuCNC)*, pp. 154–159, 2020.

[25] M. Matton, D. Van Compernolle, and R. Cools, "Minimum classification error training in example based speech and pattern recognition using sparse weight matrices," *Journal of Computational and Applied Mathematics*, vol. 234, no. 4, pp. 1303–1311, 2010. Proceedings of the Thirteenth International Congress on Computational and Applied Mathematics (ICCAM-2008), Ghent, Belgium, 7–11 July, 2008.

Collins Burton Mwakwata received MSc degree in telecommunication and networks engineering from Université de Blida 1 (Algerie) in 2017. He previously held different engineering roles in Ericsson Estonia including 5G Test Engineer, 5G Test software developer, Product Engineer for 5G massive MIMO and Test Product Owner for 5G mmWave products. Since 2018 he is with Thomas Johann Seebeck department of Electronics at Tallinn university of Technology, Tallinn (Estonia) where he is currently pursuing a PhD in ICT – Telecommunication. His research interests include radio resource management, mainly interference management for massive mMTC deployment in 5G and beyond-5G cellular systems, and radio access networks (RAN) slicing for beyond 5G networks.

Muhammad Mahtab Alam (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from Aalborg University, Denmark, in 2007, and the Ph.D. degree in signal processing and telecommunication from the INRIA Research Center, University of Rennes 1, France, in 2013. He joined the Swedish College of Engineering and Technology, Pakistan, in 2013, as an Assistant Professor. He did his postdoctoral research at the Qatar Mobility Innovation Center, Qatar, from 2014 to 2016. In 2016, he joined as the European Research Area Chair holder and an Associate Professor with the Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, where he was elected as a Professor, in 2018. In 2019, he become the head of Communication Systems Research Group and leading number of National and International projects. His research interests include the fields of wireless communications and connectivity, NB-IoT 5G/B5G smart networks and services, and low-power wearable networks for SmartHealth.

Yannick Le Moullec (Senior Member, IEEE) received the M.Sc. degree from Université de Rennes I, France, in 1999, and the Ph.D. and HDR (accreditation to supervise research) degrees from Université de Bretagne Sud, France, in 2003 and 2016, respectively. From 2003 to 2013, he successively held Postdoctoral Researcher, Assistant Professor, and Associate Professor positions with the Department of Electronic Systems, Aalborg University, Denmark. He then joined Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, Estonia: Senior Researcher (2013 to 2016) and professorship (since 2017). His research interests include embedded systems, reconfigurable systems, the IoT, and the application thereof. He has supervised or co-supervised more than 50 M.Sc. students and 11 Ph.D. students. He has been involved in more than 20 projects, including five as PI, co-PI, or co-main applicant; one such notable project was the H2020 COEL ERA-Chair project from 2015 to 2019. He is an IEEE Senior Member, and a member of the IEEE Sustainable ICT Technical Community and of the IEEE Circuits and Systems Society.

Sven Pärand received both his M.Sc. degree (2006) in telecommunications and a Ph.D. degree in electronics and telecommunication from Tallinn University of Technology, Tallinn, Estonia, in 2018. He has been an engineer at Telia Estonia Ltd since 2012, initially working on IMS and migration towards the Next Generation Network. Starting form 2018 he moved on to work as a 5G development manager with the aim of deploying the 5G network at Telia Estonia. He is currently the mobile services owner at the same company responsible for the management and development of all mobile services across all of the mobile generations.

Curriculum Vitae

1. Personal data

Name	Collins Burton Mwakwata
Date and place of birth	21 February 1989, Mbeya, Tanzania
Nationality	Tanzanian

2. Contact information

Address	Tallinn University of Technology, School of Information Technologies, Thomas Johann Seebeck Department of Electronics, Ehitajate tee 5, 19086 Tallinn, Estonia
Phone	+372 51930135
E-mail	collins.burton@taltech.ee

3. Education

2018–2022	Tallinn University of Technology, School of Information Technologies, Electronics and Telecommunication, PhD studies
2015–2017	Université de Blida 1, Faculty of Electronics, Telecommunication and Networks, MSc
2011–2015	Université de Blida 1, Faculty of Electronics, Telecommunication and Networks, BSc

4. Language competence

Swahili	native
English	fluent
French	fluent
Estonian	Intermediate

5. Professional employment

2017– 2018	Ericsson, RF Engineer
2014–2018	AIESEC, LCVP TM, LCP, MCVP TM

6. Computer skills

- Operating systems: Windows, Macintosh
- Document preparation: MS Office
- Programming languages: C#, Python, MATLAB

7. Defended theses

- 2017, MSc, University of Blida 01, Institute of Telecommunications and Networks
- 2015, BSc., University Of Saad Dahlab de Blida, Institute of Electronics and Telecommunications

8. Field of research

- Information and communication technology (ICT) - Telecommunication

9. Scientific work
Papers

1. CB Mwakwata, H Malik, MM Alam, Y Le Moullec, S Parand, S Mumtaz; Narrowband Internet of Things (NB-IoT): From physical (PHY) and media access control (MAC) layers perspectives; Sensors 19 (11), 2613.
2. CB Mwakwata, MM Alam, Y Le Moullec, H Malik, S Pärand; Cooperative Interference Avoidance Scheduler for Radio Resource Management in NB-IoT Systems; European Conference on Networks and Communications.
3. CB Mwakwata, O Elgarhy, Y Le Moullec, MM Alam, S Pärand, I Annus; Inter-cell Interference Reduction Scheme for Uplink Transmission in NB-IoT Systems; 2021 International Wireless Communications and Mobile Computing (IWCMC), 400-405.
4. CB Mwakwata, O Elgarhy, MM Alam, Y Le Moullec, S Pärand, K Trichias, K Ramantas; Cooperative Scheduler to Enhance Massive Connectivity in 5G and Beyond by Minimizing Interference in OMA and NOMA; IEEE Systems Journal
5. CB Mwakwata, MM Alam, Y Le Moullec; mMTC Users Classification Empowering Predictive Cooperative Scheduler in RAN Slicing for 5G and Beyond Networks; To be submitted

Elulookirjeldus

1. Isikuandmed

Nimi	Collins Burton Mwakwata
Sünniaeg ja -koht	21.02.1989, Mbeya, Tansaania
Kodakondsus	Tansaania

2. Kontaktandmed

Aadress	Tallinna Tehnikaülikool, Thomas Johann Seebecki elektroonikainstituut, Ehitajate tee 5, 19086 Tallinn, Estonia
Telefon	+372 519 30135
E-post	collins.burton@taltech.ee

3. Haridus

2018–2022	Tallinna Tehnikaülikool - Info- ja kommunikatsioonitehnoloogia, Elektroonika ja telekommunikatsioon, doktoriõpe
2015–2017	Blida 01 ülikool, Võrgustikud ja kommunikatsioonitehnoloogia teaduskond, magistriõpe *cum laude*
2014-2016	Tallinna Tehnikaülikool - Info- kommunikatsioonitehnoloogia, magistriõpe
2011–2015	Saad Dahlab de Blidaülikool, Elektroonika ja kommunikatsioonitehnoloogia teaduskond, BSc
2008-2013	Tallinna Tehnikaülikool - bioonika bakalaureuseõpe

4. Keelteoskus

Swahili keel	emakeel
Inglise keel	kõrgtase
Prantuse keel	kõrgtase
Eesti keel	kesktase

5. Teenistuskäik

2017– 2018	Ericsson Eesti AS, RF insener
2014–2018	AIESEC, LCVP TM, LCP, MCVP TM

6. Arvutioskused

- Operatsioonisüsteemid: Windows, Macintosh
- Kontoritarkvara: MS Office
- Programmeerimiskeeled: C#, Python, MATLAB

7. Kaitstud lõputööd

- 2022, doktorantuur, juhendaja Prof.Muhammad Mahtab Alam, Yannick Le Moulec, Tallinna Tehnikaülikool, ... Info- ja kommunikatsioonitehnoloogia Instituut.
- 2017, magistriõpe, Blida 1 ülikool, kommunikatsioonitehnoloogia Instituut
- 2015, bakalaureuseõpe, Saad Dahlab de Blida ülikool, kommunikatsioonitehnoloogia Instituut

8. Teadustöö põhisuunad

- Info- ja kommunikatsioonitehnoloogia - Telekommunikatsioonitehnoloogia

9. Artiklid

1. CB Mwakwata, H Malik, MM Alam, Y Le Moullec, S Parand, S Mumtaz; Narrowband Internet of Things (NB-IoT): From physical (PHY) and media access control (MAC) layers perspectives; Sensors 19 (11), 2613.
2. CB Mwakwata, MM Alam, Y Le Moullec, H Malik, S Pärand; Cooperative Interference Avoidance Scheduler for Radio Resource Management in NB-IoT Systems; European Conference on Networks and Communications.
3. CB Mwakwata, O Elgarhy, Y Le Moullec, MM Alam, S Pärand, I Annus; Inter-cell Interference Reduction Scheme for Uplink Transmission in NB-IoT Systems; 2021 International Wireless Communications and Mobile Computing (IWCMC), 400-405.
4. CB Mwakwata, O Elgarhy, MM Alam, Y Le Moullec, S Pärand, K Trichias, K Ramantas; Cooperative Scheduler to Enhance Massive Connectivity in 5G and Beyond by Minimizing Interference in OMA and NOMA; IEEE Systems Journal
5. CB Mwakwata, MM Alam, Y Le Moullec; mMTC Users Classification Empowering Predictive Cooperative Scheduler in RAN Slicing for 5G and Beyond Networks;

ISSN 2585-6901 (PDF)
ISBN 978-9949-83-846-2 (PDF)