Enhancing Low-Cost Cellular IoT Networks on the Sidelink and Satellite Links

by

Vishnu Rajendran Chandrika

M.Tech., Indian Institute of Technology Guwahati, 2014

B.Tech., University of Kerala, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Enhancing Low-Cost Cellular IoT Networks on the Sidelink and Satellite Links

submitted by **Vishnu Rajendran Chandrika** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Computer Engineering.**

Examining Committee:

Lutz Lampe, Professor, Electrical and Computer Engineering, UBC

Supervisor

Sathish Gopalakrishnan, Associate Professor, Electrical and Computer Engineering, UBC

Supervisory Committee Member

Bruce Shepherd, Professor, Computer Science, UBC

University Examiner

Panos Nasiopoulos, Professor, Electrical and Computer Engineering, UBC

University Examiner

Xiaodai Dong, Professor, Electrical and Computer Engineering, University of Victoria

External Examiner

Additional Supervisory Committee Member:

Victor Leung, Professor Emeritus, Electrical and Computer Engineering, UBC

Abstract

Internet-of-Things (IoT) is one of the fast growing technologies in the current era which holds a large and rapidly increasing global market size. Device-to-device (D2D) communication is a key enabler for connecting devices together to form the IoT, especially when the cellular coverage is limited. Similarly, non-terrestrial networks (NTNs) involving satellites which complement the terrestrial cellular networks to provide global coverage also play a vital role in expanding IoT. Both D2D and satellite links are essential in providing seamless universal cellular IoT (C-IoT) coverage. In this thesis, we propose enhancements for C-IoT devices which address up-to-date problems in D2D communication and NTNs.

Leveraging the unlicensed frequency bands for D2D communication reduces the costs, and offloads network traffic from the licensed spectral resources. To this end, we design a new low-cost radio access technology (RAT) protocol called Sidelink Communications on Unlicensed BAnds (SCUBA). SCUBA complements the primary RAT, and functions by reusing the existing hardware on a non-overlapping time-sharing basis. We prove the effectiveness of our protocol with analyses and simulation results of the medium access control layer of SCUBA.

One of the most critical problems faced by NTN is the uplink (UL) synchronization failure due to high Doppler offset. While NTN new radio (NR) devices rely on global navigation satellite system (GNSS) to resolve this issue, it is not always feasible for power-critical IoT user equipments (UEs). Therefore, we design Synchronization signal-based Positioning in IoT Non-terrestrial networks (SPIN) which enables the IoT UEs to tackle the UL synchro-

nization problem. Our evaluations show that SPIN positioning accuracy achieves the Cramér-Rao lower bound and meets the target accuracy required for UL synchronization, along with significant battery life savings over GNSS-based solution.

Another pertinent problem faced by C-IoT devices in NTN is the extended round-trip time resulting in a degraded network throughput. To this end, we propose smarter hybrid automatic repeat request (HARQ) scheduling methods that can increase the efficiency of resource utilization. We conduct end-to-end link-level simulations of C-IoT traffic over NTNs. Our numerical results of data rate show the improvement in performance achieved using our proposed solutions against legacy scheduling methods.

Lay Summary

Internet-of-Things (IoT) refers to the interconnected network of physical objects which collect and share data. IoT transforms unintelligent objects into intelligent devices thereby realizing a smarter world. In this thesis, we identify problems related to cellular IoT (C-IoT) devices in terrestrial and non-terrestrial networks (NTN) and propose different solutions which enable them to have seamless global coverage. There exists no device-to-device (D2D) communication technology in C-IoT which can reuse the existing radio on the devices in a time-shared manner with the primary radio technology. Further, the recently standardized C-IoT NTN experiences initial access failure due to high Doppler offset and throughput degradation due to large propagation delay. To this end, we design a time-multiplexed D2D communication protocol, a self-positioning solution, and smarter resource scheduling techniques for the C-IoT devices. Our main design criteria include low cost and low power consumption which make our solutions ideally fit into an IoT device.

Preface

This thesis is based on original research that I conducted under the supervision of Professor Lutz Lampe in the Department of Electrical and Computer Engineering (ECE) at the University of British Columbia, Vancouver, Canada. The co-authors in my publications, Mr. Gustav Vos and Mr. Serkan Dost, from Sierra Wireless Inc., have assisted me in formulating the research problems and provided necessary feedback related to industry requirements. The invention disclosure documents and the patent applications are the outcomes of collaborative effort of all the inventors listed under the patents mentioned below. The other co-authors in my works, Dr. Gautham Prasad and Dr. Jiayin Chen, from the University of British Columbia, guided me in the research by providing valuable suggestions and reviewed the manuscripts. For the publication related to Chapter 5, Dr. Gautham Prasad and I are the first authors, and we contributed equally to the manuscript. For all the other manuscripts, I am the first author, and was responsible for reviewing the literature, developing the solutions, evaluating through simulations, and writing the publication manuscripts. Professor Lutz Lampe supervised all my work.

The publications related to the work in this thesis are listed below.

Publications Related to Chapter 2

Conference Paper, Published

- **Vishnu Rajendran**, Gautham Prasad, Lutz Lampe, and Gus Vos, "On the Feasibility of In-Device Multiplexed Unlicensed D2D Communications," *IEEE Vehicular Technology Conference - Fall (VTC-Fall)*, Nov. 2020.

Journal Paper, Published

- **Vishnu Rajendran**, Gautham Prasad, Lutz Lampe, and Gus Vos, "SCUBA: An In-Device Multiplexed Protocol for Sidelink Communication on Unlicensed Bands," *IEEE Internet of Things Journal*, vol. 8, no. 22, pp. 16637-16652, Nov. 2021.

Patent, Filed

- Gus Vos, **Vishnu Rajendran Chandrika**, Gautham Prasad, Lutz Lampe, "Method and Apparatus for Device to Device Communication for Cellular Devices," *International Application No. PCT/CA2020/051632*, filed on November 27, 2020

Publications Related to Chapter 3

Conference Paper, Published

- **Vishnu Rajendran Chandrika**, Gautham Prasad, Lutz Lampe, and Gus Vos, "Synchronized SCUBA: D2D Communication for Out-of-Sync Devices," *IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1-7, June 2020.

Patent, Filed

- Gus Vos, **Vishnu Rajendran Chandrika**, Gautham Prasad, Lutz Lampe, "Method and Apparatus for Supporting Device to Device Communication for Wireless Devices," *International Application No. PCT/CA2021/051759*, filed on December 8, 2021

Publications Related to Chapter 4

Journal Paper, Submitted

- **Vishnu Rajendran Chandrika**, Jiayin Chen, Lutz Lampe, Gus Vos, Serkan Dost, "SPIN: Synchronization Signal Based Positioning Algorithm for IoT Non-Terrestrial Networks," submitted in 2022.

Patent, Filed

- Gus Vos, Serkan Dost, **Vishnu Rajendran Chandrika**, Jiayin Chen, Lutz Lampe, "Method and Apparatus for Determining Frequency Offset Between a Non-Terrestrial Network Base Station and a User Equipment," *U.S. Patent Application No. 63/422,651*, filed on November 4, 2022.

Publications Related to Chapter 5

Journal Paper, Major Revision Submitted

- Gautham Prasad, **Vishnu Rajendran Chandrika**, Lutz Lampe, Gus Vos, "Enhanced Hybrid Automatic Repeat Request Scheduling for Non-Terrestrial IoT Networks," submitted in 2022.

Patent, Filed

- Gus Vos, Gautham Prasad, **Vishnu Rajendran Chandrika**, Lutz Lampe, Serkan Dost, "Method and apparatus for HARQ scheduling for non-terrestrial networks," *U.S. Patent Application No. 63/136,969*, filed on January 13, 2021.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2-WLS	Two-step Weighted Least Squares
3GPP	3rd Generation Partnership Project
ACK	Acknowledgment message
AoA	Angle of Arrival
API	Application Program Interface
AWGN	Additive White Gaussian Noise
BLE	Bluetooth Low Energy
BLER	Block Error Rate
BS	Base Station
BW	Bandwidth
Cat-M1	Category-M1
CCA	Clear Channel Assessment
CCH	Control CHannel
CDRX	Connected mode with discontinuous reception
CD-SSB	Cell-Defining Synchronization Signal Block
CFR	Code of Federal Regulations
C-IoT	Cellular Internet-of-Things
CN	Core Network
ConA	Connected Active mode
COOS	Coverage-out-of-sync

CP	Cyclic Prefix
CQI	Channel Quality Indication
CRLB	Cramér-Rao Lower Bound
D2D	Device-to-Device communication
D2D-U	D2D in Unlicensed bands
Data-INAT	Data INActivity Timer
DCI	Downlink Control Information
DD2A	Downlink Data to Acknowledgement
DG2D	Downlink Grant to Data
DL	Downlink
DRX	Discontinuous reception
DRX-INAT	DRX INActivity Timer
DST	Destination UE
DTX	Discontinuous transmission
ECEF	Earth-Centered Earth-Fixed
E-CID	Enhanced Cell ID
eDRX	Extended DRX
EIRP	Effective Isotropic Radiated Power
eMBB	Enhanced Mobile Broadband
ERP	Effective Radiated Power
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FDOA	Frequency Difference of Arrival
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix

FOA	Frequency of Arrival
GEO	Geostationary Earth Orbit
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HAPS	High-Altitude Platform Stations
HARQ	Hybrid Automatic Repeat Request
HC	Homogeneous Coverage
HD-FDD	Half-Duplex and Frequency Division Duplex
IAT	Inter-Arrival Time
iBLER	Initial Block Error Rate
IDRX	Idle mode with discontinuous reception
IMSI	International Mobile Subscriber Identity
IoMT	Internet-of-Mobile-Things
IoT	Internet-of-Things
ISL	Inter-Satellite Links
ISM	Industrial Scientific and Medical
LBS	Location-Based Services
LEO	Low Earth Orbit
LOS	Line-of-Sight
LS	Location Server
LTE	Long Term Evolution
LTE-M	LTE-MTC
LUT	Look-Up Table
MAC	Medium Access Control
MCL	Maximum Coupling Loss

MCS	Modulation and Coding Scheme
MEO	Medium Earth Orbit
MIB	Master Information Block
MMSE	Minimum Mean Squared Error
mMTC	Massive Machine Type Communication
MOPS	Million Operations per Second
M-PDCCH	MTC Physical Data Control CHannel
MTBG	Multiple Transport Block Grant
MTC	Machine Type Communication
NACK	Negative acknowledgment message
NB-IoT	Narrowband IoT
NLOS	Non-Line-of-Sight
N-PDCCH	NB-IoT Physical Downlink Control CHannel
NPSS	Narrowand Primary Synchronization Signals
NR	New Radio
NR SL	New Radio Sidelink
NSSS	Narrowband Secondary Synchronization Signals
NTN	Non-Terrestrial Network
O2I	Outdoor-to-Indoor
OFDM	Orthogonal Frequency Division Multiplexing
OOC	Out-of-Coverage
OTA	Over-the-Air
OTDOA	Observed Time Difference of Arrival
PC	Partial Coverage
PDCCH	Physical Downlink Control CHannel

PDSCH	Physical Downlink Shared CHannel
PF	Paging Frame
PHY	Physical Layer
PO	Paging Occasion
P-RAT	Primary Radio Access Technology
PRB	Physical Resource Block
PSD	Power Spectral Density
PSS	Primary Synchronization Signals
PUCCH	Physical Uplink Control CHannel
PUSCH	Physical Uplink Shared CHannel
PWC	Power Class
QPSK	QuadriPhase Shift Keying
RA	Random Access
RAI	Release Assistance Indication
RAN	Radio Access Network
RAT	Radio Access Technology
RedCap	Reduced Capability
RLC	Radio Link Control
RMS	Root Mean Square
RP	Reference Point
RRC	Radio Resource Control
RS	Reference Signal
RSRP	Reference Signal Received Power
RTT	Round-Trip Time
Rx	Receiver

RX	Receive
SAM	Sidelink Availability Message
SAM-D	Sidelink Availability Message for Dynamic SL-PO
SAM-U	Unavailability SAM
SCH	Shared CHannel
SCUBA	Sidelink Communications on Unlicensed BAnds
SCUBA-LLM	SCUBA Low Latency Mode
SF	Sub-frame
SIB	System Information Block
SL	Sidelink
SL-DRX	Sidelink-Discontinuous reception
SL-INAT	Sidelink INActivity Timer
SL-PO	Sidelink Paging Occasion
SLSS	Sidelink Synchronization Signal
SL-U	SL in Unlicensed frequency bands
SNR	Signal-to-noise ratio
SOSN	Self Organized SCUBA Network
SPIN	Synchronization signal-based Positioning in IoT Non-terrestrial networks
S-RAT	Secondary Radio Access Technology
SRC	Source UE
SRS	Sounding Reference Signal
SS	Synchronization Signal
SSB	Synchronization Signal Block
SSS	Secondary Synchronization Signal
SSW	SCUBA Sync Window

STBG	Single Transport Block Grant
SUF	Sub-frame Utilization Factor
SW	Switching SF
TA	Timing Advance
TB	Transport Block
TBS	Transport Block Size
TDL	Tapped Delay Line
TDM	Time Division Multiplex
TDOA	Time Difference of Arrival
TimeREQ	Timing REQuest message
TimeRSP	Timing ReSPonse
TOA	Time of Arrival
TWLS	Taylor series-based Weighted Least Squares
Tx	Transmitter
TX	Transmit
UAS	Unmanned Aircraft Systems
UAV	Unmanned Aerial Vehicles
UE	User Equipment
UG2D	Uplink Grant to Data
UG	Uplink Grant
UL	Uplink
URLLC	Ultra-Reliable Low-Latency Communications
UTDOA	Uplink Time Difference of Arrival
V2X	Vehicle-to-everything
ZC	Zadoff-Chu

Acknowledgments

I would like to express my deepest gratitude to my PhD supervisor Prof. Lutz Lampe for giving me the opportunity to work with him and for the continuous guidance throughout my research. I could not have undertaken this journey without his timely help, constructive feedback, and our regular discussions. I admire the way he manages time to regularly meet with each student in his research group and conducts research on diverse topics. I couldn't have asked for a better adviser.

I am also thankful to the rest of my supervisory committee: Prof. Victor Leung and Dr. Sathish Gopalakrishnan for their time and effort in evaluating my work and providing valuable feedback and comments. Special thanks to Dr. Sathish Gopalakrishnan for the constant motivations and friendly chats, be it during the instruction of the course or whenever we met outside. I am immensely thankful to Prof. Xiaodai Dong of University of Victoria for taking the time out to serve as the external examiner for my final exam and for providing valuable input to improve the contents of my thesis. I extend my gratitude to Prof. Bruce Shepherd and Prof. Panos Nasiopoulos for serving as university examiners and for providing feedback and comments. I would also like to extend my thanks to Dr. David Michelson who served as examiner for my PhD proposal and departmental exams and gave insightful comments.

I am sincerely grateful to Mr. Gustav Vos and Mr. Serkan Dost from Sierra Wireless Inc., Richmond, BC, Canada, for helping me throughout the PhD research by giving frequent feedback and comments. Their inputs always helped me to understand the industry trends

and rightfully set the target requirements for the proposed solutions. I am thankful to them for also finding time to have weekly technical discussions with us and for their extended support in filing patents for our research works. I am grateful to MITACS, Canada, Natural Science and Engineering Research Council of Canada (NSERC), and Sierra Wireless Inc. for supporting my research projects.

I would also like to express my deepest appreciation to Dr. Gautham Prasad who always guided me both personally and at work, right from my first day in Canada. Many thanks to him for providing me accommodation, guiding me during my initial days in Vancouver, and for hosting parties. His constant guidance, help, and research collaboration definitely played a big role in the fulfillment of my PhD research. I would like to extend my thanks to Dr. Jiayin Chen for finding time to have weekly discussions and providing me insightful feedback. I also thank her for reviewing my work and manuscript, and for giving valuable comments.

I am deeply indebted to my partner and my parents for their unconditional love, unwavering support, and continuous encouragement. This endeavor would not have been possible if it was not for my mom, Mrs. Chandrika Rajendran, who is my inspiration for deciding to go for higher studies. I also thank my Masters supervisors, Prof. P.K. Bora and Dr. A Rajesh, who encouraged me for the first time to do PhD research. I would also like to acknowledge the constant support and love given by my dearest friend Yaash, who always made me feel Vancouver as my home. Special thanks to my friend Dhananjay for frequently hanging out with me and being there to share my joys and sorrows.

Finally, I would like to thank Praveen, Rohith, Maria, Amal, Elena, Mrinmoy, Razi, Faramarz, Taha, Mostafa, and Tharrane for being very friendly, helpful, and supportive, and for hosting parties. It would be remiss in not mentioning my labmates at Kaiser 4090, all my friends, and my close relatives, they all helped me in one way or the other.

Dedication

To my parents, my partner, and my sister

Chapter 1

Introduction

The Internet-of-Things (IoT) enables almost every device to connect to the internet, and thereby creates a smart world where seamless connection exists between different systems and devices. The paradigm of IoT provides connection to devices deployed in various applications demanding flexible and wide range of data rates, latency, coverage, and power constraints. The key enablers of IoT include Bluetooth, Wi-Fi, Zigbee, low-power wide-area (LPWA) technologies such as LoRa, Sigfox, and recent generations of cellular communication standards such as long term evolution (LTE) and 5G. Currently, when the world is witnessing a drastic increase in the number of IoT devices, a plethora of use cases among them demand low-cost and low-power solutions. The 3rd generation partnership project (3GPP) has introduced many variants of cellular communication standards as part of LTE and 5G which include LTE-machine-type communication (LTE-M) and narrowband IoT (NB-IoT) primarily aimed at low-power and low-cost devices. These cellular communication technologies are distinct from other IoT counterparts due to their long range connectivity, ability to support flexible performance requirements, standard deployment, and fast pace of evolution. Nevertheless, the cellular IoT (C-IoT) network is not seamless or ubiquitous. First, there might be coverage gaps or holes in between the cells served by the cellular base station (BS). However, the user equipments (UEs) located in these regions might have close proximity to other UEs which are in coverage of the cellular network. Device-to-device (D2D) communication is relevant in this scenario as it enables direct communication between the nearby UEs. D2D links help in eliminating the coverage gaps [1] either by allowing direct commu-

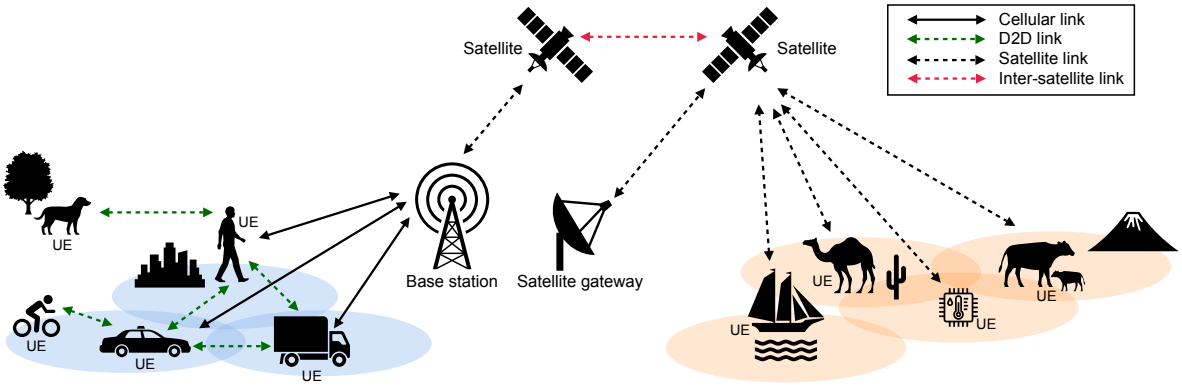


Figure 1.1: An illustration of a cellular IoT network complemented by D2D and satellite links.

nication between the end devices or by enabling the UE to communicate with the network by hopping through another UE in coverage. In addition, D2D offloads the traffic from the cellular link which allows the BS to serve more UEs, which is particularly beneficial for the C-IoT networks having large UE density. Second, there are large geographical regions on the earth such as deserts, oceans, forests, and remote villages which are currently unserved by the cellular network. It is not economical and/or practical to invest on, deploy, operate, and maintain cellular network infrastructure in these regions. However, this problem can be resolved if satellite networks complement the cellular network [1]. Toward this end, in Release 17, 3GPP standardized the first phase of enabling non-terrestrial network (NTN) access for the cellular users including LTE-M and NB-IoT devices, with the objective to eventually provide quasi-ubiquitous cellular connectivity.

As discussed above, D2D and satellite communication are two critical components of C-IoT which help in providing seamless global coverage. In Figure. 1.1, we show an illustration of a conventional cellular IoT network complemented by D2D and satellite communication links. While the UEs in close proximity communicate to each other via D2D links regardless of their cellular coverage, the UEs located in remote unserved regions communicate to the network through the satellite links. The existing D2D solutions have several shortcomings

(discussed later in Section 1.2.1) and hence they are not ideal to be implemented in C-IoT devices. On the other hand, NTN integration with C-IoT is still in its standardization phase and hence there are several problems remaining to be resolved. To this end, in this research, we focus on further enhancing the performance of low-cost C-IoT devices by proposing various solutions. First, we design a D2D communication protocol which enables the low-cost C-IoT devices to directly talk to each other on unlicensed bands. Then, we design a self-positioning solution for C-IoT devices to enable them compensate for the delay and Doppler impacts in NTN, thereby allowing initial access to the NTN and further location-based services (LBS). Finally, we propose smarter hybrid automatic repeat request (HARQ) scheduling techniques for C-IoT UEs to tackle the throughput degradation due to large propagation delay in NTN.

1.1 Background and Motivation

In this section, we discuss the relevant background information and the motivations for the cellular IoT research problems addressed in this thesis.

1.1.1 Sidelink Communication

As briefly discussed above, D2D communication is a critical constituent of the IoT. IoT and machine-type communication (MTC) devices interacting directly with each other on the sidelink (SL), in lieu of using uplink (UL) and downlink (DL) paths to talk via a central BS, require no continuous supervision from a central access point or a BS [2–4]. By allowing network nodes to communicate directly with each other, D2D offers potential reduction in the total power consumed and the network resources used, improvement in the achieved latency, and possible coverage enhancement [5–10]. In 3GPP cellular standards, SL[1] refers to the

[1]We adopt the term SL from 3GPP standards, and use the terms SL communications and D2D communications interchangeably throughout the thesis.

direct D2D communication interface which enables proximity services and relaying in LTE-Advanced [11–14], and advanced vehicle-to-everything (V2X) communication in 5G new radio (NR) [14, 15]. SL communication is a key component to realize 5G-mobile edge computing (MEC) based networks primarily aimed at V2X applications [16–20]. D2D solutions are also suitable in a wide range of MTC and IoT applications such as sensors, wearables, public lighting, container tracking, and smart metering [21, Chs. 2, 4]. Additionally, using SL on unlicensed bands (SL-U) is a cost-effective solution to obtain all the benefits of D2D communications with the potential of reducing the communication latency, while also alleviating traffic on the already congested cellular bands [22–24].

1.1.2 IoT NTN

NTN refers to a network which involves satellites in low earth orbit (LEO), medium earth orbit (MEO), or geostationary earth orbit (GEO) or unmanned aircraft systems (UAS) enabling radio frequency communication between users [25]. In Release 17, 3GPP standardized the first phase of integration of conventional terrestrial cellular networks with NTN with an objective to achieve global cellular coverage. The 3GPP NTN standardization plan for Release 17 and beyond includes NTN support for NR and IoT which covers both enhanced mobile broadband (eMBB) and massive machine type communications (mMTC) use cases. Cellular-NTN integration helps in extending the cellular communication service to currently unserved and underserved NR and IoT users which include those located in remote villages, aircrafts, ships, and cell-edge regions [26–30]. While NTN use cases in eMBB mainly refer to high data rate applications such as broadband connectivity and media and entertainment, NTN-IoT use cases consider services such as wide and local area connectivities [26, 31, 32]. The wide area NTN IoT services include global connectivity between sensors and actuators scattered over a large geographical area, which are used in a variety of applications such as

automotive and road transport, energy, and livestock management. On the other hand, local area IoT services provided by NTN encompass connectivity to sensors and actuators located in a smart grid system or a moving platform, e.g., a container on-board a vessel, a truck, or a train [26].

In a cellular-NTN integrated network, the BS is present either on-board the satellite or on the earth connected to the satellite through a gateway. The main challenges faced by NTN communication are long propagation delay due to the large distance between satellite and UE, significant difference in timing advance (TA) between UEs at different locations within the cell due to its large size, and huge Doppler shifts due to the movement of satellites. The large propagation delay results in stop-and-wait gaps in the bidirectional communication between the UE and the BS. On the other hand, the disparity of TAs between UEs situated at different locations in the cell results in a large overlap of UL signals at the BS. Furthermore, the Doppler shift due to satellite velocity results in a high frequency offset in the UL signal at the BS. While the long propagation delay affects the data rate of communication, TA disparity and Doppler shift directly impact the UL synchronization, thus making the network inaccessible to the UE.

1.1.3 Self-Positioning in IoT NTN

To solve the synchronization problem, 3GPP considers common TA and Doppler compensations with respect to a reference point (RP) in the cell. In addition, the working assumptions include global navigation satellite system (GNSS) capability in the UE. Thus, with the knowledge of its relative location with respect to the RP, the UE can compute and compensate for the residual TA and Doppler. The GNSS based solution is feasible for NR devices, whereas it is not always viable for IoT devices for several reasons including battery and cost constraints. While several state-of-the-art solutions are available to counter the NTN UL synchronization

problem [33–36], a solution specifically tailored for IoT UEs which considers both the power constraints and the required accuracy is still unavailable. To this end, we design a positioning solution which utilizes the existing synchronization signals (SSs) in IoT NTN, and operates with minimal power consumption. In the following, we explain the motivation of our work by having a detailed discussion of the UL synchronization issues and GNSS based solution and its drawbacks.

UL Synchronization Problem

To partially solve the UL synchronization problem, 3GPP has agreed to enable the BS to compensate for the common TA and Doppler offset with respect to an RP in the cell [25]. While the BS pre-compensates the common TA and Doppler offset in DL, it post-compensates the common offsets on the received UL signal. During DL synchronization process, the UE estimates the sum of residual Doppler and crystal oscillator offset. However, unless the individual offset components are estimated, the UE cannot accurately pre-compensate for the residual Doppler. Likewise, for timing synchronization, the UE is unable to estimate accurate residual TA. Consequently, the UL signal suffers from large residual TA and Doppler shift depending on the location of the UE within the cell. For instance, if the RP is the cell center, the UEs located at the cell edge experience large residual Doppler and TA. Subsequently after DL synchronization, the UE tries to establish a connection with the network by initiating random access (RA) procedure, which involves the transmission of an RA preamble. However, the frequency offset due to the residual Doppler and the timing offset due to the residual TA will cause the preambles to remain undetected at the BS, resulting in connection establishment failure. It should be noted that the residual TA and Doppler offset are significantly high for LEO satellites while they are negligible for MEO or GEO satellites [37]. Moreover, UAS also usually experiences lower delay and Doppler which makes it similar or equivalent to a

terrestrial network [25]. Therefore, we consider only LEO satellites in our design, analyses, and evaluation.

Need for UE Location

Since the network compensates common TA and Doppler offset with respect to an RP, the UE requires the knowledge of its own location, satellite location and velocity, and the coordinates of the RP to compute the residual offsets. To this end, 3GPP has decided to enable the network to broadcast the location coordinates of the RP and the ephemeris which consists of the position and velocity vectors of the satellite. In addition, 3GPP considers only those UEs with GNSS capability such that the UEs know their location and hence can compute and pre-compensate residual Doppler and TA. However, when battery-powered low-cost IoT UEs are considered, the assumption of GNSS capability poses several challenges which are discussed in the following.

GNSS Optimization Problem

3GPP has recently agreed to allow UEs to have time gaps for the purpose of performing GNSS position fix. The working assumptions for IoT NTN include UE incapability to perform GNSS and cellular operations simultaneously. Therefore, before the positioning time gaps, an IoT UE must terminate radio resource control (RRC) connected mode and transition to RRC idle mode to perform GNSS positioning. After obtaining the position fix, the UE must re-establish connection which involves synchronization and RA procedures, thus resulting in huge battery drain. Consequently, for a battery-powered mobile IoT UE which occasionally changes its location, and has long connected modes, it is not feasible to rely on the GNSS assisted UL synchronization. In this regard, the 3GPP work item description (WID) document for NTN IoT Release 18 includes an item which is aimed at optimizing the

GNSS operations involved in the UL synchronization procedure [38].

GNSS Unavailability Problem

Though 3GPP is currently focusing on IoT UEs with GNSS capability, there are many scenarios where GNSS is unavailable. First, GNSS chipset is not necessarily integrated into all cellular IoT devices for cost and battery impact reasons. Second, GNSS suffers from poor link budget due to long distance [39]. Consequently, GNSS fails even in soft indoor cases such as logistic applications involving UEs placed inside containers or carriages [40]. On the other hand, LEO satellites with antenna gains same as or more than that of GNSS, have better link budget [39][2]. Furthermore, GNSS signals are also highly susceptible to jamming and spoofing [34,40]. All the above problems result in GNSS unavailability which eventually leads to NTN communication failure in the UE.

Location-Based Services (LBS)

There is a wide range of services which essentially require the UE's location, called LBS. Among them are emergency services, e.g., E911, and infotainment services such as map services, direction to a location, and local advertising [41] [42, Ch. 1]. In terrestrial networks, UEs rely on GNSS, cellular positioning, or a hybrid of both to obtain the location. However, in an IoT NTN, these methods either do not work due to the lack of cellular coverage and the unavailability of GNSS or are not feasible due to UE power constraints. Alternatively, the focus on utilizing LEO communication satellites for LBS has been growing recently in light of the massive LEO constellations such as Starlink and Amazon-Kuiper [43].

To solve all the above discussed problems, we need a positioning solution as an alternative to GNSS which should be designed to specifically cater to the NTN-IoT requirements.

[2]This is not always true, since sometimes the LEO satellites may have lower antenna gains which results in a link budget similar to that of GNSS [25].

As briefly discussed in Section 1.1.2, another major problem faced by NTN is the degradation of network throughput due to the extended round-trip time (RTT). We find that the current HARQ specification is not optimal to deal with the huge propagation delay in NTN. In the next section, we discuss the effect of extended RTT on IoT NTN throughput and the need for enhanced HARQ techniques.

1.1.4 HARQ in IoT NTN

We investigate the impact of the extended RTT of bidirectional signals caused due to the large propagation distance in NTNs on the network throughput. Typically, the achievable data rate may not be the primary target performance indicator to focus on while designing C-IoT and mMTC systems. One of the reasons for this is the greater resource consumption (e.g., extended bandwidth requirement) and higher transmit power associated with achieving increased throughput. However, we find that the network throughput can be improved without demanding additional resources or increasing transmit power, but instead by efficiently scheduling the HARQ processes. As a result, the IoT-NTN can support a larger number of mMTC devices, which is critical in NTNs due to the significantly larger cell size compared to conventional terrestrial networks. Toward this end, we exploit the lengthy RTT in NTN links to overlap bidirectional signals in the air. We further utilize the frequency-division full-duplex nature of BSs to allow simultaneous transmission and reception at the satellites and BSs.

1.2 Related Work

In the previous section, we discussed our research problems and also provided necessary background information. In this section, we provide a summary of the related works in each

of the above research problems.

1.2.1 Sidelink Communication

While several commercial D2D communication radio access technologies (RATs), such as Bluetooth [44], ZigBee [45], and Wi-Fi-Direct [46], which operate on unlicensed frequency bands, already exist, and various other solutions to this end have been proposed in the past [47–55], including the LTE D2D and 5G NR sidelink (NR SL) by 3GPP [11–15], we identify the following major shortcomings in these prior-arts.

First, many of the existing D2D solutions demand a dedicated radio in addition to the already existing cellular radio in the UE. This results in increased hardware overhead and additional cost, which are undesirable, especially for low-cost and low-power MTC and IoT devices [56, 57]. The SL-U method proposed in [53] based on an optimal scheduling and resource allocation algorithm, the D2D clustering based technique in [54] that achieves SL communication using an SL device with the best channel interacting with the BS on behalf of every other user in its cluster, and the millimeter and microwave based D2D communication solution in [52], all require Wi-Fi interfaces to enable SL operation.

Second, several of the prior arts require continued aid from the network BS for operation assistance and control, e.g., resource scheduling, which undermines the gains achieved by transitioning SL to unlicensed bands. The millimeter wave unlicensed band D2D communications solutions in [52, 55], the SL-U on television white space solution in [47], the D2D clustering based technique in [54], and several other prior arts in [48–51], all rely on a central BS for scheduling, data collection and processing, and/or resource grant allocation.

Furthermore, without further enhancements, commercially available technologies such as Bluetooth and Wi-Fi-Direct require manual device pairing and repeated user interventions to re-establish connections when disruptions occur [58–60].

1.2.2 UL Synchronization in IoT NTN

To address the NTN UL synchronization problem during GNSS unavailability, a positioning solution using time difference of arrival (TDOA) measurements, is proposed in [33]. This method uses NR primary synchronization signals (PSS) from the serving LEO communication satellite to get the position fix. However, this solution has several shortcomings. First, the method targets 5th generation (5G) NR UEs and hence cannot be directly adopted for NTN IoT standards such as LTE-M or NB-IoT. For instance, when an NB-IoT UE uses narrowband primary synchronization signals (NPSS) and/or narrowband secondary synchronization signals (NSSS) in place of NR PSS, the accuracy will be worse due to their smaller bandwidth. Second, the method directly uses the time of arrival (TOA) estimates obtained from the cellular synchronization step for positioning, which is limited by the sampling rate. Despite using a high sampling rate of 30.72 Msps in NR UEs, the achieved accuracy is found to be insufficient. Directly adopting this solution for NB-IoT and LTE-M UEs which use an acquisition sampling rate of 1.92 Msps makes it completely unfit for UL synchronization purpose. Moreover, the method considers only NR PSS for which the accuracy could be improved by including NR secondary synchronization signal (SSS). The next problem is that the method uses unweighted least squares for solving the first order Taylor series based linearized equations. In an NTN network consisting of LEO satellites which move at a high velocity relative to the UE, the signal-to-noise ratio (SNR) of the received signal varies significantly over time. Under the highly varying SNR, a weighted least squares (WLS) solution undoubtedly gives better accuracy. Furthermore, the solution does not utilize the frequency difference of arrival (FDOA) measurements which could significantly improve the position estimation [61] [62, Ch. 5]. Another pertinent problem with the method is that it uses beam center broadcast by the network as the initial location for the iterative positioning algorithm. However, we find that the beam center for set-3 and set-4 satellites, the newly introduced

satellite scenarios [63], can be very far from the true UE location. This may cause the positioning algorithm to wrongly converge to a local minimum. More importantly, this solution defines only position acquisition which is performed before initiating an RRC connection. It does not define any position tracking, which is essential for maintaining the accuracy when the UE is mobile and stays in connected mode for long time. Finally, we also find some shortcomings regarding the evaluation assumptions including the number of satellites, measurement noise model, and the UE location. The method considers only a single satellite for positioning which largely affects the achievable accuracy and the time taken to get a position fix. It takes approximately 4.2 minutes to achieve a positioning accuracy of around 700 m. The achieved accuracy is insufficient for the purpose of UL synchronization in NTN especially when the synchronization error attributed by other sources, e.g., crystal offset, is high. In addition, the large positioning time results in high latency and huge energy consumption which are not feasible for an IoT UE. Currently planned and recently deployed LEO satellite constellations, e.g., Starlink [64, 65] and Amazon-Kuiper [65], are large such that the UEs may get visibility to several LEO communication satellites simultaneously or sequentially in a specified duration of time. Therefore, the positioning solution and evaluation should consider a reference constellation which is a realistic scenario.

Another state-of-the-art solution for the UL synchronization problem is the network assisted closed-loop frequency and TA control [66]. This method can work even when the UE is in RRC connected mode. This is in contrast with the GNSS solution which cannot function simultaneously with cellular connected mode, especially in low-cost IoT devices. However, closed-loop control requires configuration of measurement signals, e.g., sounding reference signal (SRS), and frequency and TA offset signaling via media access control (MAC) layer control element (CE) and/or downlink control information (DCI). In addition, closed-loop control can work only after the UL transmission of msg1 [67], which demands the support

for enhanced physical random access channel (PRACH) formats and/or preambles [25]. The signaling overhead, the specification impact, and the requirement for enhanced PRACH make the closed-loop frequency and TA control an unfavorable choice.

Another solution recently introduced to tackle the UL synchronization problem during GNSS unavailability is to directly estimate the Doppler shift using the cellular reference signals transmitted at more than one frequency [34]. However, the method requires the network to transmit additional synchronization signal blocks (SSBs) at frequencies sufficiently far from that of the cell-defining SSB (CD-SSB). In addition, the network is also required to include indication of time and frequency locations of the additional SSB in the system information broadcast 1 (SIB1) associated with the CD-SSB. This signaling associated with the additional SSBs results in network overhead. Moreover, the SSB based solution is applicable only to NR UEs, and hence cannot be directly adopted to IoT UEs.

1.2.3 HARQ in IoT NTN

IoT-NTN studies have thus far, rightly, focused predominantly on initial access aspects, such as analysis of link budget [68, 69], challenges of Doppler effects and solutions to counter them [69–71], and issues related to random access [69, 72–74]. Enhancing scheduling for IoT-NTN networks has also been considered in the past, but from a system-level perspective, i.e., enhancements at the BS-end for scheduling transmissions among users [75, 76]. The solutions proposed are beneficial in countering the impact of increased Doppler shift in IoT-NTN systems, but do not address the link-level packet scheduling issue that we consider in this work to enhance data rates.

In the context of NTN, solutions that have previously been proposed to counter the impact of large propagation delay can be roughly classified into three categories:

A) targeting a low initial block error rate (iBLER),

B) disabling the use of HARQ feedback at the MAC layer, and

C) increasing the number of HARQ processes.

In the following, we consider solutions from each of these categories and discuss the reasons why they are not applicable to solve the problem at hand.

Low Target iBLER

The issue of stop-and-wait gaps produced due to the lengthy RTT is recognized in [77, 78] for GEO systems. These works acknowledge that packet errors and the resulting re-transmissions can significantly increase transmission gaps especially for high altitude GEO links and thereby reduce the data rate. While [77] only discusses the open issue of stop-and-wait gaps, [78] proposes a solution to counter this by simply improving robustness, i.e., targeting a low iBLER to avoid re-transmissions. Although using low iBLERs can help in reducing re-transmissions, previous simulation studies have shown that a high target iBLER instead is optimal to reduce the overall UE power consumption. This is because a larger target iBLER demands fewer repetitions and can accommodate a higher coding rate [79]. The study in [79] however only considers terrestrial networks operating with negligible propagation delay. When a large iBLER is used in NTNs, it does result in increased stop-and-wait gaps.

Disabled HARQ

Although link-level scheduling research in IoT-NTN is still in its infancy, the issue has been studied extensively for supporting eMBB applications in NR-NTN. An overview of these methods is reported in [25]. We evaluate if the methods suggested for NR-NTN can be adapted also for IoT-NTN. One solution to solve the issue of stop-and-wait gaps in HARQ

transmissions is to disable the use of HARQ feedback [25, Sec. 6], [80]. In this case, re-transmissions can be handled at the radio link control (RLC) layer. However, several problems exist with this technique. First, the benefits of HARQ combining at the physical (PHY) or MAC layer are absent, and thus, more re-transmissions may be needed, which then results in poor overall spectral efficiency. Furthermore, relying on RLC layer for re-transmission introduces additional user-plane latency and increased jitter when errors occur due to the high re-transmission timeout. This is because, without HARQ feedback, the network relies on the RLC feedback. In this case, the status report, which contains the acknowledgment message (ACK) or negative ACK (NACK) feedback, can be severely delayed. For example, when there is no UL data to transmit, the UE must send a scheduling request on PUCCH to request the scheduling of UL data just to send the status report [11]. This can take longer than three times the RTT, depending on PUCCH configuration. Furthermore, when there are burst errors, the status report can only be sent after a valid packet is received, which also causes increased latency. Additionally, if the last physical data unit is in error, the status report is not sent until the transmitter requests it. A possible solution to counter the delayed status reporting is to let the transmitter request an RLC status report often via the *poll bit* [81]. While this reduces signaling latency, it introduces a large signaling overhead, which may be tolerable for eMBB applications but is significant for IoT-NTN scenarios.

Increasing the Number of HARQ Processes

An alternative solution to reduce or eliminate the stop-and-wait gaps is to increase the number of HARQ processes, N_{HARQ}, together with using multiple transport block grant (MTBG) and bundled ACK to reduce the associated signaling overheads [25]. However, there are a few issues with this approach. First, the use of MTBG and ACK bundling increases the system complexity and may not be implemented in all networks, especially those catering to low-cost

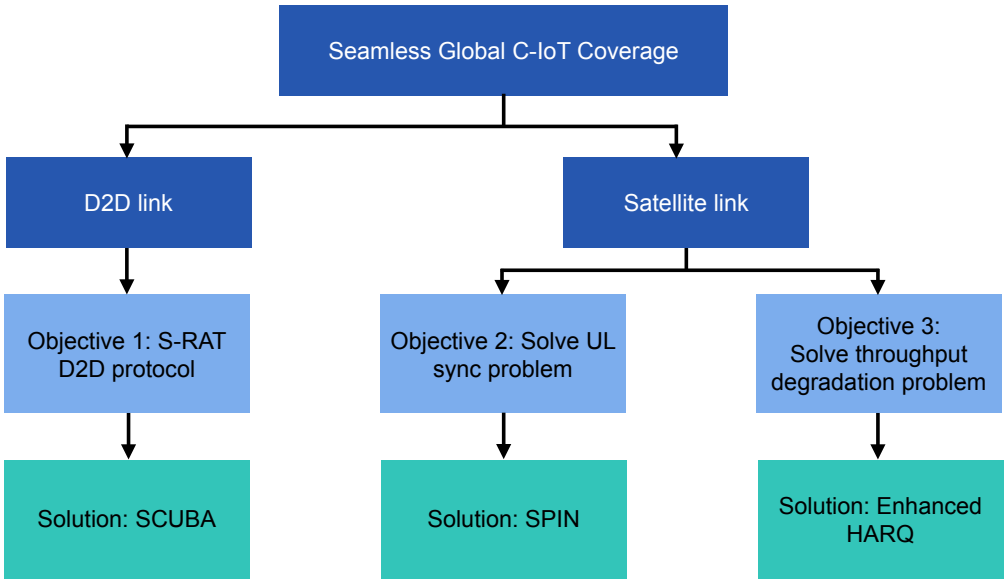


Figure 1.2: The overview of our research objectives and proposed solutions.

IoT devices. Furthermore, ACK bundling is only efficient when the iBLER is small. But as discussed earlier, a lower iBLER results in poorer spectral efficiency, lower speed, and higher UE power consumption. Therefore, increasing the N_{HARQ} to fill the stop-and-wait gaps is not ideal for IoT-NTN systems.

1.3 Research Objectives, Challenges, and Contributions

In this section, we provide a summary of our research objectives based on the discussions in the previous sections. We also discuss the challenges associated with the research problems before listing the main contributions of our thesis. An overview of the research objectives and our proposed solutions to address the problems are shown in Figure. 1.2. As discussed at the beginning of this chapter, our main research goal is to enable seamless global coverage to all C-IoT UEs which can be achieved by using D2D and satellite links. As discussed in the previous section, we identify several problems associated with the C-IoT UEs in D2D/sidelink and satellite links. In this thesis, we develop and analyze solutions to address these problems.

1.3.1 Sidelink Communication

To implement D2D links for C-IoT UEs, our objective is to design a sidelink protocol which acts as a secondary RAT (S-RAT) and can coexist with the primary RAT (P-RAT) on the devices. As C-IoT UEs are low-cost devices, the main challenge is to reuse the single radio on the device and operate S-RAT in a time division multiplexed (TDM) manner with the P-RAT. To this end, we need to analyze the feasibility by assuming appropriate traffic models for both P-RAT and S-RAT. As the prospective protocol operates on unlicensed frequency bands, it is also important to study the regulations governing the concerned bands and consider that in the protocol design. In addition to the insights obtained from the feasibility analysis, we need to consider the power and latency constraints associated with C-IoT applications. More importantly, to provide seamless D2D connectivity, the protocol design should additionally make sure that it works regardless of the network coverage scenarios. To this end, we design a new low-cost RAT protocol, called Sidelink Communications on Unlicensed BAnds (SCUBA), which can be implemented on cellular devices such that it coexists with the legacy cellular protocol by operating as an S-RAT in a time division multiplexed manner using the existing radio hardware. Our key contributions in the context of SCUBA are as follows:

I We present a detailed feasibility analysis of an in-device TDM SL protocol for low-cost half-duplex frequency division duplexing (HD-FDD) devices, where we identify unoccupied time-slots in cellular UEs to accommodate an S-RAT for SL communications, and also specify requirements for the new SL-U protocol to meet regulatory requirements in the European and North American regions.

II We develop a novel MAC protocol called SCUBA, the key features of which include SL discontinuous reception (DRX) mechanism to provide a flexible battery-latency trade-off, and synchronization methods to establish successful SCUBA links between devices that are out-of-sync with each other. To the best of our knowledge, our protocol

SCUBA is the first D2D protocol which is particularly designed as a secondary RAT which can function by sharing the radio in a TDM manner with the primary RAT.

1.3.2 Self-Positioning in IoT NTN

In the above, we discussed the challenges associated with the design of an S-RAT protocol to realize the D2D links and also listed the corresponding contributions of this thesis. While D2D communication enabled by SCUBA provides seamless coverage to the C-IoT UEs in close proximity, satellite communication enables global coverage by extending the service to C-IoT UEs located in remote regions and/or scattered over large geographical area. As discussed in Section 1.1, we identify two main problems associated with the C-IoT UEs in the satellite link. The first problem in IoT NTN is the UL synchronization failure due to the large Doppler offset caused by the velocity of the satellite. We design a self-positioning algorithm for the UE based on the available DL broadcast signals such that the location information can be used to compute and pre-compensate the Doppler offset in UL signals. One of the main challenges in designing such an algorithm is to achieve sufficient accuracy which is often limited by the low sampling rate in the C-IoT UE. Additionally, we need to ensure that the positioning solution does not result in significant battery drain or demand high computational complexity. To this end, we design Synchronization signal-based Positioning in IoT Non-terrestrial networks (SPIN) which enables the IoT UEs to tackle the UL synchronization problem. We list our key contributions related to this work as follows:

I We design a new positioning algorithm called SPIN, by exploiting appropriate state-of-the-art signal processing techniques including fast Fourier transform (FFT), correlation, and fine resolution curve fitting, and joint TDOA and FDOA based positioning, with an objective to achieve the theoretical bounds of position and velocity estimation accuracy.

II A major novelty of SPIN is the capability to perform position tracking in a time-

multiplexed manner with the underlying communication. Such a tracking mechanism provides continuous positioning and hence guarantees uninterrupted cellular connection in a mobile IoT UE with long connected mode.

III In the best case, SPIN results in $29 - 64\%$ battery life savings over GNSS based solution, whereas it is $12 - 43\%$ in the worst case. The computational complexity in any step of SPIN is less than 30 million operations per second (MOPS), which is less than the existing most computationally demanding case in NB-IoT.

1.3.3 HARQ in IoT NTN

The second problem associated with the C-IoT UEs in NTN is the reduction in the throughput due to the extended time of flight between the UEs and the BS. Since the C-IoT UEs are low-complexity and power-critical devices, the solution to this problem should not demand high computational complexity or signaling overhead. To tackle the throughput degradation problem associated with large propagation delay in NTN, we design smarter HARQ scheduling techniques. The key contributions of our thesis in this context are listed as follows:

I We identified that the fixed processing delays between grant and data, and data and acknowledgment message are the key limiting factors to increasing the throughput by utilizing the bidirectional overlap of signals in the air. Based on this finding, we design flexible HARQ scheduling techniques for IoT-NTN systems to increase the UL and DL resource utilization efficiency and throughput. To this end, we develop flexible DL data to acknowledgment (DD2A) delays and UL grant to data (UG2D) delays for more efficient bidirectional signaling.

II Our proposed HARQ scheduling solution with flexible UG2D delay delivers throughput gain of 28% and 31% for LTE-M and NB-IoT, respectively. The major highlights of

our solution are the throughput gain obtained without extra resources, and the minimal or no signaling overhead.

1.4 Organization of the Thesis

The remainder of this thesis is organized as follows. We present the feasibility analysis for sidelink communication and the SCUBA protocol design, analyses, and numerical results in Chapter 2. In Chapter 3, we extend the application of SCUBA to all cellular coverage scenarios by designing different synchronization mechanisms. In Chapter 4, we provide the design of our positioning algorithm, SPIN, which primarily solves the UL synchronization problem in IoT NTN. We develop enhanced HARQ scheduling techniques for IoT NTN in Chapter 5. Finally, we conclude the thesis in Chapter 6 by summarizing our work and providing possible future research directions.

Chapter 2

SCUBA: Sidelink Communication on Unlicensed Bands

2.1 Introduction

Towards the final goal of enhancing the performance of C-IoT networks by closing the network coverage gaps, we first focus on designing a D2D protocol for low-cost IoT UEs. As part of reducing the cost and power consumption of IoT and MTC devices, the use of HD-FDD operation is often considered [56, 57, 82–84]. Applying SL on unlicensed bands for HD-FDD devices using existing commercial D2D protocols such as Bluetooth [44], Zigbee [45], and Wi-Fi-Direct [46] is expensive, since these radio access technologies require an additional radio chain. Furthermore, these technologies also require manual device pairing and repeated user interventions [58–60], which reduce the appeal of D2D communications for IoT and MTC applications. Alternative D2D solutions that have been proposed in the literature may counter a few of the above challenges, but still fall short of meeting the needs of incorporating SL communications in low-cost low-power IoT devices (as explained in Section 1.2.1). For example, the solutions in [47–49, 51] achieve D2D communication in unlicensed bands, but require significant and continued assistance from a centralized BS for successful operation. This undermines the benefits achievable with the use of SL. Therefore, we develop a novel protocol called Sidelink Communications on Unlicensed BAnds (SCUBA) for D2D communications in unlicensed bands.

To this end, we first explore the feasibility of a cost-efficient solution that integrates D2D-U along with cellular operation using a single radio that can operate even on, say, category-M1 (Cat-M1) LTE UEs, which typically support only one radio due to cost constraints. Towards developing an in-device multiplexed D2D-U protocol, we investigate the feasibility of peer-to-peer communication as an overlay RAT, with legacy cellular operation as the P-RAT. Therefore, we require a prospective D2D-U protocol to be compatible with all forms of duplexing applied within the device, including the elementary HD-FDD used by Cat-M1 LTE-M devices. These prerequisites also restrict us from directly adopting existing radio interface technologies, such as Bluetooth, ZigBee, or Wi-Fi-Direct. Considering these targets, we initiate the process of designing a customized D2D-U protocol by conducting a feasibility study and suggesting operating guidelines for the protocol. Such a protocol ideally fits within the framework of MulteFire, which aims at enabling new wireless networks with LTE-like operations in the unlicensed bands [85], for standardization, or for use as a proprietary solution.

We begin by examining the regulations that govern the use of our chosen unlicensed frequency bands in terms of the permissible operating bandwidth (BW), transmission power, and spectrum usage. We then employ standard MTC traffic models to determine the traffic handling requirements to be specified in the prospective D2D-U protocol to ensure duty cycle compliance mandated by regulations. Further, we provide an analysis of the MAC procedure of the legacy cellular operation to identify time opportunities for multiplexing D2D-U, and suggest possible features to be incorporated in a custom D2D-U protocol to ensure coexistence with cellular communication. We verify our analysis by considering MTC as the P-RAT and simulate the MAC layer of the LTE-M standard with realistic MTC traffic. Based on our results, we suggest techniques to be incorporated within the custom protocol to enhance D2D-U communications.

Based on the insights obtained from the feasibility analysis, we then design a new protocol called SCUBA, which

- is compatible with UEs that use HD-FDD operation,
- uses the existing single cellular radio to function in a time division multiplexed (TDM) manner as the S-RAT on the device,
- meets typical SL latency and battery-life targets,
- conforms to regulatory requirements in various geographical regions, and
- provides flexibility for latency-power trade-off.

While transitioning SL operation to the unlicensed bands and functioning with a single radio architecture solution enabled by the TDM operation reduce the monetary cost associated with SCUBA, our proposed protocol also presents a low opportunity cost by coexisting with the underlying P-RAT without requiring any modifications to the legacy cellular operation.

The rest of the chapter is organized as follows. In Section 2.2, we discuss the preliminaries which include the PHY and the UE modes. We present the feasibility analysis in Section 2.3 and its numerical results in Section 2.4. The system model is provided in Section 2.5 and the traffic model is shown in Section 2.6. We present the SCUBA design in Section 2.7. We provide power consumption and collision rates analyses in Section 2.8, and their simulation results in Section 2.9. In Section 2.10, we discuss our protocol by highlighting its salient features and identifying potential future work. Finally, a summary of the chapter is provided in Section 2.11. Furthermore, for ease of reading, we have provided lists of important notations in SCUBA and their meanings in Table 2.3.

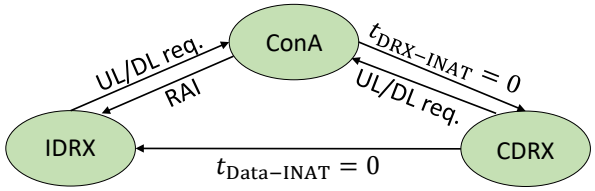


Figure 2.1: State transition diagram for UE operation modes.

2.2 Preliminaries

2.2.1 Physical Layer

We consider low-cost LTE-M Cat-M1 UEs that use a maximum of six physical resource blocks occupying a BW of $W \leq 1.4$ MHz. The UL and DL operations in LTE-M are specified in units of 1 ms long subframes (SFs). Further to time division duplexing, Cat-M1 devices also separate the UL and DL signals in frequency to operate in a HD-FDD fashion. The UL and DL SFs are separated in time by a guard period to allow the hardware to switch between transmitter (Tx) and receiver (Rx) paths. Note that since we consider HD-FDD operation for the underlying cellular RAT, much of our analyses also hold true when constraints are relaxed to accommodate other forms of duplexing.

2.2.2 UE Modes

Most cellular standards, such as LTE and LTE-M, allow for flexible battery life and latency tradeoff by using DRX, which enables UEs to sleep periodically for pre-defined cycles. Accordingly, a UE may be in one of two states: RRC connected or RRC idle discontinuous reception (IDRX) state. In IDRX, the UE sleeps for a majority of the time and wakes up only periodically to page on pre-defined UE-specific SFs called paging occasions (POs). When the UE either receives a valid DL page from the BS during its PO, or when it has UL data itself, it transitions back into the RRC connected state. An illustration of the mode transi-

tions is provided in the state diagram in Figure. 2.1. The RRC connected state begins with an active period (ConA) during which the UE may have continuous UL and DL communications with the BS. During ConA, the UE maintains two timers, the DRX inactivity timer ($t_{\text{DRX-INAT}}$) and the data inactivity timer ($t_{\text{Data-INAT}}$), with $t_{\text{DRX-INAT}} \leq t_{\text{Data-INAT}}$, to switch its operating modes. Both these timers restart every time a UL/DL data transmission begins. Upon encountering $t_{\text{DRX-INAT}} = 0$, the UE transitions from ConA to the RRC connected discontinuous reception mode (CDRX). CDRX consists of extended periods of time when the UE sleeps to conserve power, and periodically turns on during the CDRX-ON period. Upon the further expiry of $t_{\text{Data-INAT}}$, it then transitions into IDRX. On the other hand, if the UE receives a DL transmission during CDRX-ON, it reverts back to ConA and restarts $t_{\text{DRX-INAT}}$ and $t_{\text{Data-INAT}}$. Optionally, the UEs may also transition directly from the ConA to IDRX mode upon receiving such an instruction from the BS using release assistance indication (RAI). As part of our feasibility analysis, we examine each of the above discussed modes to identify time opportunities for D2D-U.

2.3 Feasibility Analysis

We begin our feasibility analysis by studying the regulatory requirements imposed on accessing the unlicensed bands.

2.3.1 Regulatory Requirements

A primary aspect of communications in the unlicensed band is to comply with regulations governing the usage of the chosen spectrum that are designed to ensure fair use by all participants. We consider the use of $865 - 868$ MHz and $902 - 928$ MHz bands for D2D-U in Europe and United States (US), respectively, as these frequencies are relatively less occupied

portions of the unlicensed spectra. Nevertheless, depending on coverage requirements and hardware constraints within the devices, alternative unlicensed bands, such as the 2.4 GHz Industrial Scientific and Medical (ISM) band, could also be used, and a similar analysis is applicable for any considered frequency spectrum.

902-928 MHz

Section 15.247 of US Code of Federal Regulations (CFR) [86] specifies regulations imposed on operations in the $902-928$ MHz band for digitally modulated systems, such as operations based on LTE [87]. CFR requires that the 6 dB bandwidth (W_{6dB}) of all digitally modulated systems to be $W_{\text{6dB}} \geq 500$ kHz, without a restriction on the maximum occupied BW. The regulations limit the peak conducted output power of the transmitter to a maximum of 1 W, and the 100 kHz band power in the undesired frequencies to 20 dB lower than the peak power in the desired 100 kHz bands. Furthermore, apart from the total signal power, the power spectral density (PSD) of the signal is also mandated to be $\psi \leq 8$ dBm/3 kHz. All the above requirements can be met with physical layer power and BW specifications of the prospective D2D-U protocol.

865-868 MHz

The regulatory requirements in Europe for the frequency band $865-868$ MHz are provided by the European Telecommunications Standards Institute (ETSI) [88, 89]. The regulations require the effective radiated power, $P_{\text{erp}} \leq 25$ mW and the spectrum to be accessed with a duty cycle, $\tau \leq 1\%$. As an alternative to complying with limits on τ, devices may also choose to use clear channel assessment (CCA) with adaptive frequency agility while using the spectrum [88, 89]. However, since implementing CCA requires high sensitivity at the receiver to sense transmissions over long distances [85], we suggest applying limits on τ to

Table 2.1: MTC traffic characteristics [90, Annex. A].

Traffic type	Traffic model	IAT (s)	Data size (bits)
Regular report	Periodic	$300 - 3600$, 60 (optional)	$n_{\text{reg}} = 1000$, 10000 (optional)
Triggered report	Poisson	Mean $= 30$	$n_{\text{trig}} = 256, 1000$

be a more suitable strategy for low-cost MTC devices. Unlike US regulations that can be addressed by physical layer protocol specifications, addressing limit on τ requires evaluating spectrum usage for various traffic scenarios encountered in D2D communications.

Duty Cycle Analysis

Depending on the D2D traffic being served, the protocol may need to additionally ensure that $\tau \leq 1\%$ is met by, say, introducing buffer-times for overflowing packets. In the following, we analyze the duty cycle using different D2D traffic and the time duration required to complete data transmission. To this end, we adopt the MTC traffic models specified by 3GPP [90], with the notion that traffic in MTC cellular devices follows the MTC traffic model in both the cellular link and the sidelink. We summarize the pertinent models in Table 2.1 with their packet inter-arrival time (IAT) and data size[3].

For UEs involved in periodic reporting, the duty cycle can be computed as

$$\tau_{\text{reg}} = \frac{t_{\text{reg}}}{T_{\text{IAT}}}, \tag{2.1}$$

where t_{reg} is the time duration over which the D2D-U data is transmitted and T_{IAT} is the data IAT. As t_{reg} is the time duration corresponding to the number of transport blocks (TBs)

[3]Note that we consider the traffic model specified in TR 36.888 [90] since they present a worse traffic condition, i.e., lower IAT, than that specified in the updated traffic model [91].

spanning the transmission of n_{reg} data bits, it can be computed as

$$t_{\text{reg}} = \left(\frac{n_{\text{reg}}}{N_{\text{TBS}} - N_{\text{OH}}} \right) t_{\text{TBS}}, \tag{2.2}$$

where t_{TBS} is the time occupied by 1 TB, N_{TBS} is the transport block size (TBS) chosen based on the transmission BW and the modulation and coding scheme (MCS) used, and N_{OH} is the total number of bits consumed by the cumulative cross-layer overheads, such as headers from multiple layers. The worst-case duty cycle can then be computed using (2.1) with the lowest IAT from Table 2.1.

In contrast to the analysis for periodic MTC traffic, since data packets arrive randomly when the UE is involved in triggered reporting, the duty cycle computation further requires determining the number of arrivals in an observation window, T_{obs}. For a Poisson traffic model where the packet IAT is an exponentially distributed random variable with a mean of $\frac{1}{\lambda_{\text{trig}}}$, we obtain the probability of observing m arrivals in T_{obs} as [92, Ch. 4]

$$p(m, T_{\text{obs}}, \lambda_{\text{trig}}) = \frac{(\lambda_{\text{trig}} T_{\text{obs}})^m \mathrm{e}^{-\lambda_{\text{trig}} T_{\text{obs}}}}{m!}. \tag{2.3}$$

For an exponentially distributed IAT, the total number of packet arrivals, M, within T_{obs} could approach infinity, and therefore a finite duty cycle cannot be guaranteed with absolute certainty. However, the specification indicates that the duty cycle requirement suffices to be met by the devices 99% of the time [88]. Thus, we determine the value of M that achieves the 99th percentile condition by solving

$$\sum_{m=0}^{M} p(m, T_{\text{obs}}, \lambda_{\text{trig}}) = 99\%. \tag{2.4}$$

We then compute the duty cycle as

$$\tau_{\text{trig}} = \frac{\widetilde{M} \cdot t_{\text{trig}}}{T_{\text{obs}}}, \tag{2.5}$$

where $\widetilde{M}$ is the solution of (2.4), and $t_{\text{trig}} = \left(\frac{n_{\text{trig}}}{N_{\text{TBS}} - N_{\text{OH}}}\right) t_{\text{TBS}}$, similar to (2.2). We present the results of (2.1) and (2.5) later in Section 2.4.

2.3.2 Sidelink Transmission Opportunities

Since we consider D2D-U as an in-device time division multiplexed S-RAT in practice, we identify the time opportunities available between cellular operation for D2D-U. We perform this analysis for each of the three modes that the UE operates in, i.e., ConA, CDRX, and IDRX. To this end, we include the time consumed in setting up the connection when the UE transitions from IDRX to ConA (t_{RRC}) within the ConA mode. Similarly, we include the DRX inactivity time also within ConA. Therefore, the cumulative time spent by the UE in ConA and RRC connected modes are

$$T_{\text{ConA}} = t_{\text{RRC}} + t_{\text{data}} + t_{\text{DRX-INAT}} \tag{2.6}$$

$$T_{\text{CONN}} = t_{\text{RRC}} + t_{\text{data}} + t_{\text{Data-INAT}}, \tag{2.7}$$

respectively, where t_{data} is the total time required to complete the UL/DL transmission.

Connected Active Mode

The first time opportunity we identify for D2D-U is the guard period included for switching between cellular UL and DL when using HD-FDD or TDD operation. This guard period varies between 1 to 3 SFs depending on the use-case and device complexity. However, our chipset suppliers report that it takes only 80 μs and 250 μs $-$ 450 μs for Tx-Tx/Rx-Rx and

Tx-Rx/Rx-Tx switching, respectively, even without including any switching hardware optimization. Thus, we identify at least $\frac{1}{2}$ SF or 0.5 ms of time to be available for a short D2D-U message transfer. If the actual switching time in the UE is similar to the aforementioned values, this D2D-U opportunity is nearly guaranteed with bidirectional cellular link traffic. Furthermore, the UE may also often enjoy free SFs in between its cellular operation when it has no UL/DL for communication. Additionally, the DRX inactivity period, when the UE is constantly listening for a potential DL message, may also be used to transmit a short D2D-U message if the possible risk of losing a DL message is acceptable.

Connected DRX Mode

The free time available in the CDRX mode is significantly less stochastic when compared to the ConA mode. A UE in CDRX is in either CDRX-ON or CDRX-OFF state, with the CDRX-ON period being the only time duration when the UE is busy with the cellular link. While the CDRX-OFF duration is entirely available for D2D-U, their positions in absolute time are not always fixed and are customizable by the network. Therefore, the D2D-U protocol requires a mechanism where a UE can advertise its CDRX-OFF locations, possibly using a broadcast signal.

Idle DRX Mode

Unlike operation in the RRC connected modes, the UE activities in the IDRX mode are deterministic in nature. In every IDRX cycle, the UE wakes up during its IDRX PO and is available at all other times for D2D-U.

2.3.3 Mode Probabilities

As the next step in determining the feasibility of achieving an in-device D2D-U, we compute the mode probabilities for each of ConA, CDRX, and IDRX modes. This analysis drives the design of the D2D-U protocol by suggesting UE operating conditions that potentially need to be prioritized. The mode probabilities depend on the cellular traffic that the UE serves and the system parameters set by the network. In the following, we compute the probabilities individually for both kinds of cellular traffic, Poisson and periodic, served by the UE.

Poisson Cellular Traffic

The UE is present in ConA at any point in time as long as it has obtained one or more UL/DL request(s) in the previous T_{ConA} duration of time. Therefore, the probability of encountering a UE to be in ConA is

$$\begin{aligned} p_{\text{ConA,trig}} &= \sum_{m=1}^{\infty} p(m, T_{\text{ConA}}, \lambda_{\text{trig}}) \\ &= 1 - \mathrm{e}^{-\lambda_{\text{trig}} T_{\text{ConA}}}. \end{aligned} \tag{2.8}$$

Next, we determine the probability that the UE is in the RRC connected mode, i.e., either ConA or CDRX modes. As described in Section 2.3.2, the UE remains in the RRC connected mode until the expiry of the data inactivity timer. Hence, the probability of encountering the UE in connected mode is

$$\begin{aligned} p_{\text{CONN,trig}} &= \sum_{m=1}^{\infty} p(m, T_{\text{CONN}}, \lambda_{\text{trig}}) \\ &= 1 - \mathrm{e}^{-\lambda_{\text{trig}} T_{\text{CONN}}}. \end{aligned} \tag{2.9}$$

Since CDRX fills the remainder duration of the RRC connected mode after encountering no

UL/DL for a period of DRX inactivity time, the CDRX probability can be expressed as

$$\begin{aligned} p_{\text{CDRX,trig}} &= p_{\text{CONN,trig}} - p_{\text{ConA,trig}} \\ &= \mathrm{e}^{-\lambda_{\text{trig}} T_{\text{ConA}}} - \mathrm{e}^{-\lambda_{\text{trig}} T_{\text{CONN}}}. \end{aligned} \tag{2.10}$$

Finally, as the UE is either in the RRC connected or the IDRX mode, the IDRX probability can be calculated to be

$$\begin{aligned} p_{\text{IDRX,trig}} &= 1 - p_{\text{CONN,trig}} \\ &= \mathrm{e}^{-\lambda_{\text{trig}} T_{\text{CONN}}}. \end{aligned} \tag{2.11}$$

Periodic Cellular Traffic

With periodic cellular traffic being served, the probability of encountering a UE in ConA, CDRX, and IDRX modes are

$$p_{\text{ConA,reg}} = \frac{T_{\text{ConA}}}{T_{\text{IAT}}}, \tag{2.12}$$

$$\begin{aligned} p_{\text{CDRX,reg}} &= p_{\text{CONN,reg}} - p_{\text{ConA,reg}} \\ &= \min\left\{\frac{T_{\text{CONN}}}{T_{\text{IAT}}}, 1\right\} - \frac{T_{\text{ConA}}}{T_{\text{IAT}}}, \end{aligned} \tag{2.13}$$

$$\begin{aligned} p_{\text{IDRX,reg}} &= 1 - p_{\text{CONN,reg}} \\ &= \max\left\{\left(1 - \frac{T_{\text{CONN}}}{T_{\text{IAT}}}\right), 0\right\}, \end{aligned} \tag{2.14}$$

respectively. Note that the $\min\{\cdot\}$ and $\max\{\cdot\}$ conditions in (2.13) and (2.14), respectively, handle the rare scenarios where $T_{\text{CONN}} > T_{\text{IAT}}$, i.e., when the reporting data is large and/or the IAT is short enough resulting in the data inactivity timer not expiring before the beginning of the next reporting activity. In such cases, the UE remains in the RRC connected mode at all

Table 2.2: Simulation settings for feasibility evaluation.

Parameter	Value	Parameter	Value
$\lambda_{\text{trig}}^{-1}$	30 s	CDRX-ON duration	20 ms
T_{IAT} (periodic)	5 min	Long CDRX cycle	640 ms
t_{data}	250 ms	IDRX cycle	640 ms
$t_{\text{Data-INAT}}$	10 s	t_{RRC}	100 ms

times and never enters the IDRX state. However, we impose the condition that $T_{\text{ConA}} < T_{\text{IAT}}$, which generally holds true for practical application scenarios and network settings.

The design of any D2D-U protocol must consider the above discussed mode probabilities to ensure satisfactory latency targets. In the next section, we present the numerical results of mode probabilities and their impact on the protocol design.

2.4 Feasibility Evaluation

In this Section, we verify and validate the analysis in Section 2.3 by simulating the MAC layer of LTE-M using the traffic models provided in TR 36.888 [90]. We summarize the network parameters we used in Table 2.2, whose values conform with the set of values allowed by the 3GPP specification [81].

2.4.1 Duty Cycle

For each MTC use-case, we compute the worst-case τ using the largest data size and the shortest IAT to examine if any additional measures are required to be implemented by the D2D-U protocol to ensure compliance with regulations. Note that we do not use the "optional" IAT and packet sizes specified in the traffic models (see Table 2.1), as they only represent rare operating conditions. To evaluate the worst-case duty cycle values, we assume that the UE is free from cellular traffic, i.e., UE is always available for D2D-U. We also consider the

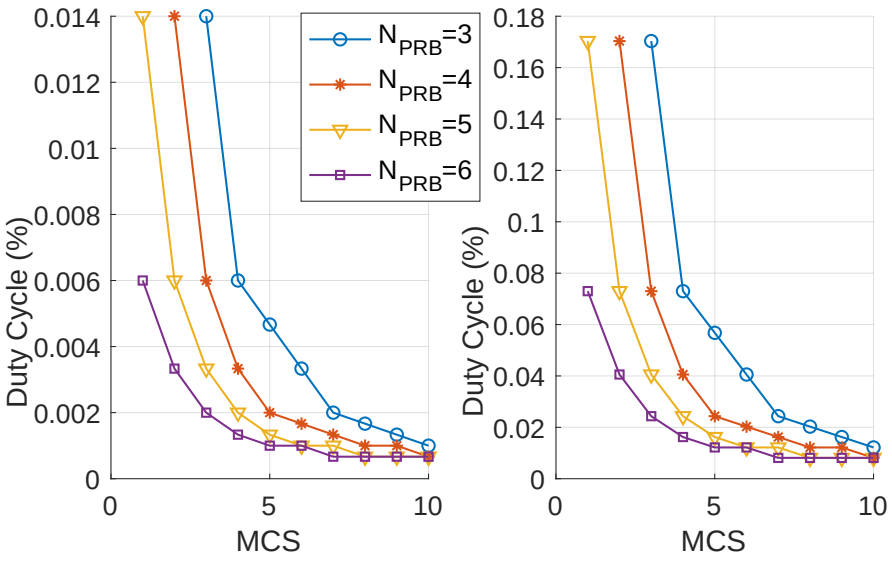


Figure 2.2: Duty cycle vs MCS for periodic traffic with packet size = 1000 bits and IAT = 5 min (left), and Poisson traffic with packet size = 1000 bits and mean IAT = 30 s (right).

D2D-U protocol to be built along the lines of the LTE specifications, similar to LTE-D2D. Accordingly, we employ MCS and TBS values from TS 36.213 [93, Table 7.1.7.2.1-1, Table 8.6.1-1]. To investigate transmission in poor channel conditions, we only examine using lower order MCS schemes. We also limit the number of physical resource blocks (PRBs), N_{PRB}, to $3 \leq N_{\text{PRB}} \leq 6$, considering the hardware limitations of LTE-M devices. Further, we use $N_{\text{OH}} = 19$ bytes [94], and evaluate the duty cycle only for $N_{\text{TBS}} > N_{\text{OH}}$ to ensure non-zero number of data bits within a TB.

Periodic MTC traffic: Using the worst-case values from Table 2.1, i.e., $n_{\text{reg}} = 1000$ bits and $T_{\text{IAT}} = 300$ s in (2.1) and (2.2), we obtain the results shown in the left-half of Figure. 2.2 for a standard TBS length of 1 SF, i.e., $t_{\text{TBS}} = 1$ ms. We observe that across different values of N_{PRB} and MCS schemes, the duty cycle associated with complete data transmission including overheads is well below the $\tau \leq 1\%$ limit imposed by regulations.

Poisson MTC traffic: The duty cycle for random traffic arrival is defined over $T_{\text{obs}} = 1$ hour [88]. We solve (2.4) with $\epsilon = 0.01$ and find that $\widetilde{M} = 146$ corresponds to the 99th

percentile duty cycle value. Next, we use the worst-case data traffic size of $n_{\text{trig}} = 1000$ bits as given in Table 2.1 to compute the duty cycle using (2.5). We plot the results for varying values of the chosen N_{PRB} and MCS schemes on the right-half of Figure. 2.2. We notice that $\tau \leq 1\%$ is guaranteed under all conditions. Therefore, recalling the regulatory limits from Section 2.3.1, if $P_{\text{erp}} \leq 25$ mW is applied on D2D-U transmission, no special accommodations need to be enforced in the protocol. Note that every D2D transmission may also be further reciprocated with an acknowledgment message, which does not impact the duty cycle computations as the acknowledgment is transmitted by a different UE.

2.4.2 Mode Probabilities

For establishing the cellular mode probabilities, we perform a Monte-Carlo simulation by generating 10^5 message arrivals each for both types of traffic models. We consider a pair of UEs, the packet arrivals for which are independent of each other, and measure the probabilities of the UE pair in all combinations of cellular modes. We compare these probabilities against the analytical values computed as

$$p_{\alpha\beta} = \begin{cases} p_\alpha^2, & \text{for} \quad \alpha = \beta \\ 2p_\alpha p_\beta, & \text{for} \quad \alpha \neq \beta \end{cases} \tag{2.15}$$

where $\alpha, \beta \in \{\text{ConA, CDRX, IDRX}\}$, and the individual probabilities are computed using (2.8)-(2.14). Across our simulations, we vary the DRX inactivity timer between $1 \leq t_{\text{DRX-INAT}} \leq 2560$ ms in steps specified by TS 36.311 [81]. We show the mode-pair probabilities in Figure. 2.3 and Figure. 2.4 for Poisson and periodic MTC traffic models, respectively. We notice that for both types of traffic, a pair of UEs are in IDRX for the majority period of time, which indicates sufficient time opportunity for D2D-U. For the busier Poisson traffic with a mean IAT of 30 s, we encounter a larger number of conditions where at least

one UE is in ConA. Since time opportunities available for D2D-U is limited in ConA, the D2D-U latency performance may suffer when UEs are busy in the cellular link. Therefore, applications that involve periodic reporting with lower IAT, e.g., smart metering and health monitoring, potentially enjoy superior latency performance while talking between devices in the unlicensed band. On the other hand, the UEs associated with triggered reporting, such as container tracking or smart public lighting, may occasionally face relatively poorer latency in the D2D-U link. Nevertheless, we note from Figure. 2.3 that the worst-case $t_{\text{DRX-INAT}} = 2.56$ s still experiences over 80% of the cases where both UEs are in one of the DRX modes, indicating sufficient time opportunity for D2D-U.

We further focus on the condition where $t_{\text{DRX-INAT}} = 100$ ms, which is reported in the literature to be a commonly used value for the DRX inactivity timer [95]. From Figure. 2.3 and Figure. 2.4 we notice that this setting guarantees that a pair of UEs are in one of the DRX modes for at least 97% of the time for either type of cellular traffic. UEs in the DRX modes are largely free from cellular operation. Therefore, based on the mode probabilities in Figure. 2.3 and Figure. 2.4, and the results in Figure. 2.2, which show that the worst-case duty cycle is lower than 0.2%, it is clear that D2D-U transmission can be completed within the available free time. These results also indicate that devising a protocol that utilizes the DRX sleeping duration in CDRX and IDRX is potentially sufficient to integrate low-power low-latency D2D-U operation within legacy cellular devices.

In the above, we presented a detailed feasibility analysis of an in-device TDM SL protocol for low-cost HD-FDD devices, where we identified unoccupied time-slots in cellular UEs to accommodate an S-RAT for SL communications. We also specified requirements for the new SL-U protocol to meet regulatory requirements in the European and North American regions. Despite addressing these issues, we still face the following additional challenges for SL-U. Since we target deploying SL-U to operate on low-cost and low-complexity devices, such as

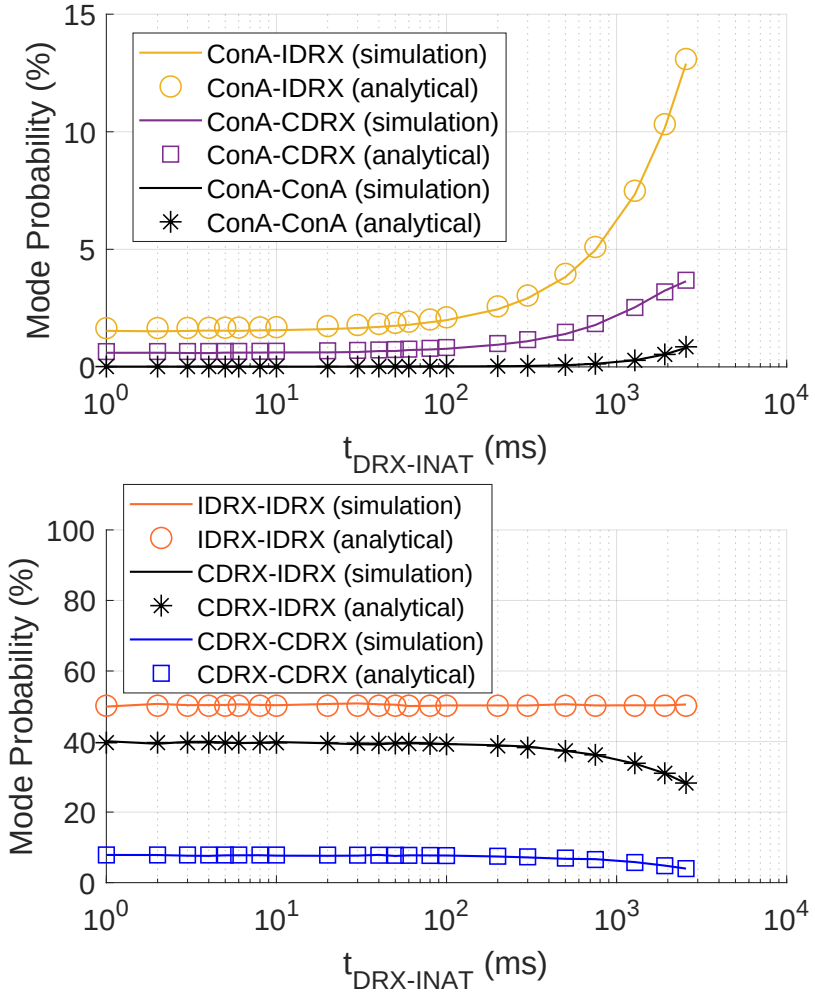


Figure 2.3: Probability of RRC states vs $t_{\text{DRX-INAT}}$ for a Poisson MTC traffic.

MTC and NB-IoT UEs, SL-U should ensure extended UE battery life while also providing reasonable control- and user-plane latency. Further, since we aim to operate independently without constant control and assistance from a central BS, SL-U must enable UEs to operate in a distributed manner, e.g., determining SL availability status of the destination UE.

Considering the above design criteria and challenges, we develop SCUBA as an additional S-RAT to cellular communication in a TDM fashion, such that it *exploits the existing hardware* radio resource during idle times. In the rest of this chapter, we present the design of SCUBA, its performance analysis, and numerical evaluations. For ease of reading, we

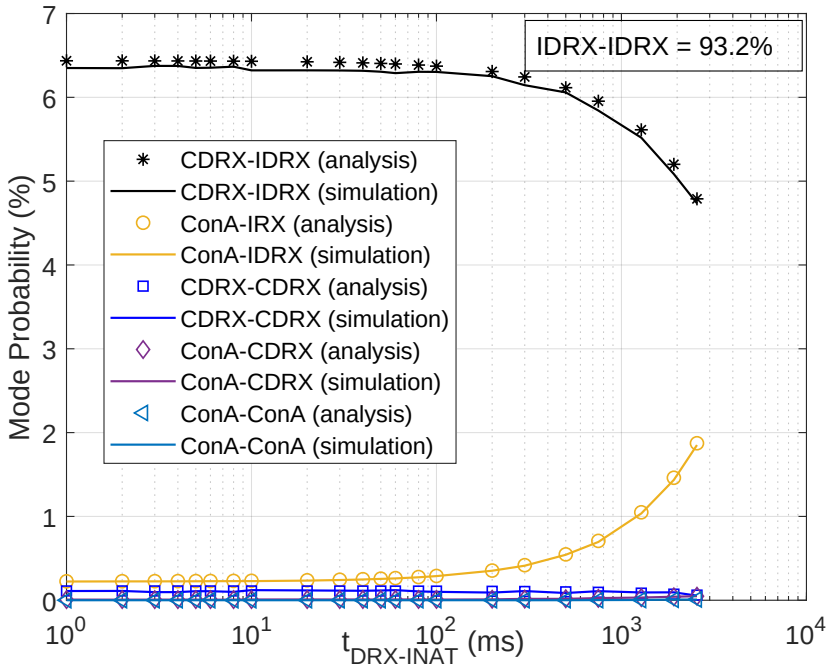


Figure 2.4: Probability of RRC states vs $t_{\text{DRX-INAT}}$ for periodic MTC traffic.

summarize the mathematical notations used in SCUBA in Table 2.3.

2.5 System Model

We consider a hybrid network of UEs with SCUBA as the S-RAT that is integrated with a primary *LTE-like* RAT such as LTE, LTE-Advanced, LTE-M, or NB-IoT. In the following, we choose LTE-M operating on an elementary HD-FDD radio as the P-RAT for the design and analyses, such that constraints adopted in this study can be relaxed to extend the analysis for devices with greater capability, e.g., full-duplex FDD as in 5G NR UEs. A conceptual representation of a SCUBA integrated network with LTE-M as the P-RAT and SCUBA enabling D2D on the unlicensed frequencies as an S-RAT is shown in Figure. 2.5.

Table 2.3: List of important notations in SCUBA.

Notation	Meaning
α_{UE}	UE identity derived from IMSI
E_{NoData}	Energy consumption during SCUBA idle state
E_{RXData}	Energy consumption during SCUBA reception
$E_{\text{SL-TX}}$	Energy consumption during SCUBA transmission
$i_{\text{SL-PO}}$	SF index of the SL-PO
N_{cluster}	Number of clusters in SL-PO
n_{dist}	SF separation between clusters in SL-PO
N_{frame}	SL-HARQ frame length in units of SFs
$N_{\text{SL-grant}}$	SF separation between SL data and SL grant
n_{SL}	SL data duration in units of SFs
N_{HARQ}	Number of parallel synchronous SL HARQ processes
N_{CDRX}	Cellular CDRX cycle in units of SFs
N_{IDRX}	Cellular IDRX cycle in units of SFs
N_{SAM}	SAM period
$N_{\text{SAM-U}}$	SAM-U transmission interval in units of SFs
$N_{\text{SAM-D}}$	SAM-D transmission interval in units of SFs
$n_{\text{SAM-U}}$	SAM-U duration in units of SFs
$n_{\text{SAM-D}}$	SAM-D duration in units of SFs
$n_{\text{SL-PO}}$	SL-PO duration in units of SFs
$n_{\text{SL-INAT}}$	SL inactivity time in units of SFs
$N_{\text{SL-DRX}}$	SL-DRX cycle in units of SFs
$T_{\text{SL-DRX}}$	SL-DRX cycle in units of radio frames

2.5.1 UE Operation in Primary RAT

We consider low-cost and low-complexity HD-FDD LTE Cat-M1 UEs [96]. Since we focus on the example of LTE-M being applied in the P-RAT, we follow the mode operations and frame structures accordingly. It can be seen that adopting the study to other P-RATs is straightforward. The UE modes of operation in P-RAT is shown in Figure. 2.1 and is discussed in detail in Section 2.2.2.

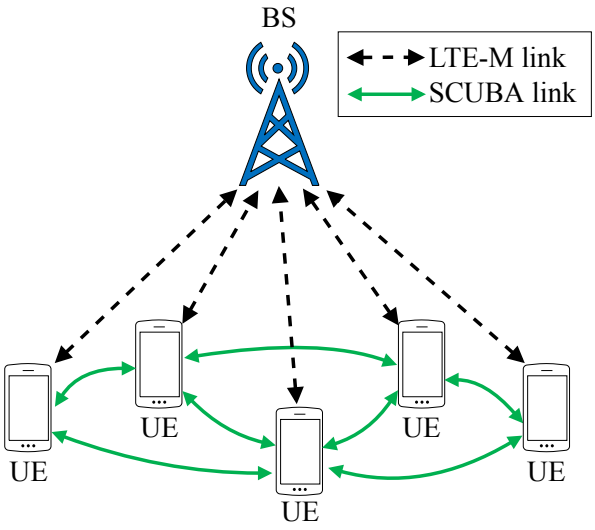


Figure 2.5: A realization of SCUBA coexisting in an LTE-M network.

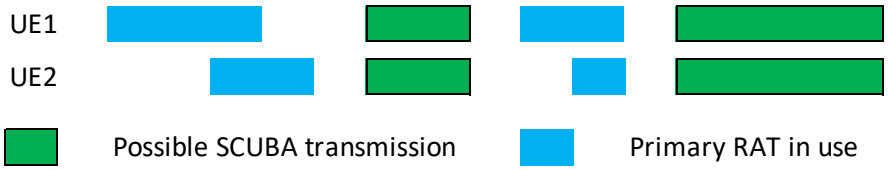


Figure 2.6: An illustration of the TDM nature of SCUBA, which operates as an S-RAT only when both devices are idle in their P-RATs.

2.5.2 UE Operation in Secondary RAT

SCUBA is an S-RAT that allows UEs to communicate directly with each other in the SL. To this end, SCUBA ensures that it operates in a TDM manner with the P-RAT, such that the cellular communication, i.e., the P-RAT, operates unimpeded considering the HD-FDD constraint. An illustration of TDM operation of SCUBA along with P-RAT is shown in Figure. 2.6, where the UEs communicate with each other via SCUBA only when both are free from their respective P-RAT.

Table 2.4: Regulatory requirements.

Parameter	Europe (865-868 MHz)	US (902-928MHz)
Maximum ERP	25 mW	1 W
Maximum duty cycle	1%	-
Minimum bandwidth	-	500 kHz

We design SCUBA to operate in the unlicensed bands of $865 - 868$ MHz in Europe and $902 - 928$ MHz in US, as they are comparatively less occupied by other unlicensed RATs. We provide the summary of the relevant regulatory requirements discussed in Section 2.3.1, which govern the choice of effective radiated power (ERP), duty cycle, and bandwidth of SCUBA in Table 2.4. Nevertheless, SCUBA is also applicable across other unlicensed spectra together with any applicable upper-layer traffic shaping that may be required to satisfy the duty cycle regulations associated with the used frequencies. These restrictions drive the physical layer design and the applicability of SCUBA across communication systems. For example, since the US Federal Communications Commission (FCC) regulates a minimum transmission bandwidth of 500 kHz to be used in the $902 - 928$ MHz band, SCUBA may not be applicable in some devices which only support NB-IoT operating bandwidth of 180 kHz.

2.6 Traffic Model

Table 2.5: Traffic model [90, Annex. A].

RAT	Traffic model	IAT (s)
Primary RAT e.g., LTE-M	Periodic	300
	Poisson	Mean = 30
SCUBA	Poisson	Mean = 30

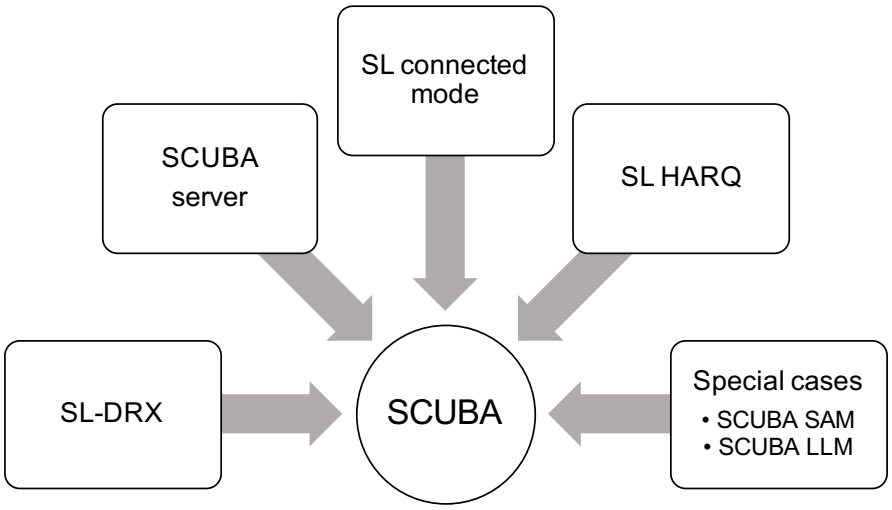


Figure 2.7: Building blocks of SCUBA protocol.

We consider the MTC traffic models recommended by 3GPP [90, Annex. A] for the design, analysis, and simulation of SCUBA by applying MTC traffic at both cellular link and SL. As shown in Table 2.5, we consider both types of traffic models with the respective IATs as specified in [90, Annex. A], i.e., periodic and Poisson data arrivals for the cellular link. A typical application of SCUBA is a network of sensors reporting to a central node or tracking devices attached to a number of assets. The data traffic between a pair of UEs in such a scenario is expected to be random in nature. Therefore, we apply Poisson traffic model from [90, Annex. A] for SCUBA traffic to evaluate its performance. Since the above Poisson traffic model with a small IAT simulates a busy traffic scenario, it also represents the worst-case condition and hence the resultant analysis and evaluations guarantee performance targets.

Table 2.6: Opportunities available for SCUBA in each cellular mode.

Cellular Mode	Cellular Operation	Availability for SCUBA
ConA	DL/UL	No
	Switch SF	Yes
	DRX inactivity	No
	RRC connection setup	No
CDRX	DRX-ON	No
	DRX-OFF	Yes
IDRX	IDRX PO	No
	IDRX sleep	Yes

2.7 SCUBA

As SCUBA is intended to operate on a device that uses an *LTE-like* protocol as the P-RAT, it borrows the PHY signal waveforms and also higher layer specifications, except the MAC layer, from the baseline protocol used within the UE. We thus focus on filling this gap by defining the MAC layer of SCUBA in this section. We define the MAC protocol by first introducing each of its fundamental structural components, as shown in Figure. 2.7, that finally integrate to provide the overall SCUBA MAC protocol.

Prior to describing our protocol, we first highlight the challenges that we face and system requirements that need to be met by SCUBA. SCUBA is intended to provide an integrated single-radio SL-U solution for cellular UEs by incorporating a TDM SL as an S-RAT that reuses the existing hardware. Therefore, the first challenge is to identify time occasions in the UE that are available for SCUBA transmission, i.e., the time instances where the UEs are free from cellular operations. For LTE-M as the P-RAT, a compiled list of the available opportunities in each of ConA, CDRX, and IDRX modes is shown in Table 2.6. The UE is available for SCUBA transmission and reception in CDRX and IDRX when it is scheduled to *sleep* for cellular communication. Therefore, the UE is free for SCUBA in all times except when it listens for a possible DL message during the CDRX-ON period or IDRX PO.

Considering typical MTC traffic types [90], these periods are sufficient for efficient SCUBA communications. Nevertheless, the switching SFs that are allocated to switch the UE hardware from the transmission to reception modes, several of which may be available when the UE is in the ConA mode, can also be used for short SCUBA message transmissions, since these durations are typically over allocated, i.e., most UEs can switch their hardware well within an entire SF. Once these available time slots are determined, a straightforward SL implementation is to let all network UEs listen for an SL message at all opportunities. Consider a portion of the network shown in Figure. 2.5, where a source UE (SRC) has SCUBA data to transmit to a destination UE (DST). A continuous listening technique is beneficial to achieve near-instant communication and therefore provides negligible control- and user-plane latency, since a DST is nearly guaranteed to be listening on a particular time-slot, given the sparse nature of MTC traffic in the cellular link [90]. However, listening for a SCUBA message on all available time slots significantly drains the UE battery. Therefore, to provide a flexible trade-off between the desired battery and latency, we borrow and extend the idea of DRX from cellular communications into SCUBA to introduce the concept of SL-DRX.

2.7.1 SL-DRX

Inspired by DRX modes present in conventional cellular communication [97, 98], we introduce SL-DRX cycles in SCUBA to provide a flexible application-controlled trade-off between network latency and UE battery life. As a solution, we specify SL paging occasions (SL-POs) during which the device wakes up to page for a potential SL message from an SRC. Every UE pages for a duration of $n_{\text{SL-PO}}$ SFs once in every $N_{\text{SL-DRX}}$ SFs, where $N_{\text{SL-DRX}}$ is the SL-DRX cycle in units of subframes. We let the application program interface (API) choose $N_{\text{SL-DRX}}$ according to the latency-power consumption trade-off it desires, including the option of setting $N_{\text{SL-DRX}} = 0$ for latency critical applications (see Section 2.7.5).

As the number of UEs in the network grows, the probability of two or more UEs sharing the same SL-PO also increases, which results in an increased rate of packet collisions. Recall that unlike other proposed SL-U solutions [47–49, 51], SCUBA does not rely on a central BS for resource scheduling or operation monitoring. Therefore, we introduce an inherent elementary collision control mechanism in choosing the SL-POs. The collision control method we propose reduces potential slot overlaps, i.e., overlap of SL-POs of different UEs and also an overlap of SL-POs with the IDRX PO of the UE P-RAT.

We begin with a brief overview of selecting the IDRX-PO, which also guides us in designing the SL-PO slot allocation. The IDRX-PO location in conventional cellular operation is computed by first locating the paging frame (PF) as [99]

$$\begin{aligned} i_{\text{PF}} &= i_{\text{SFN}} \in \mathcal{I}_{\text{SFN}} \\ \text{s.t. } & i_{\text{SFN}} \bmod T_{\text{IDRX}} = N_{\text{ID}}, \end{aligned} \tag{2.16}$$

where $\mathcal{I}_{\text{SFN}}$ is the set of system frame numbers (SFNs), whose range of values is $\mathcal{I}_{\text{SFN}} = \{0, 1, .., 1023\}$ for *LTE-like* standards, i_{SFN} is one SFN from the set, i_{PF} are those SFNs which qualify to be the paging frames,

$$N_{\text{ID}} = \frac{T_{\text{IDRX}}}{N_{\text{min}}}(\alpha_{\text{UE}} \bmod N_{\text{min}}), \tag{2.17}$$

$$N_{\text{min}} = \min(T_{\text{IDRX}},\ N_{\text{control}}), \tag{2.18}$$

$$\alpha_{\text{UE}} = \alpha_{\text{IMSI}} \bmod \beta, \tag{2.19}$$

T_{IDRX} is the IDRX paging cycle in radio frames, N_{control} is a control parameter signaled in the system information block-2 (SIB2), α_{IMSI} is the international mobile subscriber identity (IMSI) of the UE, and the value of β is 1024 for LTE, 4096 for NB-IoT, and 16384 for NB-IoT on non-anchor carrier and LTE-M. Next, the pointing index i_{s} that points to the exact

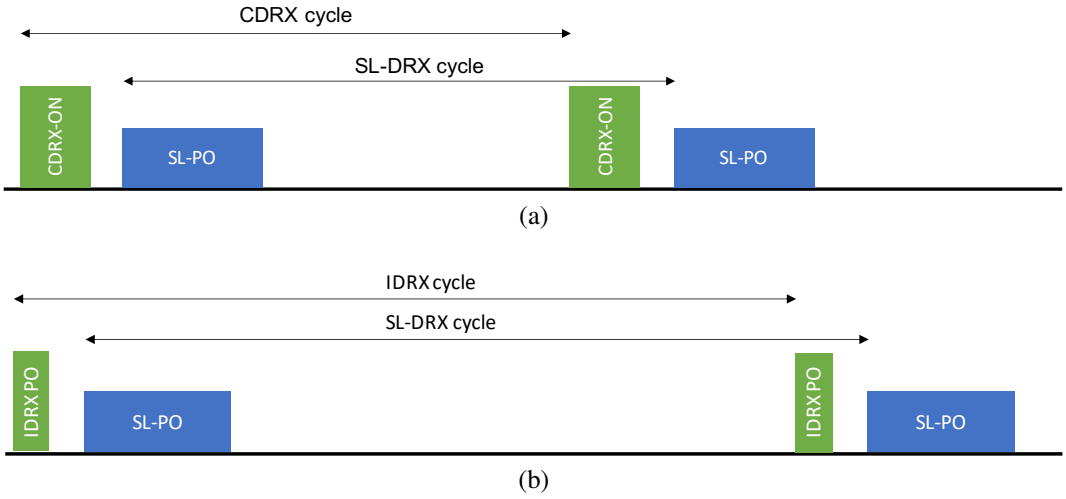


Figure 2.8: SL-DRX cycles when the UE is in (a) CDRX mode (b) IDRX mode.

location of the paging occasion inside an i_{PF} is calculated as

$$i_{\text{s}} = \left\lfloor \frac{\alpha_{\text{UE}}}{N_{\text{min}}} \right\rfloor \bmod \max\left(1, \frac{N_{\text{control}}}{T_{\text{IDRX}}}\right). \tag{2.20}$$

Finally, the SF index of the PO, i_{PO}, inside the paging frame is obtained from a lookup table (LUT) based on the values of i_{s} and N_{control} [99].

With this backdrop, we design the SL-PO for an SL paging cycle of $T_{\text{SL-DRX}}$ radio frames, such that it does not overlap with i_{PO}. We compute $i_{\text{SL-PF}}$, the SL paging frame which accommodates the SL-PO, as

$$i_{\text{SL-PF}} = i_{\text{SFN}} \in \mathcal{I}_{\text{SFN}} \text{ s.t.}$$
$$i_{\text{SFN}} \bmod T_{\text{SL-DRX}} = \begin{cases} N_{\text{ID}}, \text{for } T_{\text{SL-DRX}} \geq T_{\text{IDRX}} \\ N_{\text{ID}} \bmod T_{\text{SL-DRX}}, \text{otherwise.} \end{cases} \tag{2.21}$$

We then refer to the same LUT used for i_{PO} to obtain the SF index of the SL-PO, $i_{\text{SL-PO}}$, but offset it by n_{off} SFs to assure no collision between IDRX PO and SL-PO. This UE IMSI-

dependent SL-PO positioning randomizes the locations of the SL-PO and reduces the probability of inter-SL-PO overlaps. At the same time, an SRC can determine the SL-PO location of the DST using the a-priori DST IMSI information. We further propose a method to optionally modify the locations of SL-PO in Section 2.7.5 to avoid the rare occurrence of an overlap of the SL-PO with ConA UL/DL communication and/or CDRX-ON duration.

For the paging cycles to be consistently periodic across different hyper-frames, which consists of 1024 radio frames, (2.21) requires $T_{\text{SL-DRX}}$ to be a factor of the hyper-frame duration of 10.24 s, and thus SCUBA sets the allowed set of SL-DRX cycle values accordingly. Note that the same computation can be adapted when the P-RAT uses other forms of DRX, e.g., extended DRX (eDRX) as in LTE [99], which can then also support SL-DRX cycles higher than 10.24 s.

The SL-DRX cycles and SL-PO locations when the UE is in CDRX and IDRX are shown in Figure. 2.8(a) and Figure. 2.8(b), respectively. While IDRX PO is typically fixed to be a single SF, SCUBA provides a flexible SL-PO of $n_{\text{SL-PO}} \geq 1$ to facilitate quick retransmissions. The multiple SFs in an SL-PO may be either continuous or interleaved to support efficient transmission of SL messages of different lengths.

In general, we divide every SL-PO opportunity consisting of $n_{\text{SL-PO}}$ SFs into N_{cluster} separate clusters of SFs whose starting SFs are n_{dist} SFs apart, with $n_{\text{dist}} N_{\text{cluster}} \geq n_{\text{SL-PO}}$. Larger SL messages prefer higher values of $n_{\text{SL-PO}}$ and lower n_{dist} since multiple transport blocks can be transmitted with low latency, while smaller values of N_{cluster} are desirable for short messages since the UE can go back to sleep without requiring to complete paging on all $n_{\text{SL-PO}}$ SFs. At the same time, larger values of n_{dist} also provide greater time diversity but with a possible increase in network latency. SCUBA provides UE APIs with the flexibility to choose these parameters based on the intended use-case. By carefully choosing desired values of parameters $\{n_{\text{SL-PO}}, N_{\text{cluster}}, n_{\text{dist}}\}$, the resultant SL-PO may either be interleaved or consecutive.

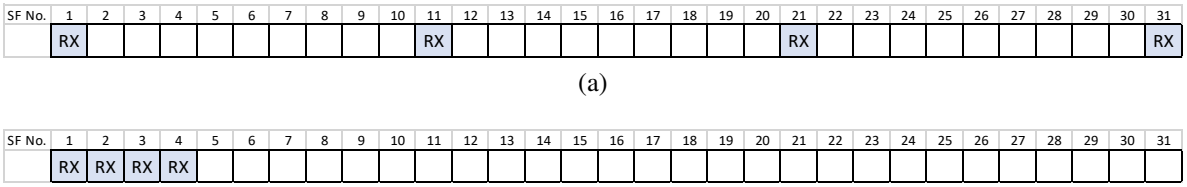


Figure 2.9: Timing diagram for (a) Interleaved and (b) Consecutive SL-PO.

We show the examples of interleaved and consecutive SL-POs in Figure. 2.9(a) and (b) with $\{n_{\text{SL-PO}}, N_{\text{cluster}}, n_{\text{dist}}\} = \{4, 4, 10\}, \{4, 1, \text{N/A}\}$, respectively, where the UE listens for paging on SFs marked as RX (receive).

2.7.2 SCUBA Server

As evident in Section 2.7.1, SCUBA requires an SRC to have a-priori knowledge of the $N_{\text{SL-DRX}}$ and α_{UE} of the DST to determine its $i_{\text{SL-PO}}$ to transmit on. One solution to this end is to embed these parameters within every UE and update it as necessary when new UEs join the network. Such a method is suitable for a largely static network with few participating UEs. However, this technique introduces significant signaling overhead when the number of network nodes increases and/or UEs are mobile. Therefore, we propose the use of a central SCUBA server with the database of all UE and network parameters that can be accessed and updated using the conventional cellular link. A SCUBA server could be the network BS or a dedicated SCUBA infrastructure node. Since updating and accessing these parameters by UEs are typically infrequent (α_{UE} is unique to every UE and hence is acquired only once per DST, and $N_{\text{SL-DRX}}$ is fairly static), occasionally communicating with the SCUBA server introduces negligible signaling overhead.

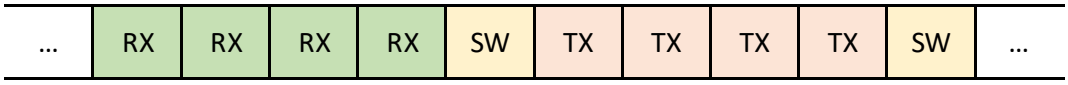


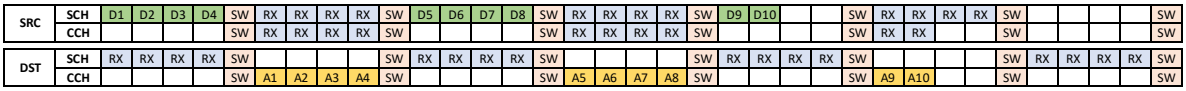
Figure 2.10: An example of SL SF pattern for 4 parallel HARQ processes

2.7.3 SL Connected Mode

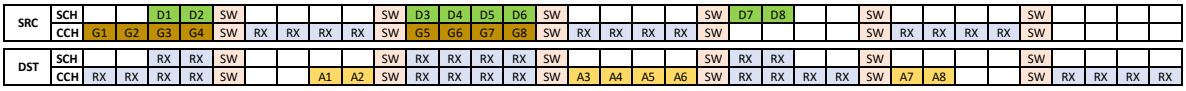
The latency associated with the SL communication is directly proportional to $N_{\text{SL-DRX}}$ chosen by the DST. But after a SCUBA message exchange is initiated, waiting for the next SL-PO to continue transmission introduces unnecessary delays. Therefore, SCUBA includes an SL connected mode, where the SRC and DST UEs engage in interactive communication until both their associated SL buffers are emptied. Note that this operation continues to be contingent on the UE being free from its P-RAT, as is the case with all SCUBA operations. SL connected mode includes N_{HARQ} continuous transmit (TX) SFs followed by a switching (SW) SF and additional N_{HARQ} RX SFs and so on, with as many cycles as required for data transfer to complete. An example of the RX-TX SF pattern with $N_{\text{HARQ}} = 4$ is shown in Figure. 2.10. Following completion, we allot an SL inactivity time, controlled by the SL inactivity timer (SL-INAT), during which both UEs monitor for a potential SL message. This period potentially provides further reduction in network latency, similar to the DRX inactivity time seen in LTE cellular link.

2.7.4 SL HARQ

To counter varying channel and noise conditions, we incorporate adaptive MCS and backward error correction in the form of HARQ in SCUBA. We support N_{HARQ} parallel synchronous HARQ processes in the SL connected mode to efficiently utilize the processing time of a transport block (TB) for reception of further TBs. Additionally, we also propose the use of the following two HARQ schemes in SCUBA which the UE API can choose from, along with a suitable value for N_{HARQ}.



(a)



(b)

Figure 2.11: Timing diagram for (a) Fixed MCS and (b) Grant-based HARQ schemes.

Fixed MCS/TBS

In this method, the SRC transmits SL data and control channels using a fixed MCS and TB size (TBS) chosen from a small set of pre-defined values. The grant-free nature of this scheme results in blind decoding at the DST, which adds to the processing complexity and time, but improves latency as signaling overhead introduced by grants is avoided. The pre-defined set of MCS and TBS values can either be pre-programmed on all UEs, or can be obtained from the SCUBA server. After successfully decoding every SL TB, the DST responds to the SRC individually with an ACK N_{HARQ} SFs after receiving the message. Therefore, the value of N_{HARQ} must also be chosen based on the computational ability of the UEs. The SRC retransmits any unacknowledged TB using the same pre-defined MCS and TBS, with the associated sequence number in the RLC header distinguishing a retransmission from a fresh transmission, similar to the use of RLC sequence number in LTE [100].

Grant-based HARQ

In this scheme, SRC begins transmission by sending a grant in the control channel for the upcoming data transmission, indicating its MCS and retransmission flag. The MCS is chosen adaptively using the channel quality indication (CQI) feedback provided by the DST, or with a conservative MCS for a cold transmission. Although this scheme may increase network

latency due to the signaling overhead introduced by transmitting a grant preceding every data SF, it provides the ability to adapt transmission to the varying channel and noise conditions to maximize the overall throughput. To ease grant decoding complexity at the DST, we impose the SL grant to be transmitted at least $N_{\text{SL-grant}}$ SFs before the SL data is transmitted.

For defining grant and data locations, we define an SL-HARQ frame length as, $N_{\text{frame}} = 2(N_{\text{HARQ}} + 1)$, which is the time taken to complete one cycle of TX-RX together with an SF for switching between TX and RX. Figure. 2.11 shows the timing diagrams for the fixed MCS scheme (Figure. 2.11(a)) with $N_{\text{HARQ}} = 4$, and the grant-based HARQ (Figure. 2.11(b)) with $N_{\text{SL-grant}} = 2$ and $N_{\text{HARQ}} = 4$. Similar to LTE, we use shared channel (SCH) to transmit the SL data, and control channel (CCH) to send grant and ACK. The notations Gi, Di, and Ai in the figures represent the ith grant, data, and ACK, respectively. For a grant transmitted on the ith SF in an SL-HARQ frame, the corresponding SL data is transmitted on the kth SF, where

$$k = \begin{cases} i + N_{\text{SL-grant}}, & \text{for } i < N_{\text{HARQ}} - 1 \\ i + 2N_{\text{HARQ}}, & \text{otherwise.} \end{cases} \tag{2.22}$$

2.7.5 Special Cases

In the following, we describe two modifications to native SCUBA to accommodate special needs of dedicated use-cases.

SCUBA-SAM

When a UE is relatively busy with the P-RAT characterized by long durations of ConA, several of its SL-PO opportunities are overlapped by the cellular link, which in turn may lead to a massive increase in the control plane latency for SCUBA. This situation can be exacerbated when the DST uses a large SL-DRX cycle. To counter this problem, in addition to the

existing SL-POs, we introduce dynamic SL-POs, which are positioned at the earliest available time opportunity when the UE is free from the cellular link. Such opportunities could either be the CDRX-OFF states, or the IDRX sleep duration when RAI is enabled. Since the locations of these dynamic POs are stochastic in nature and are unknown to other UEs, they need to be specifically advertised. We thus propose a short broadcast message called SL Availability Message for Dynamic SL-PO (SAM-D), as shown in Figure. 2.12, which notifies UEs of an upcoming dynamic SL-PO. Any SRC can then decode a SAM-D to identify the SL availability of a DST. For a rational SAM scheme that does not introduce extended SAM-listening periods, both the availability and unavailability statuses of a UE needs to be regularly communicated. Accordingly, we also use SAM to indicate the unavailability of a DST UE, e.g., during the ConA mode. To this end, we utilize the short time opportunities available during the ConA, e.g., radio switching SFs and/or other idle SFs, to transmit an Unavailability SAM (SAM-U), as shown in Figure. 2.13. We set the length of SAM to be $n_{\text{SAM-D}} = n_{\text{SAM-U}} \leq 0.5$ SF to ensure that it fits within the shortest available time-frame in the ConA mode (≈ 0.6 ms). Since continuous listening for a SAM introduces unreasonable power consumption, we define a SAM period of N_{SAM} SFs within which every UE must transmit a SAM, and whose value can be obtained from the SCUBA server. We choose $N_{\text{SAM}} >$ DRX-INAT to guarantee that at least one SAM can be sent by every UE within N_{SAM}. Note that since SAM transmission and listening results in increased power consumption, we recommend SCUBA-SAM to be used only by UEs with busy cellular traffic, where ConA operation often blocks the SL-POs.

SCUBA-LLM

UEs that are not limited by battery life, e.g., alternating current (AC) powered devices, and are used for latency critical applications can use a special option of our protocol, called

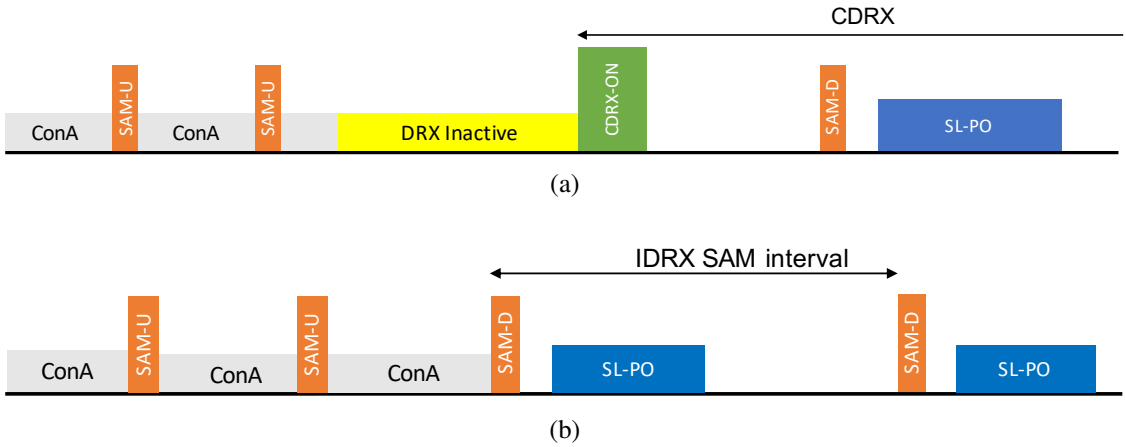


Figure 2.12: SAM-D transmissions and the associated dynamic SL-POs when the UE enters (a) CDRX mode, and (b) IDRX mode directly, when RAI is activated.

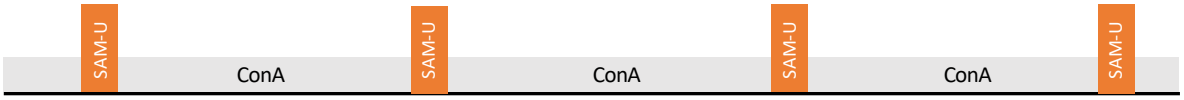


Figure 2.13: SAM-U transmissions when the UE is in ConA.

SCUBA low latency mode (SCUBA-LLM). This setting assigns $N_{\text{SL-DRX}} = 0$, and thus the UE listens for a potential SL message on all available time opportunities (see Table 2.6) without entering the sleep state. SCUBA-LLM subsumes SCUBA-SAM to enable SAM-U and SAM-D advertisements for control plane latency enhancement in busy devices, since the additional power required to transmit and receive SAMs is not a factor of consideration to these UEs that are not constrained by power consumption.

2.7.6 Data Transfer

We now specify the rules for SCUBA transmission considering the features introduced in Sections 2.7.1 to 2.7.5. Since SCUBA is an S-RAT, its data transmission is always contingent on both the UEs, SRC and DST, being free from cellular operation. However, SCUBA does not impose any further traffic shaping measures as we already showed previously that, with

Algorithm 1: SCUBA transmission procedure at the SRC.

```
1  Start: SL transmission request from the application layer
2  Compute i_SL-PO of DST using (2.21) and the LUT from [99];
3  while SRC free from cellular link do
4      Transmit grant/data on the next i_SL-PO of DST;
5      if ACK received then
6          Enter SL connected mode;
7          if SL-INAT expires then
8              break;
9      else
10         go to: Step 3;
```

the right choice of TBS, all types of MTC traffic specified by 3GPP [90] conforms well within the duty cycle limits introduced by regulatory authorities.

SCUBA Transmission

An SRC that receives transmission request from the application layer uses the values of $N_{\text{SL-DRX}}$ and α_{UE} (either obtained newly from the SCUBA server or that it knows a-priori) to compute $i_{\text{SL-PO}}$ of the DST. Based on the HARQ method used and the SL-PO type applied at the DST, the SRC initiates transmission on the DST SL-PO. Upon positive reaffirmation of a successful transmission by way of ACK reception, the SRC enters the SL connected mode with the DST, and exits after SL data transfer completion. If no ACK is received, the SRC retries at the next available DST SL-PO. This operation is summarized in Algorithm 1.

SCUBA-SAM Transmission

Upon receiving an SL transmission request from the application layer, an SRC UE begins listening for a SAM for N_{SAM} SFs as long as the SRC is not in the connected mode. If it receives a SAM-U from the DST indicating that the DST is in ConA mode, the SRC goes to sleep for a period of DRX-INAT, since that is the minimum period that the DST would remain in ConA.

If the SRC receives a SAM-D instead, it initiates SL communication with the DST on the dynamic SL-PO indicated in the SAM-D. On the other hand, if the SRC does not receive any SAM, it concludes that the DST is in the IDRX state. It then computes the SL-PO of the DST and attempts transmission on $i_{\text{SL-PO}}$. A successful SL initiation is indicated by an ACK transmitted by the DST. If the SRC does not receive an ACK, it retries transmission in the next $i_{\text{SL-PO}}$. A summary of the SCUBA-SAM transmission is also provided in Figure. 2.14.

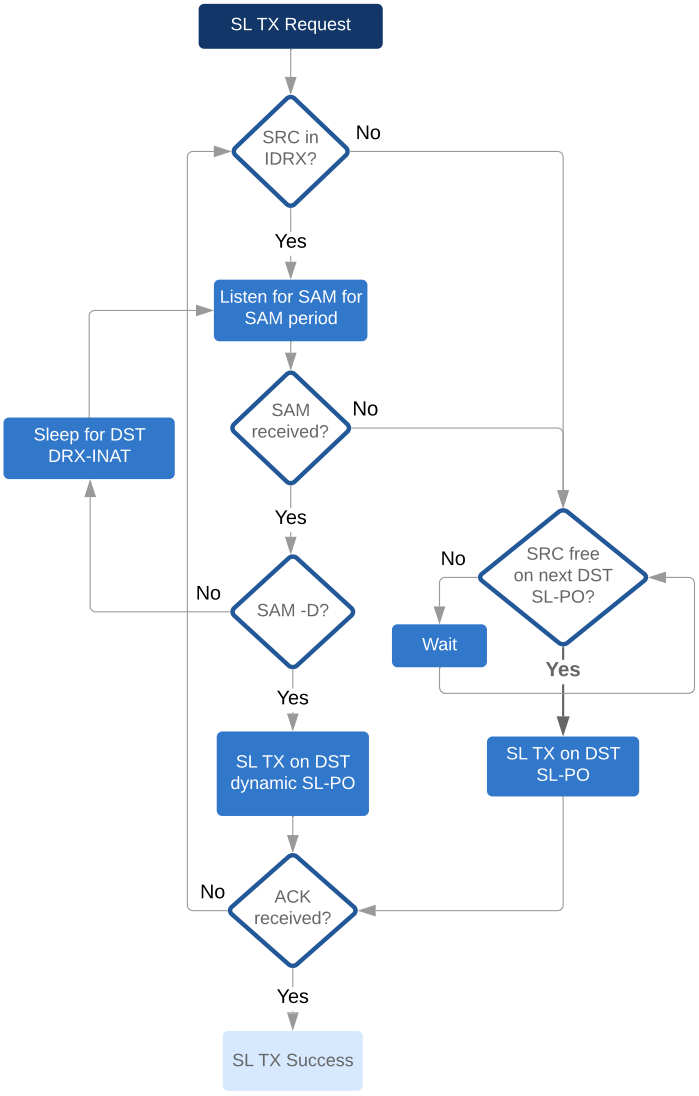


Figure 2.14: Flowchart for SCUBA-SAM transmission.

SCUBA-LLM Transmission

The transmission procedure in LLM follows the same rules of SCUBA-SAM transmission as shown in Figure. 2.14. When the DST also applies SCUBA-LLM, the DST SL-PO spans a large duration, which provides near-instant transmission opportunity to the SRC.

2.8 Performance Analysis

In this section, we analyze the performance of our proposed protocol in terms of the power consumed in each UE and the packet collision rate in a SCUBA network for varying traffic conditions. Throughout the analysis, we consider fixed MCS/TBS scheme as defined in Section 2.7.4 since conventional SL communication protocols like LTE-D2D are typically grant-free in nature. The analysis can be extended in a similar way for the grant-based HARQ scheme by making appropriate changes to include SL grant SFs.

2.8.1 Power Consumption

We analyze the power consumption for all three modes of operation. In general, the average overall power consumption can be computed as

$$P = \lambda_{\text{TXData}} E_{\text{TXData}} + \lambda_{\text{RXData}} E_{\text{RXData}} + \lambda_{\text{NoData}} E_{\text{NoData}}, \tag{2.23}$$

where λ_{TXData}, λ_{RXData}, and λ_{NoData} are the average event rates for UE having SL data to transmit, to receive, or neither, respectively, and $\lambda_{\text{TXData}} + \lambda_{\text{RXData}} + \lambda_{\text{NoData}} = 1$. E_{TXData}, E_{RXData}, and E_{NoData} are the total energy consumed by the UE during SCUBA transmission, reception, and idle states, respectively. We begin with computing the energy consumption for completing the SCUBA transmission and reception, $E_{\text{SL-TX}}$ and $E_{\text{SL-RX}}$, respectively, for

variable packet sizes as

$$E_{\text{SL-TX}} = t_{\text{SF}}\Bigg(\underbrace{P_{\text{TX}} n_{\text{SL}}}_{\text{TX SL-U data}} + \underbrace{P_{\text{RX}} n_{\text{SL}}}_{\text{RX ACK}} + \underbrace{P_{\text{switch}} \left\lceil \frac{n_{\text{SL}}}{N_{\text{HARQ}}} + 1 \right\rceil}_{\text{Switch b/w TX and RX}} + \underbrace{P_{\text{RX}} n_{\text{SL-INAT}}}_{\text{SL inactivity}} \Bigg),$$

$$E_{\text{SL-RX}} = t_{\text{SF}}\Bigg(\underbrace{P_{\text{RX}}(N_{\text{HARQ}} + \max(0, n_{\text{SL}} - N_{\text{HARQ}}))}_{\text{RX SL data}} + \underbrace{P_{\text{TX}} n_{\text{SL}}}_{\text{TX ACK}} + \underbrace{P_{\text{switch}} \left\lceil \frac{n_{\text{SL}}}{N_{\text{HARQ}}} + 1 \right\rceil}_{\text{Switch b/w TX and RX}} + \underbrace{P_{\text{RX}} n_{\text{SL-INAT}}}_{\text{SL inactivity}} \Bigg), \tag{2.24}$$

where P_{TX}, P_{RX}, and P_{switch} represent the UE power consumption for any data TX, RX, and switching between TX and RX operations, respectively, n_{SL} and $n_{\text{SL-INAT}}$ are the number of SFs consumed by SL data and SL inactivity time, and t_{SF} is the time duration of a SCUBA SF. Next, we compute E_{TXData}, E_{RXData}, and E_{NoData} for each of three modes of operation.

Native SCUBA

Under the native SCUBA operation, UEs only TX, RX, and listen for SL-POs during P-RAT inactivity. Therefore,

$$E_{\text{TXData}} = (1 - p(\text{ConA}))E_{\text{SL-TX}}, \tag{2.25}$$

$$E_{\text{RXData}} = (1 - p(\text{ConA}))E_{\text{SL-RX}}, \tag{2.26}$$

$$E_{\text{NoData}} = (1 - p(\text{ConA}))\frac{P_{\text{RX}} n_{\text{SL-PO}} t_{\text{SF}}}{N_{\text{SL-DRX}}}, \tag{2.27}$$

where $p(\text{ConA})$, $p(\text{CDRX})$, and $p(\text{IDRX})$ represent the probabilities of the UE to be in ConA, CDRX, and IDRX modes, respectively, in the P-RAT.

SCUBA-SAM

When UEs use the SAM mode, they additionally also transmit SAM in the connected modes to indicate their availability status. Since the use of SAM-D is limited in IDRX mode, its probability is negligible and hence not considered in the analysis. Thus,

$$\begin{aligned} E_{\text{TXData}} = p(\text{ConA})\frac{P_{\text{TX}} n_{\text{SAM-U}} t_{\text{SF}}}{N_{\text{SAM-U}}} + (1 - p(\text{ConA}))\Big(E_{\text{SL-TX}}\Big) \\ + p(\text{IDRX})\Bigg(P_{\text{RX}} t_{\text{SF}}\Big(p(\text{ConA})\frac{k_{\text{SAM-U}} N_{\text{SAM-U}} + N_{\text{SAM-D}}}{2} \\ + p(\text{CDRX})\frac{N_{\text{SAM-D}}}{2} + p(\text{IDRX}) N_{\text{SAM}}\Big)\Bigg), \end{aligned} \tag{2.28}$$

$$E_{\text{RXData}} = p(\text{ConA})\frac{P_{\text{TX}} n_{\text{SAM-U}} t_{\text{SF}}}{N_{\text{SAM-U}}} + (1 - p(\text{ConA}))E_{\text{SL-RX}} + p(\text{CDRX})\frac{P_{\text{TX}} n_{\text{SAM-D}} t_{\text{SF}}}{N_{\text{SAM-D}}}, \tag{2.29}$$

$$\begin{aligned} E_{\text{NoData}} = p(\text{ConA})\frac{P_{\text{TX}} n_{\text{SAM-U}} t_{\text{SF}}}{N_{\text{SAM-U}}} + p(\text{CDRX})\left(\frac{P_{\text{TX}} n_{\text{SAM-D}} t_{\text{SF}}}{N_{\text{SAM-D}}} + \frac{P_{\text{RX}} n_{\text{SL-PO}} t_{\text{SF}}}{N_{\text{SL-DRX}}}\right) \\ + p(\text{IDRX})\frac{P_{\text{RX}} n_{\text{SL-PO}} t_{\text{SF}}}{N_{\text{SL-DRX}}}, \end{aligned} \tag{2.30}$$

where $k_{\text{SAM-U}}$ is the average number of SAM-Us received by the SRC UE before receiving a SAM-D. $N_{\text{SAM-U}}$ and $N_{\text{SAM-D}}$ are the transmission intervals of SAM-U and SAM-D in SFs units, respectively, with preferably $N_{\text{SAM-D}}, N_{\text{SAM-U}} \leq \frac{N_{\text{SAM}}}{2}$ which guarantees SAM-D and SAM-U to be received by the SRC listening for N_{SAM} SFs.

SCUBA-LLM

The energy consumption in LLM is similar to that of SCUBA-SAM with the exception that the SL-POs extend throughout all available times in CDRX and IDRX modes, which impacts the power consumption when the UE is idle. Therefore, the transmission and reception

energies are identical to (2.28) and (2.29), respectively, and

$$E_{\text{NoData}} = p(\text{ConA})\frac{P_{\text{TX}} n_{\text{SAM-U}} t_{\text{SF}}}{N_{\text{SAM-U}}} + p(\text{CDRX})\left(\frac{P_{\text{TX}} n_{\text{SAM-D}} t_{\text{SF}}}{N_{\text{SAM-D}}} + \frac{P_{\text{RX}} n_{\text{CDRX,SL-PO,LLM}} t_{\text{SF}}}{N_{\text{CDRX}}}\right) + p(\text{IDRX})\frac{P_{\text{RX}} n_{\text{IDRX,SL-PO,LLM}} t_{\text{SF}}}{N_{\text{IDRX}}}, \tag{2.31}$$

where N_{CDRX} and N_{IDRX} are the CDRX and IDRX cycle values in SFs of the P-RAT in the UE, and $n_{\text{CDRX,SL-PO,LLM}}$ and $n_{\text{IDRX,SL-PO,LLM}}$ are the free time duration in SFs in CDRX and IDRX modes, respectively, which are all used for paging SL messages.

2.8.2 Packet Collisions

SCUBA inherently includes an elementary collision control strategy by incorporating the SL-POs to be dependent on α_{UE}. Nevertheless, packets can still collide when two or more SRC UEs transmit data simultaneously to the same UE or to different UEs sharing overlapping SL-POs and SL bands. To compute the probability of SL data collision, we define the following events:

- A: two or more SRC UEs have data to transmit at an SL-PO
- B: two or more SRC UEs transmit at the same SL-PO
- C: SRC and DST UEs are not in ConA.

The probability of collision can then be expressed as

$$p_{\text{c}} = \frac{1}{N_{\text{B}}} p(A \cap B \cap C), \tag{2.32}$$

where N_{B} is the number of orthogonal unlicensed frequency bands that can be used, depending on the total available bandwidth and number of PRBs applied for SCUBA transmission. To evaluate the worst-case collision scenario, we set $p(C) = 1$ to emulate conditions where

UEs are mostly free from cellular communication and available for SL-U[4]. Therefore, we can express

$$p_{\mathrm{c}} = \frac{1}{N_{\mathrm{B}}} p(A) p(B|A), \tag{2.33}$$

with

$$p(A) = \sum_{k=2}^{N_{\mathrm{UE}}} \binom{N_{\mathrm{UE}}}{k} p_{\text{SL-TX}}^{k} (1 - p_{\text{SL-TX}})^{N_{\mathrm{UE}}-k}, \tag{2.34}$$

where N_{UE} is the total number of UEs in the SCUBA network and $p_{\text{SL-TX}}$ is the probability of a UE having a non-empty SL buffer. The latter can be computed as

$$p_{\text{SL-TX}} = \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\gamma N_{\text{SL-DRX}} t_{\text{SF}}} (\gamma N_{\text{SL-DRX}} t_{\text{SF}})^{n}}{n!}, \tag{2.35}$$

where $\frac{1}{\gamma}$ is the mean inter-arrival time of the Poisson packet arrivals in SCUBA. Furthermore,

$$p(B|A) = \sum_{k=2}^{N_{\mathrm{UE}}} \left(\frac{N_{\text{SL-PO}}}{N_{\text{SL-DRX}}} \right)^{k} \tag{2.36}$$

under the assumption that an SRC UE transmits on an SL-PO with uniform random probability and that SL-POs are allotted in slots with the same SL-DRX cycles used by all UEs. The above analysis considers the case of consecutive SF SL-POs as shown in Figure. 2.9(b), but similar analysis also follows for the case where SL-POs are interleaved. Note that for applications where UEs frequently communicate with a central receiver, e.g., a data aggregator, we have $p(B|A) = 1$ and therefore, $p_{\mathrm{c}} = \frac{1}{N_{\mathrm{B}}} p(A)$. We show results for p_{c} in Section 2.9.3 across varying number of network participants that will guide the choice of SL-DRX cycle based on the SL use-case and the traffic being served.

Similar to SCUBA data collisions, the SAMs transmitted as part of the UE state discovery procedure may also collide leading to a false positive for UE availability. Determining the

[4]We have shown previously that Event C is overwhelmingly the most probable scenario encountered with typical MTC traffic.

SAM collision probability across various network conditions contributes to the choice of SAM transmission intervals along with the acceptable latency and power consumption trade-off. The SAM collision probability can be expressed similar to (2.34) as

$$p_{\text{c,SAM}} = \frac{1}{N_{\text{B}}} \sum_{k=2}^{N_{\text{UE}}} \binom{N_{\text{UE}}}{k} p_{\text{SAM}}^{k} (1 - p_{\text{SAM}})^{N_{\text{UE}}-k}, \tag{2.37}$$

where p_{SAM} is the SAM transmission probability computed as

$$p_{\text{SAM}} = p_{\text{ConA}} \frac{n_{\text{SAM-U}}}{N_{\text{SAM-U}}} + p_{\text{CDRX}} \frac{n_{\text{SAM-D}}}{N_{\text{SAM-D}}}. \tag{2.38}$$

The SAM collision rates along with the other performance indicators of SCUBA are presented in the following section.

2.9 Numerical Results

We evaluate the performance of SCUBA by simulating its MAC layer timing behavior in MATLAB. Along with power consumption and collision rates, we also use network latency as a performance indicator for evaluation. We define network latency as the time taken to complete an acknowledged transmission of 100 bytes of SL data, which is the typical data size considered for triggered as well as periodic reports in MTC smart metering applications [90, 101]. We also include a comparison of the power consumption of native SCUBA against that of other D2D solutions for different sleep cycles.

2.9.1 Simulation Settings

For the underlying P-RAT, we simulate both Poisson and periodic cellular traffic models as suggested in TR 36.888 [90], whereas the SCUBA data arrivals are modeled as a busy Pois-

son traffic in all our simulations to evaluate the worst-case performance. To better investigate the performance of SCUBA in special scenarios, e.g., for SCUBA-SAM, we consider additional test-cases with lengthy ConA duration. Thus we fundamentally classify the simulation cases into short-data ($t_{\text{data}} = 250$ ms) and long-data ($t_{\text{data}} = 5$ s) cellular traffic conditions. The long t_{data} in cellular traffic is used to evaluate the performance gain of SCUBA-SAM, i.e., the latency gains obtained using SAM-D. The simulation settings applicable to both the cases are given in Table 2.7 whereas the case-specific simulation settings are defined in Table 2.8. For the P-RAT, we choose the values of network parameters such as RRC connection set-up time (t_{RRC}) and DRX-INAT conforming with the set of values allowed by the 3GPP specification [81]. Based on ERP regulations in Table 2.4, and considering 45% power amplifier efficiency and 60 mW power consumption in the support circuitry [102], we set the UE power consumption for SCUBA transmission as 100 mW.

The choice of MCS and PRB size impacts the SCUBA network latency. To evaluate the worst case performance, we consider the lowest possible MCS and the fewest PRBs, which result in the slowest possible transmission. However, the choice of MCS and the number of PRBs is further restricted by the following regulations governing the use of unlicensed bands. The North American regulations impose that the 6 dB bandwidth of all digitally modulated systems must be greater than 500 kHz without any restriction on the maximum usable bandwidth [86]. With LTE-M as the underlying primary RAT whose parameters and specifications are also reused for SCUBA, a minimum of three PRBs are required to ensure that the 6 dB bandwidth is greater than 500 kHz. For the choice of MCS, we consider the duty cycle limitations imposed by European unlicensed band usage regulations. Our prior analysis has established that the minimum MCS that satisfies the duty cycle regulations is MCS ≥ 3. However, this analysis considers that the transmissions are error-free. For practical communication conditions with non-negligible block errors, erroneous transport blocks

are typically retransmitted. The number of transport blocks retransmitted is dependent on the operating signal-to-noise ratio conditions, the chosen target block error rate, and the retransmission scheme used, e.g., whether Layer-1 HARQ is used or RLC retransmissions are used. To accommodate for several possible retransmissions, we choose a retransmission-agnostic conservative choice of MCS $= 6$ for our evaluations. With the chosen MCS and number of PRBs, we extract the TBS from [93, Table 7.1.7.2.1-1, Table 8.6.1-1]. Consequently, for the analysis and simulations, we choose an MCS of 6, TBS of 256 bits corresponding to a PRB size of 3, and cross-layer overhead of 19 bytes, for the SCUBA traffic to comply with duty-cycle limits. The selected MCS value corresponds to data modulated with quadriphase shift keying (QPSK) and coded with a rate of approximately 0.3.

For power and latency simulations, we consider only a single pair of SRC and DST UEs and hence the results do not reflect the impact of collisions resulting from different UEs transmitting to the same DST. We show in Section 2.9.3 that this is a valid assumption for most MTC system architectures. We focus on the unlicensed bands of $865 - 868$ MHz and $902 - 928$ MHz for Europe and United States, respectively, and therefore, a bandwidth of at least 3 MHz is available for SCUBA. Since we choose LTE-M as the P-RAT for simulations, the signals are limited to 6 PRBs (1.08 MHz). Thus the considered band allows us to have a minimum of 2 non-overlapping signal bandwidths. Hence for the collision evaluation, we further set $N_{\mathrm{B}} = 2$ to investigate the worst-case. When fewer PRBs are used and/or larger bandwidth is available for SCUBA resulting in a higher N_{B}, the collision rates are scaled down accordingly.

Table 2.7: Common simulation settings for SCUBA evaluation [56, 102, 103].

Parameter	Value	Parameter	Value
$\gamma^{-1} = \lambda_{\text{TXData}}^{-1} = \lambda_{\text{RXData}}^{-1}$	30 s	P_{TX}	100 mW
P_{switch}	80 mW	P_{RX}	80 mW
$n_{\text{SL-INAT}}$	0	n_{SL}	8
N_{SAM}	150 ms	n_{SAM}	0.5 ms
$N_{\text{SAM-D}}$	75 ms	$n_{\text{SL-PO}}$	4
$N_{\text{SAM-U}}$	20 ms	N_{B}	2
Mean IAT (Poisson, P-RAT)	30 s	IAT (periodic, P-RAT)	5 min
CDRX-ON duration	20 ms	CDRX cycle	640 ms
t_{RRC}	100 ms	IDRX cycle	640 ms
DRX-INAT	100 ms	RAI enabled	No

Table 2.8: Case-specific simulation settings for SCUBA evaluation.

Case	Short t_{data}	Long t_{data}
t_{data}	250 ms	5 s
Data-INAT	10 s	5 s

2.9.2 Simulation Results: Power and Latency

Short-Data Cellular Traffic

We present the SCUBA power consumption results in Figure. 2.15 (left). Since the UE listens on SL-POs more frequently at lower SL-DRX cycle values, the power consumption significantly reduces with increase in SL-DRX cycle. The power consumed for SL data transmission and reception remains the same regardless of the SL-DRX cycle, and therefore has little impact on the power variation. For both Poisson and periodic traffic in P-RAT, native SCUBA has similar power consumption, since the underlying cellular traffic plays little role, except in blocking rare SCUBA traffic on occasion, resulting in retransmission. The effect of such transmission failures and resultant retransmissions is negligible compared to the overall power consumption due to SL-PO listening.

We show the latency incurred by the SCUBA traffic in Figure. 2.15 (right), where L_{99} and

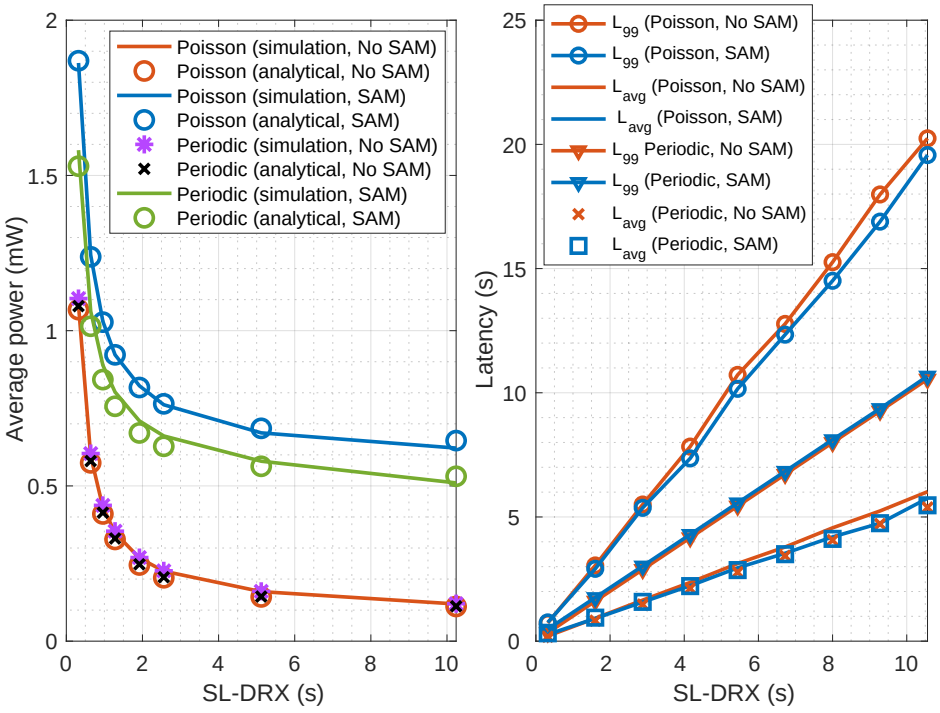


Figure 2.15: Average power consumption vs SL-DRX (left), and latency vs SL-DRX (right), for short data duration.

L_{avg} correspond to 99th percentile and average latencies respectively. We observe that the latency increases nearly linearly with the SL-DRX cycle duration. Along with providing no improvement in latency, using SAM results in higher power consumption due to SAM transmission and listening. In the presence of SAM, the power consumption in Poisson traffic model is high compared to the periodic case because of the higher number of SAM transmissions due to dominant ConA and CDRX modes. This clearly indicates that the use of SCUBA-SAM is not suitable when UEs operate with short data lengths in the P-RAT.

Long-Data Cellular Traffic

The results of power consumption and network latency for long data traffic in the P-RAT are shown in Figure. 2.16. Under the absence of SAM, SCUBA power remains the same as in the short-data case since there are no additional factors which contribute to the power

Table 2.9: Latency simulation results for LLM.

Cellular traffic	Short data		Long data	
	99th percentile	Average	99th percentile	Average
Poisson	340.8 ms	28.2 ms	14.11 s	1.392 s
Periodic	38.0 ms	20.6 ms	3.717 s	108.9 ms

Table 2.10: Average power consumption results for LLM.

Cellular traffic	Short data		Long data	
	Simulation	Analysis	Simulation	Analysis
Poisson	78.2 mW	78.6 mW	67 mW	67 mW
Periodic	79.8 mW	80.1 mW	78.6 mW	78.9 mW

consumption. However, SAM in long-data case results in much higher power consumption than in short-data case due to the higher percentage of ConA and CDRX modes causing more number of SAM transmissions. This increased power consumption due to SAM is also associated with an improvement in the achieved latency. The achieved latency gain increases with SL-DRX cycle duration, and is higher for SCUBA multiplexed with Poisson cellular traffic as compared to periodic cellular transmission. The result shows that a gain of more than 23% is observed in 99th percentile latency with the use of SAM when the SL-DRX cycle is 10.56 s. The results clearly show that SCUBA-SAM is particularly useful when the UEs are busy with cellular traffic causing ConA and CDRX modes to occupy a majority of the time. In busy cellular traffic, the ConA and CDRX-ON modes overlap with most of the fixed SL-POs, which introduces higher latency. This latency can be reduced with the use of SAM by assisting the SRC to transmit on an earlier dynamic SL-PO without waiting for the next fixed SL-PO.

To avoid repetition, we do not show the simulation results for the case where SAM is transmitted in cellular IDRX mode when RAI is enabled. We observed similar latency gains for SAM mode in the long-data case when RAI is enabled because of SAM-D in IDRX allowing for an early SL transmission.

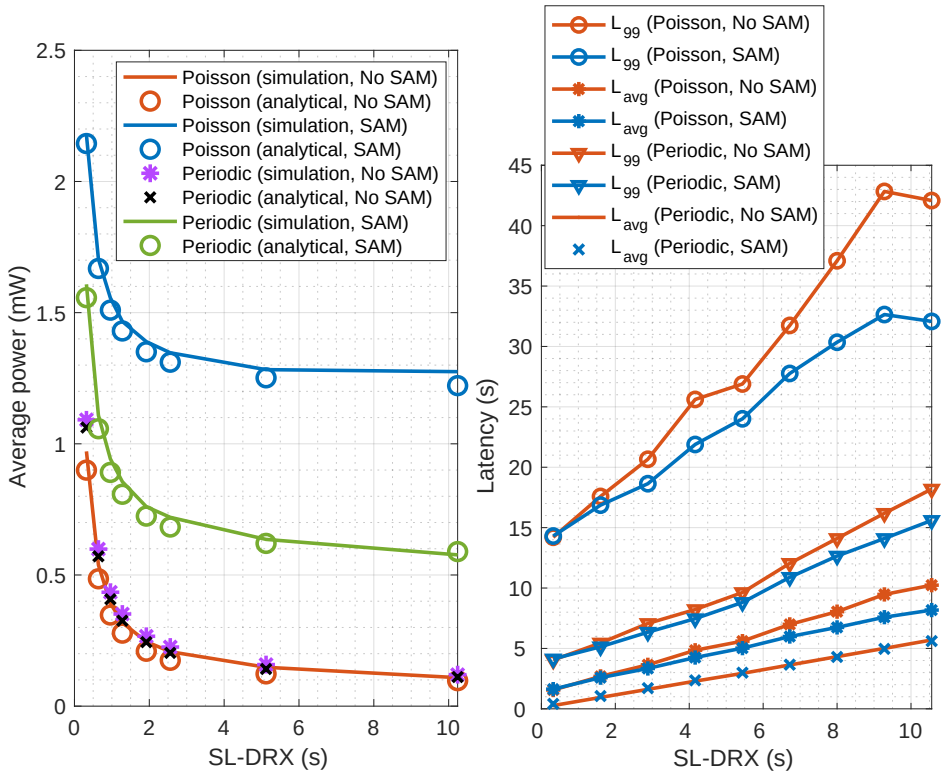


Figure 2.16: Average power consumption vs SL-DRX (left), and latency vs SL-DRX (right), for long data duration.

Since the SL-DRX cycle is 0 for SCUBA-LLM, we tabulate the latency and power values in Table 2.9 and Table 2.10 separately, respectively. LLM latency is significantly smaller for SCUBA multiplexed with the longer IAT periodic data transmission than a higher rate of arrival Poisson cellular traffic, because the latter hinders SL listening more with frequent data arrivals. We observe that the SCUBA-LLM power consumption is higher in case of short cellular data than in the longer data traffic case. The power difference is also significant for Poisson cellular traffic, whereas it is marginal with periodic arrivals because the former allows little SL listening resulting from more frequent long data arrivals in the P-RAT. Overall, the results show that SCUBA-LLM allows nearly instant SL communication at the expense of high power consumption compared to the version of SCUBA with non-zero SL-DRX cycles.

Comparison with Other D2D Solutions

To put the performance of SCUBA in perspective, we consider the power consumption values of other D2D technologies reported in the literature [104]. For comparison with SCUBA, we focus on Bluetooth low energy (BLE) and ZigBee, since they consume lower power than the other state-of-the-art commercialized D2D solutions [104]. Note that the transmitter power we have chosen for SCUBA is the maximum transmit power allowed in the $865 - 868$ MHz band. The power consumption in SCUBA will however be different when transmitting in a different regulated unlicensed band.

Prior to comparing SCUBA with other D2D solutions, we emphasize two major differences in the nature of the protocols. First, transmit power of signals in BLE and ZigBee are regulated to be fixed between $1 - 100$ mW [44], which severely limits the achievable coverage when compared to that of SCUBA. Furthermore, unlike SCUBA, BLE and ZigBee are incompatible to function as a TDM solution in low-cost LTE-M HD-FDD UEs using a single radio architecture. Nevertheless, the power consumption of BLE and ZigBee devices are shown in Figure. 2.17. We present the average power consumption comparison between BLE, ZigBee, and native SCUBA over different sleep cycles. The power consumption values of BLE and ZigBee devices are from [104], and correspond to periodic 8 byte data transmissions at 1 mW between sleep cycles of t_{sleep} duration. On the other hand, the SCUBA power consumption results correspond to an LTE-M device transmitting an 8 byte data at a transmit power of 25 mW. The power consumption in SCUBA is evidently lower than BLE and ZigBee in spite of latter solutions having lower transmit power. Since SCUBA operates as an S-RAT utilizing the synchronization acquired through P-RAT, and transmits on predefined SL-POs, it does not require power-expensive advertising procedures as in BLE and Zigbee [104]. Considering the higher transmission power, SCUBA also has the potential to provide much wider coverage than BLE and ZigBee .

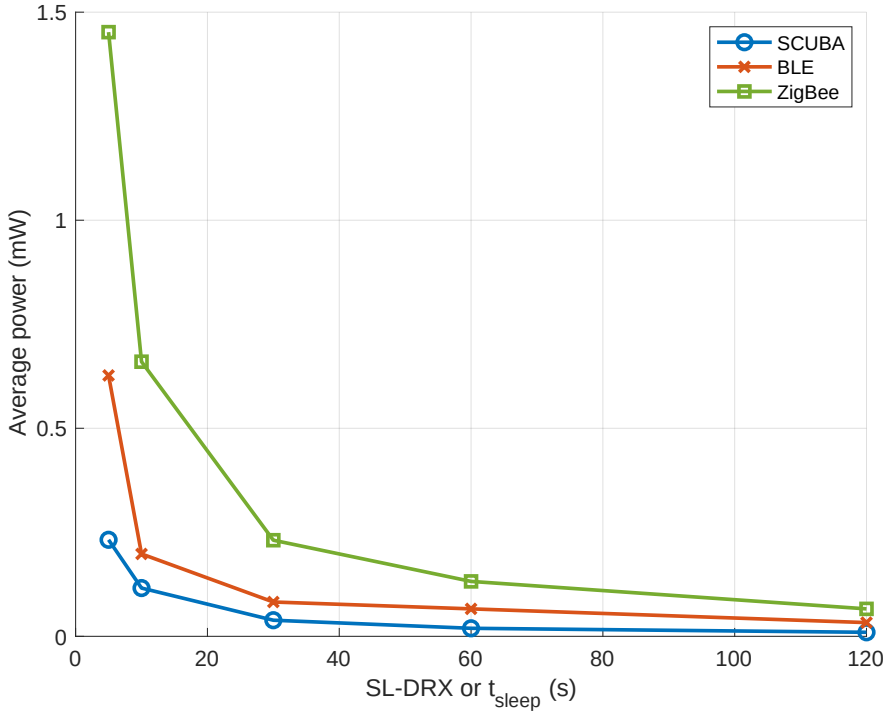


Figure 2.17: Average power consumption vs SL-DRX or t_{sleep} for native SCUBA, BLE, and ZigBee [104].

Additionally, SCUBA also facilitates LLM for latency-critical applications, where it provides the earliest possible transmission opportunity while coexisting with a P-RAT. For a transmission of 376 bits, SCUBA LLM average latency for cellular periodic short-data case is 9.97 ms against the combined discovery-connection mode average latency of 11 ms in BLE [105, 106]. Furthermore, unlike BLE which operates only in 2.4 GHz band, SCUBA can function in any unlicensed frequencies of choice, by following the regulatory requirements set for each band. In summary, different from other existing D2D standards, SCUBA can coexist with a cellular RAT in a TDM manner with considerably lower power consumption, improved coverage expansion, and comparable latency performance.

Battery life

Since we have chosen LTE-M as the P-RAT for our SCUBA simulations and analysis, we use the energy consumption analysis of MTC devices provided in the 3GPP technical report [91] to estimate the impact of SCUBA on the battery life of a UE. We consider a UE having an ideal battery source of capacity $E_{\text{b}} = 5$ Wh with no power leakage [91]. The energy consumption per day by SCUBA and LTE-M operating together can be computed as

$$E_{\text{day}} = E_{\text{SCUBA}} + E_{\text{LTE-M}} \text{ Wh}, \tag{2.39}$$

where E_{SCUBA}, $E_{\text{LTE-M}}$ are per-day energy consumption by SCUBA, LTE-M respectively. Then, the UE battery life can be computed as, $D = \frac{E_{\text{b}}}{E_{\text{day}}}$ days.

For SCUBA, first we consider Poisson traffic with the same values of parameters as in Table 2.7. Since LLM and SAM are intended for delay sensitive applications tailored for UEs mostly connected with alternating current (AC) power source or devices without battery-life constraints, we analyze the battery life impact of SCUBA in its native mode only. For LTE-M, we consider the battery-life analysis as given in the literature [56, 107]. For LTE-M standalone operation, if a device with battery capacity of 5 Wh transmits packets worth 200 bytes in a cellular link at 164 dB maximum coupling loss (MCL) at a frequency of 2 hours, the battery lasts for 328.5 days [56]. SCUBA power consumption values are independent of the underlying cellular traffic. Based on the average SCUBA power consumption value for SL-DRX cycle of 10.24 s obtained from simulation results (Figure. 2.15), we estimate the UE battery life for LTE-M-SCUBA coexisting scenario. We find that the battery life reduces to 279 days when SCUBA coexists together with LTE-M. Note that these numbers are obtained for a fairly busy SCUBA traffic of Poisson packet arrivals with mean IAT of 30 s, and are therefore indicative of the worst-case scenario.

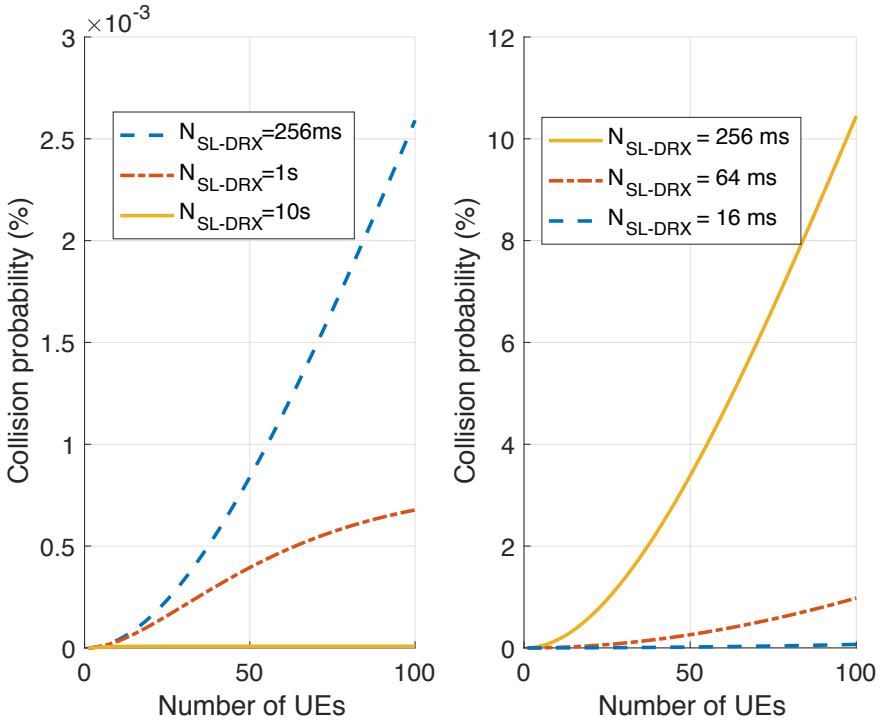


Figure 2.18: SCUBA data collision probability when UEs communicate with each other randomly (left), and when all SRC UEs always report to a central DST (right).

For a more realistic practical indication, we also investigate the battery life when SCUBA traffic is as infrequent as in LTE-M, i.e., IAT of 2 hours. The battery life is found to be 328.3 days for SCUBA-LTE-M coexistence as compared to 328.5 days in the absence of SCUBA. This shows the effectiveness of the low-power design of SCUBA, where it consumes power only for SL transmissions and receptions, and reuses the synchronization achieved in the cellular link. This energy consumption performance highlights the appeal of SCUBA to be ideally suited as an integrated low-cost and low-power solution for SL-U in low-complexity devices, particularly LTE-M Cat-M1 UEs used for MTC and IoT applications.

2.9.3 Simulation Results: Collision

For power and latency simulations discussed in the previous sections, we did not consider the collisions resulting from different UEs transmitting at the same SL-PO of a DST. If collisions occur in SCUBA transmission, the power consumption and latency will increase accordingly. In this section, we present the collision results obtained using the analysis in Section 2.8.2. We first show the results of SCUBA data collisions in Figure. 2.18. As expected, we observe an increase in the rate of collision with the number of participating UEs. For the case of devices communicating with each other at random, seen in Figure. 2.18 (left), we notice a higher probability of collision with lower SL-DRX cycle as it increases the possibility of SL-PO overlap among DST UEs. The absolute value of collision rates under all these conditions are in the order of 10^{-5}, indicating a nearly collision-free SCUBA communication across different operating conditions. However, when all source UEs are reporting to a common central DST, the collision rates increase with higher SL-DRX cycle, as seen in Figure. 2.18 (right), since a longer interval between SL-POs allows for more SRC UEs to have SL data to transmit. Therefore, DST UEs that are expected to receive data from several/all network nodes should use a lower SL-DRX cycle to avoid SCUBA packet collisions. Note that these collision results are independent of short- or long-data in the P-RAT, since we evaluate the worst-case collision with $p(C) = 1$. Given that the SCUBA packet collision rates are negligible for both categories of network architectures, we do not compare these values against the probabilities of collision obtained with other D2D solutions, since it does not provide any further meaningful insights.

Similarly, we present the results of SAM collisions for the case of SCUBA-SAM in Figure. 2.19. We observe that type of traffic in the P-RAT significantly impacts the collision rates, with long-data at the cellular link resulting in higher SAM collisions and short-data causing fewer collisions. This behavior is a result of UE spending a higher amount of time

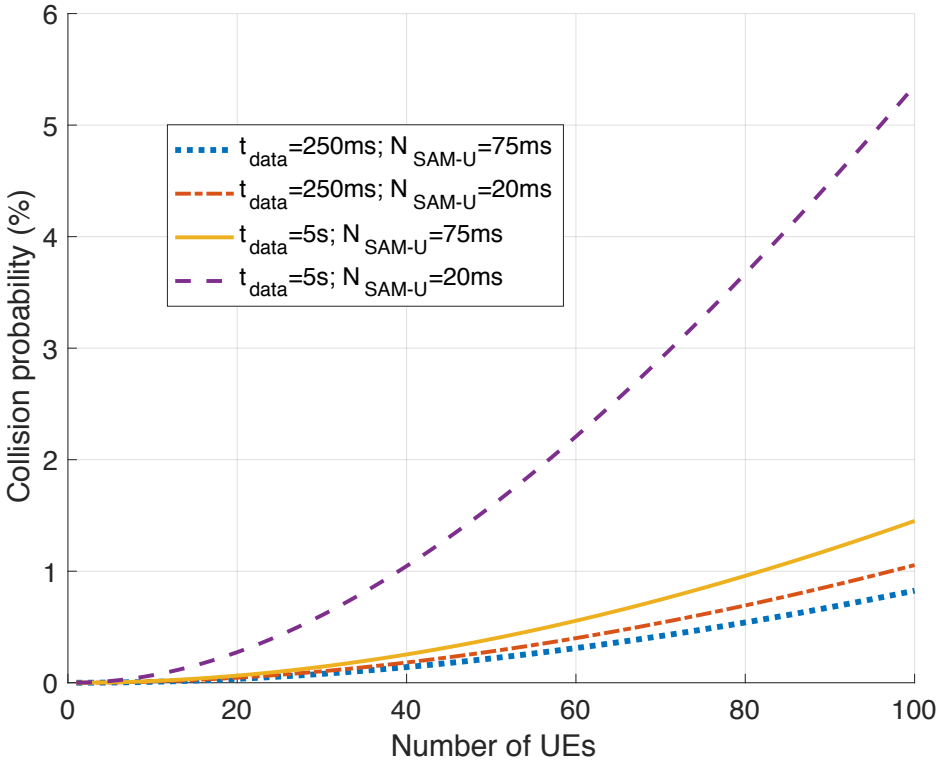


Figure 2.19: SAM collision probability for SCUBA-SAM mode.

in ConA mode, during which it regularly transmits SAM-U. These results indicate that the minor gains in network latency (less than 50 ms) achievable with use a SAM transmission interval of 20 SFs incur a high cost of collision of up to a $3\times$ increase in collision rates, especially for the case of long-data. This suggests that using a higher SAM transmission interval of, e.g., $N_{\text{SAM-U}} = 75$ ms is suitable. It also ensures that at least 2 SAMs are transmitted within a SAM period of 150 SFs, which nearly guarantees that a listening UE does not miss a SAM due to collision.

2.10 Discussion

In this section, we reflect on SCUBA by discussing its salient features and identifying potential future work required for a practical realization.

2.10.1 SCUBA Customization

SCUBA is a customizable protocol that can be further adapted based on the application scenario. For example, native SCUBA in its current form works on UEs that are in one of the DRX modes for cellular communication. However, if their ConA transmission is deterministic in nature, i.e., if the SRC is aware a-priori of idle times in ConA (e.g., switch SF) of the DST, SCUBA can also further be modified to transmit SL data to a UE in ConA mode, and subsequently expect an ACK back in the next idle time of the DST.

2.10.2 Other Layer Specifications

This work focuses on the MAC layer specifications of SCUBA. Its PHY characteristics are largely driven by the regulations governing the access of the unlicensed spectrum used. For example, we listed the PHY specifications for the use of $865 - 868$ MHz band in Europe and $902 - 928$ MHz band in the United States. For upper layer specifications, SCUBA borrows them from the underlying P-RAT used in the UE. For LTE-M devices, SCUBA uses the radio link control, radio resource control, packet data convergence, and the non-access stratum protocols from LTE-M.

2.10.3 Standardization of SCUBA

SCUBA is an ideal candidate to be integrated into the fold of MulteFire specifications, which primarily provides solutions for operating cellular technologies in shared and unlicensed spectrum [85]. In addition, specific features of SCUBA find applicability in enhancing the performance of 3GPP LTE D2D and 5G NR SL. For example, 3GPP plans to standardize ACK/NACK-based HARQ feedback scheme to ensure reliability for NR V2X protocol which is built upon NR SL [108]. Similarly, the notion of SL DRX has been incorporated currently as a work item by 3GPP to be introduced in NR SL [109]. The targets of the work item in-

clude designing collision avoiding and device-aware SL-POs, similar to the idea we introduce for SCUBA in Section 2.7.1.

2.10.4 Co-existence with Non-SCUBA Devices

This work analyzed the collision of only SCUBA transmissions with each other in the unlicensed frequency bands. However, it should be noted that there is possibility that SCUBA communication gets interfered with the transmissions from non-SCUBA devices operating in the same band. In Europe, radio frequency identification (RFID) devices which are mainly used in manufacturing and distribution, individual item tagging, and asset tracking, operate in 865-868 MHz band. On the other hand, in North America, RFID, ZigBee, location and monitoring transmitters, sensors, and wireless headphones operate in 902-928 MHz band. All the above devices including SCUBA UEs are expected to operate by conforming to the regulatory requirements, e.g., limiting the duty cycle and/or ERP, which act as inherent interference mitigation mechanism. However, when the number of devices increases, the SCUBA UEs may experience high interference from non-SCUBA UEs. We need to analyze the severity and impact of such interference and devise additional methods to mitigate the effects. To this end, we can explore using conventional interference mitigation mechanisms such as power control, frequency hopping, and retransmission after a random back-off.

2.11 Summary

In this chapter, we first investigated the feasibility of designing an in-device time division multiplexed D2D-U protocol for cellular devices. By studying regulatory requirements for the unlicensed bands across geographical locations, we devised guidelines for a prospective D2D-U protocol to function under various underlying cellular traffic models. We identified

time opportunities for D2D-U in each of the operating cellular modes and analyzed the mode probabilities with different types of machine-type communication traffic being served. We simulated the MAC layer of the LTE-M protocol to show that a potential D2D-U protocol must prioritize designing efficient communications when the UEs are in one of the DRX modes. Our feasibility analysis and simulations suggested that integrating D2D-U within conventional cellular UEs, including in low-cost HD-FDD Cat-M1 LTE-M devices, is feasible, and is a suitable cost-effective alternative to either using licensed D2D communication or introducing an additional radio for accommodating one of the already existing D2D-U protocols.

Then, we designed the protocol SCUBA which enables direct communication between cellular devices in the unlicensed frequency bands. It offers the unique benefit over the state-of-the-art unlicensed D2D RATs that it coexists with the underlying legacy cellular protocol while *reusing the existing hardware*. We provided the PHY/MAC layer specifications of SCUBA, including the SL-DRX technique that provides an application controlled flexibility between latency and power consumption. SCUBA also includes optional features of a low latency mode for near instantaneous SL transmission and a SCUBA-SAM mode for achieving reasonable SL latency targets under busy cellular traffic conditions. The collision rates, power consumption, and latency analyses and simulation results prove that SCUBA is an appealing low-cost and low-power solution for SL communication on unlicensed bands.

Chapter 3

Synchronization for SCUBA

3.1 Introduction

In Chapter 2, we designed a novel protocol called SCUBA for D2D communications in unlicensed bands. SCUBA operates in a TDM manner with the underlying cellular RAT, e.g., LTE-M. Due to the TDM nature, SCUBA functions as an S-RAT and utilizes the existing radio hardware to maintain a single radio architecture in low-cost IoT and MTC devices. Furthermore, it supports operation in the HD-FDD devices and is also upward compatible with other modes of duplexing. Additionally, SCUBA does not require any guidance (e.g., resource allocation) from a centralized BS as is the case with several of the existing D2D communication schemes [47–49, 51].

Although SCUBA solves several existing challenges for achieving SL communication on unlicensed bands for low-cost UEs, a fundamental requirement for SCUBA communication is to have the communicating UEs synchronized with each other. This is because an SRC communicates with a DST by transmitting SCUBA data on a dedicated time-slot, called the SL-PO of the DST, which it computes using a pre-defined relation. To determine the exact location of the SL-PO in time and to further communicate with each other, the two UEs must be in sync with each other. When both the UEs are in homogeneous coverage (HC), i.e., being served by the same BS, the devices are perfectly synchronized. However, there are at least three other types of scenarios where this is not the case:

1. Out-of-coverage (OOC): where both devices are out of cellular coverage area
2. Partial coverage (PC): where one of the UEs is in coverage while the other is not
3. Coverage-out-of-sync (COOS): where the two UEs are served by two different BSs that are not synchronized with each other.

In terrestrial networks with static BSs, stationary UEs always remain in one of the four coverage scenarios and can therefore use tailored solutions. For example, when two devices are in HC, legacy SCUBA can be successfully used, whereas two UEs in OOC can use other commercial D2D techniques. However, mobile UEs are likely to encounter a different condition at different instances of time. This requires a unified solution that can adapt and operate in any given coverage scenario that could change at any time. Recall from Chapter 1 that our primary goal is to ensure seamless network coverage for C-IoT UEs by eliminating the coverage gaps and unserved regions. Therefore, SCUBA still requires enhancements to achieve the main objective of extending the cellular coverage. To this end, we propose an enhancement to legacy SCUBA, which has already been shown to be the superior D2D RAT among prior arts with respect to power consumption, network latency, and heterogeneous interoperability. Our synchronized SCUBA protocol is compatible with both stationary UEs as well as nodes in an Internet-of-Mobile-Things (IoMT) environment to ensure seamless operation under all of HC, OOC, PC, and COOS conditions.

Furthermore, we note that due to the lack of synchronization mechanisms and methods to estimate timing advance, the coverage range of legacy SCUBA is restricted by the length of cyclic prefix used. Using LTE-M specifications with an orthogonal frequency division multiplexed (OFDM) sub-carrier spacing of 15 kHz [12], the maximum range of SCUBA is limited to 1400 meters, which is smaller than typical non-urban macro cell sizes [110]. To counter this issue, we present unified methods that do not only provide synchronization to SCUBA UEs but also extend the range to any arbitrary value that is not limited by time-

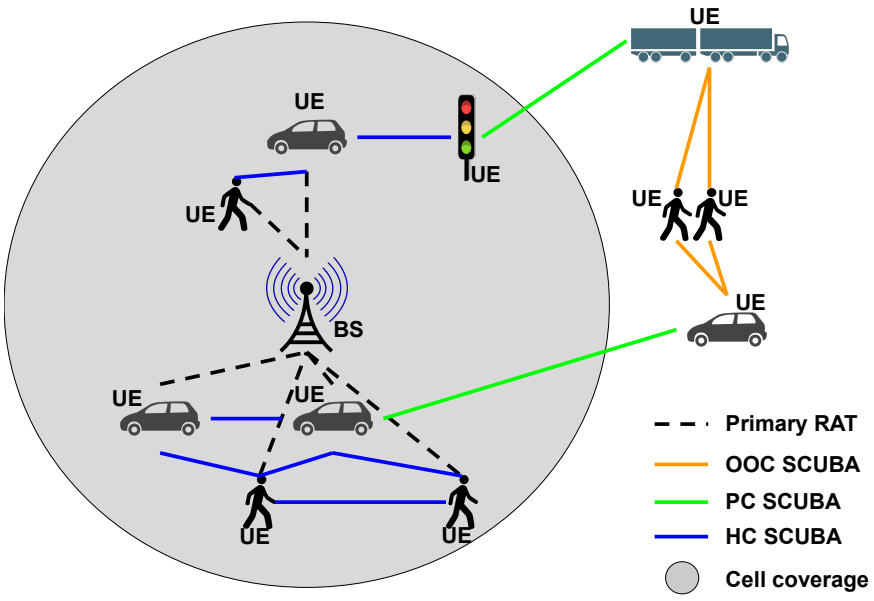


Figure 3.1: An illustration of SCUBA coexisting with P-RAT network.

of-flight. In the following, we present our solution along with a review of prior arts and its potential adaptations. We first begin by describing our system model.

The rest of this chapter is organized as follows. The system model is presented in Section 3.2. We provide the design of SCUBA synchronization methods in Section 3.3 and their performance analyses in Section 3.4. In Section 3.5, we show the numerical results, followed by a summary of the chapter in Section 3.6.

3.2 System Model

We consider a hybrid system model, where the UEs operate in two TDM modes of RATs. The UEs operate an *LTE-like* protocol such as, LTE-M, or NB-IoT as the P-RAT, and SCUBA as the S-RAT. An illustration of heterogeneous SCUBA network coexisting with P-RAT is shown in Figure. 3.1 for different coverage scenarios such as HC, PC, and OOC, as discussed

Table 3.1: UE specification for different PWCs.

Class	Transmit Power	Bandwidth	Geo Regulation
PWC1	14 dBm	$865 - 868$ MHz	Europe
PWC2	23 dBm	$902 - 928$ MHz	North America

in Section 3.1. While the UEs which are in HC communicate to each other via SCUBA by utilizing the common sync achieved through the cellular coverage, the UEs in OOC, PC, and COOS require new sync methods to have successful communication. Within the devices themselves, a SCUBA SRC transmits to a DST only when both the UEs are free from operation in their P-RATs. Figure. 2.6 illustrated this operation with the TDM nature of SCUBA that allows it to coexist with the P-RAT using the same radio architecture. We also provided the details of P-RAT and S-RAT in Section 2.5.

3.2.1 SCUBA UE Power Classes

Based on the regulations governing the use of different unlicensed frequency bands, we categorize SCUBA UEs into two power classes (PWCs): PWC1 and PWC2. The physical layer specifications of the UEs in these two PWCs are shown in Table 3.1, which are derived from the geo- and band-specific regulations.

3.3 Synchronization for SCUBA

We begin by investigating if the existing synchronization strategies in the literature can be adopted to SCUBA. Wireless technologies that include a centralized BS, e.g., LTE, usually have the BS broadcasting a periodic synchronization signal (SS) that enables the UEs in the network to maintain sync by resynchronizing frequently. However, this is a power-expensive sync scheme to be directly employed in a distributed network like SCUBA. Technologies such as Bluetooth and Wi-Fi transmit a preamble [44,111] attached to every data packet which

assists the receiver UE to synchronize to the sender and further decode the data. Alternatively, in-coverage LTE-D2D [13] utilizes the periodic synchronization provided by the LTE BS. However, the OOC and PC LTE-D2D UEs follow a synchronization approach similar to Bluetooth and Wi-Fi, where data is preceded by an SS. Since SCUBA is entirely designed by considering the underlying cellular P-RAT as the baseline protocol, i.e., reusing radio and chipset of the underlying P-RAT, and since SCUBA network operates similar to the LTE-D2D network under different coverage scenarios, we consider LTE-D2D as a starting point for our SCUBA sync design.

LTE-D2D uses a technique where a cell-edge or OOC SRC always transmits an SL synchronization signal (SLSS) before sending SL data [13]. The DST then synchronizes to the SRC using the SLSS before trying to decode the SL data. In LTE-D2D, the DST listens for SLSS within a time range, called SL synchronization window (SLSW) [81], which can be configured to either 5 ms or half of the normal cyclic prefix (CP). However, this method requires the UE to resynchronize frequently to maintain the synchronization error within the SLSW, resulting in a high power consumption. Furthermore, since the SLSW duration is fixed, it cannot be varied to either increase it to accommodate for larger sync errors or decrease to reduce power consumption for smaller sync errors.

To quantify the extent of achieved synchronization, we define *coarse* sync to be the condition when the time sync between the UEs, Δ_{sync}, follows

$$t_{\text{CP}} \leq \Delta_{\text{sync}} \leq t_{\text{SLSW}}, \tag{3.1}$$

where t_{CP} is the cyclic prefix length and t_{SLSW} is the length of the SLSW, both in time. Similarly, we define two UEs to be in *fine* sync with each other when

$$\Delta_{\text{sync}} < t_{\text{CP}}. \tag{3.2}$$

For successful SCUBA operation that ensures that an SRC can precisely compute the SL-PO of the DST for SL data transfer, it is essential to meet the condition (3.2).

With this backdrop, we analyze the potential adaptations of the LTE-D2D compatible synchronization methods and their applicability to SCUBA.

3.3.1 Adaptations of Prior Art

Inter-Cell Synchronization

UEs in COOS may be synchronized due to the inter-cell synchronization available between BSs of different cells [112]. However, this inter-cell synchronization is not always guaranteed [41], and thus falls short of being a feasible solution for SCUBA synchronization.

Sync Range Extension

UEs that are OOC may often be able to synchronize to a nearby BS since they can decode SLSSs far beyond the boundary of the supported user-plane range [41]. However, this method requires significant time duration in the order of several seconds to synchronize given the low SNR conditions [41]. This in turn critically impacts the battery life of the UE.

GNSS-Sync

GNSS based synchronization is shown to achieve $\Delta_{\text{sync}} \gtrsim 40$ ns [113]. While it meets the condition in (3.2), GNSS-based synchronization has several drawbacks. First, it is a power hungry scheme consuming several seconds for synchronization. Second, due to link budget constraints of commercial GNSS configurations, it can only work in outdoor environments. Furthermore, it also potentially requires additional hardware for GNSS signal processing. As a result, GNSS-based synchronization is unsuitable for low-cost and low-power C-IoT applications.

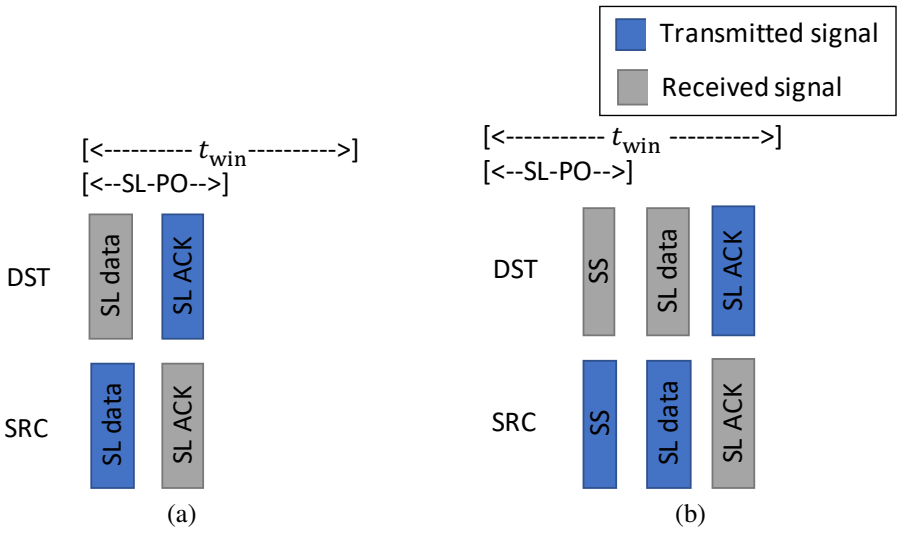


Figure 3.2: SCUBA synchronization when (a) the SRC is in fine sync with the DST, and (b) when the SRC is not.

TX-Beacon Method

LTE-D2D includes a provision to configure specific UEs, regardless of their cellular coverage, to periodically transmit SLSS, so as to extend synchronization to OOC UEs. These UEs act as SL TX beacons which periodically transmit SLSS on pre-defined SFs followed by a broadcast message which includes the timing information such as SFN [13]. Such a method may be suitable for SCUBA. However, we show later in Section 3.5.2 that an adaptation of the beacon-based method, which uses an RX beacon based strategy, is potentially more suitable for low-power C-IoT applications.

3.3.2 Proposed Solutions

Flexi-Sync Method

We present our first solution, called flexi-sync method, for two types of pre-sync scenarios.

When UEs are coarsely synchronized: Coarse synchronization can be achieved when

the UEs are in the COOS and their respective BSs provide coarse synchronization between them. We borrow the idea of SLSW from LTE-D2D, and enforce SCUBA DST UEs to listen for a duration of t_{win} around its SL-PO. However, unlike LTE-D2D, we let t_{win} to be customizable by the UE application and the extent of coarse sync achievable. Accordingly, every SCUBA transmission is preceded by the SRC transmitting an SLSS on the SL-PO of the DST, which the SRC computes using its coarse timing. We replace the DST SL-PO with two listening occasions. The first is the conventional SL-PO where the DST listens for and decodes potential SL data from UEs that are already in fine sync with the DST. The second is a SCUBA sync window (SSW), which overlaps with the SL-PO, where the UE looks for and decodes an SLSS. We choose t_{win} such that it accommodates the total of all types of timing synchronization errors, ϵ_{t}, which is

$$\epsilon_{\text{t}} = \epsilon_{\text{coarse}} + \epsilon_{\text{SRC}} + \epsilon_{\text{DST}} + t_{\text{d}}, \tag{3.3}$$

where ϵ_{coarse} is the synchronization error resulting from coarse synchronization, ϵ_{SRC} and ϵ_{DST} are the SRC and DST crystal clock errors, respectively, and t_{d} is the delay due to the time of flight between SRC and DST. By accommodating t_{d} in (3.3), our method ensures synchronization between UEs that are spaced arbitrarily far away from each other. The value of ϵ_{coarse} is dependent on the type of coarse synchronization achieved, as discussed in Section 3.3.1. The crystal clock errors, ϵ_{SRC} and ϵ_{SRC}, are given by

$$\epsilon_{\text{SRC}} = x_{\text{SRC}} \cdot t_{\text{coarse}}, \tag{3.4}$$

$$\epsilon_{\text{DST}} = x_{\text{DST}} \cdot t_{\text{coarse}}, \tag{3.5}$$

respectively, where x_{SRC} and x_{DST} are the SRC and DST crystal clock inaccuracies per unit time respectively, and t_{coarse} is the time elapsed since the latest coarse synchronization. The

value of t_{coarse} can either be set beforehand or varied dynamically. Therefore, the SRC needs to resynchronize close to when it transmits SL data to ensure that its error is within the defined maximum ϵ_{SRC}. In Figure. 3.2, we show different examples of SCUBA operation with our flexi-sync method. In Figure. 3.2 (a), we illustrate the case where $\Delta_{\text{sync}} < t_{\text{CP}}$. In such a case, a transmission of SLSS by the SRC is not required. Therefore, the SRC can directly transmit the SL data, which can be perfectly decoded at the DST. This option is suitable when the SRC is aware that it is in fine sync with the DST, either due to a previous successful SCUBA transmission in the near past or due to the prior knowledge that the SRC and DST are in HC. Thereby, we ensure power-optimized downward compatibility with HC SCUBA. The second case in Figure. 3.2 (b) is the generic scenario where the SRC transmits an SLSS before its SL data transmission. The DST listens for SLSS during its SSW and upon reception of an SLSS, synchronizes itself to the SRC and then decodes the SL data that arrives in the following time slots.

When UEs are not coarsely synchronized: We propose an adaptation to the flexi-sync method that caters to the scenarios where UEs do not have coarse sync with each other. Such a situation may be encountered at cold-start, after losing coarse sync due to inactivity, or when COOS UEs do not have coarse synchronization. When the UEs are not even coarsely synchronized with each other, an SRC is not only unaware of the SL-PO of the DST but can also not reach its SSW. Therefore, we let an SRC UE transmit an SLSS followed by an SL timing request message (TimeREQ) until it receives an SL timing response (TimeRSP) from any UE in the listening neighborhood.

The number of attempts, N_{A}, required by the SRC to encounter an SSW of the DST is dependent on t_{win}. A longer t_{win} requires a smaller N_{A} and vice versa. The choice of N_{A} and t_{win} drives the power consumption at the SRC and DST, respectively. Figure. 3.3 (a) and Figure. 3.3 (b) show instances of varying t_{win} to demonstrate the impact on N_{A}. Our

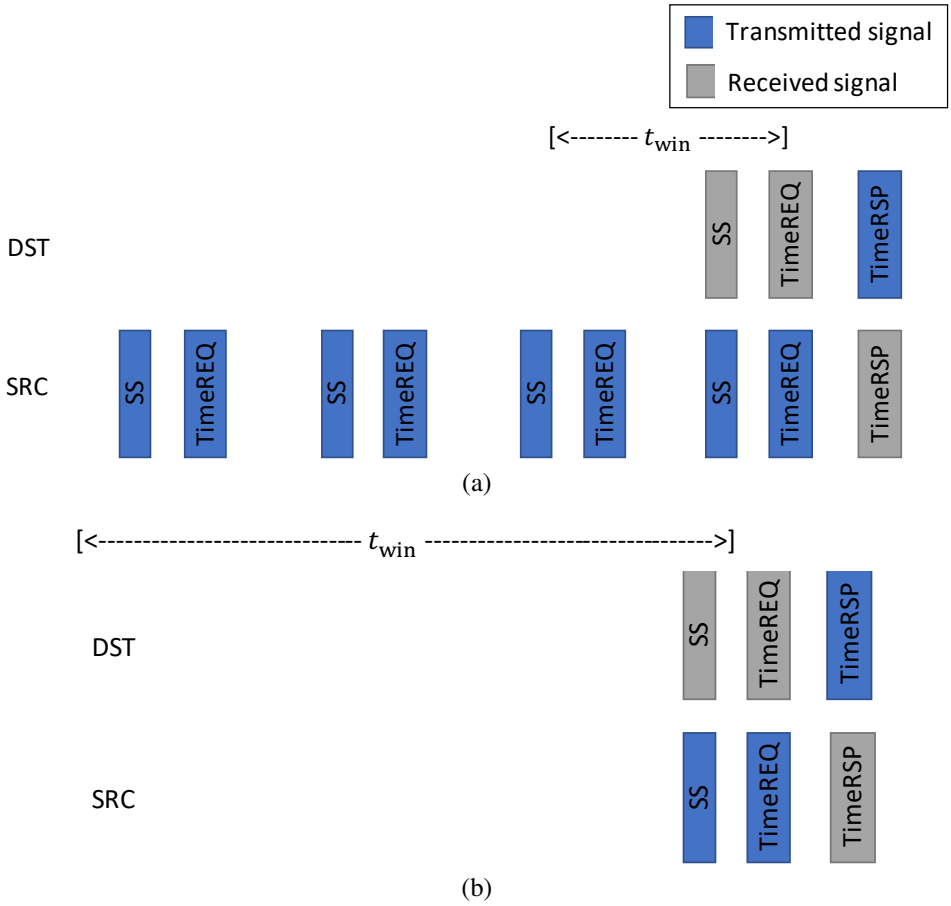


Figure 3.3: Flexi-sync operation with TimeREQ and TimeRSP for (a) shorter t_{win} with $N_{\text{A}} = 4$, and (b) extended t_{win} with $N_{\text{A}} = 1$.

method can also be modified to piggyback TimeREQ message with the SLSS to further reduce signaling by the SRC at the expense of increased decoding complexity at the DST. The trade-offs between the power consumption of an SRC and DST is typically chosen during system configuration based on the traffic type and battery life constraints at the UE.

Sync Beacon Method

For our second proposed method, we borrow the beacon-based synchronization scheme from LTE-D2D. This technique is suitable for UEs which do not have a strict constraint on power consumption, e.g., UEs with large batteries or alternating current (AC) powered devices. Bea-

con based synchronization is also applicable for UEs with or without coarse synchronization.

We define two types of beacons, TX and RX beacons, which transmit and receive beacons, respectively. A TX beacon approach is borrowed directly from LTE-D2D standard [13], where a SCUBA UE transmits beacons at regular intervals for synchronization by any other SCUBA UE. However, this method is inefficient for SCUBA UEs synchronizing to the beacons by listening for long SSW, especially for PWC1 SCUBA UEs having comparable TX and RX powers. Instead, we propose an RX beacon method, where beacon UEs meant to provide synchronization to other SCUBA devices listen for a TimeREQ message that may be transmitted by a SCUBA UE intending to obtain timing information. This is similar to the TimeREQ method with the modification that the receiving UE is an RX beacon that is meant to serve the SCUBA network to provide timing information on demand. In Section 3.5.2, we show that an RX beacon method provides superior energy efficiency when used with low transmit power devices, e.g., PWC1 SCUBA UEs. It should be noted that the beacon devices are also SCUBA UEs which have their P-RATs operating in TDM manner with SCUBA, and hence beacon periodicity or continuity is not always guaranteed. In cases where RX beacons are incapable of listening for sync requests continuously or TX beacons are unable to send SLSS at frequent intervals, the synchronizing SCUBA UE may make multiple attempts to achieve a successful sync.

3.3.3 Choosing Sync Methods

Our proposed solutions are independent of each other and are capable of providing synchronization in all scenarios, including cold-start. While the beacon-based synchronization methods allow UEs to maintain sync by resynchronizing periodically with a beacon node that has no restrictions on power consumption, the flexi-sync method is a data-driven technique that requires UEs to sync only when exchanging SCUBA messages. However, the network may

also choose to use a mixture of the two methods. For example, a network can use a beacon node with a large synchronization interval, with the option of also using flexi-sync between two UEs. This allows the UEs to use a smaller t_{win} and still encounter low values of N_{A} for beacon-based synchronization. This provides the network with greater flexibility in choosing latency and battery life trade-off.

3.3.4 SLSS Design

The design of SLSS can be adopted directly from LTE-D2D as the signals largely serve the same purpose. However, we present two modifications in the following to adapt it for SCUBA applications.

SCUBA Server Substitution

The use of our proposed methods not only solves the issue of synchronization in UEs for OOC, PC, and COOS conditions, but also eliminates the need for a central SCUBA server. Legacy SCUBA relies on a central server to extract DST information such as DST UE ID and SL-DRX cycle values to compute SL-PO. On the other hand, using our synchronization methods, where an SRC-based synchronization is supported, eliminates the need for SRC to compute the SL-PO of the DST beforehand prior to initiating SCUBA transmission. Furthermore, the UE ID and the SL-PO of the DST is also embedded within the SLSS so that the SRC can use them for subsequent transmissions without contacting a central server.

Pseudo-Unique SLSS

The SLSSs are meant to be received by a DST or an RX beacon in the SSW that is positioned around the SL-PO. SCUBA allocates SL-PO to be pseudo-unique by having them be dependent on the UE ID of the DST. However, while a 1 ms long SL-PO, as defined in SCUBA,

can be largely non-overlapping between DST UEs even in a crowded network, the SSW is considerably larger in time than the SL-PO. Therefore, the probability of inter-SSW overlap is higher, and so is the rate of SLSS collision. To counter this, we propose pseudo-uniqueness to be embedded within the SLSS using the DST UE ID, when known. This reduces the probability of false alarms of detecting an SLSS at the DST UE.

3.4 Performance Analysis

In this section, we analyze the power consumption for achieving synchronization using our proposed methods.

3.4.1 Flexi-Sync

For the flexi-sync method, we consider the performance independently for SRC and DST UEs since they consume power asymmetrically. This allows us to prioritize the performance individually. For an SRC UE that requires N_{A} attempts to achieve synchronization, the average power consumed is

$$P_{\mathrm{SRC}} = \frac{N_{\mathrm{A}}}{T_{\mathrm{data}}}\Big(P_{\mathrm{TX}}\big(t_{\mathrm{SS}} + t_{\mathrm{req}}\big) + P_{\mathrm{RX}} t_{\mathrm{rsp}}\Big), \tag{3.6}$$

where P_{TX} is the SCUBA transmission power at the UE, P_{RX} is the SCUBA reception power at the UE, T_{data} is the mean inter-arrival time of SCUBA data, and t_{SS}, t_{req}, and t_{rsp} are the time durations of SLSS, TimeREQ, and TimeRSP signals, respectively. The corresponding power consumption in the DST is given by

$$P_{\mathrm{DST}} = \frac{1}{T_{\mathrm{data}}}\Big(P_{\mathrm{RX}} t_{\mathrm{win,\,eff}} + P_{\mathrm{TX}} t_{\mathrm{rsp}}\Big), \tag{3.7}$$

where

$$t_{\text{win, eff}} = \frac{t_{\text{win}}}{N_{\text{A}}} \tag{3.8}$$

is the effective reduced SSW when multiple attempts are performed.

3.4.2 Resync Using TX and RX Beacons

For the beacon method, we analyze the power consumption in a SCUBA UE synchronizing periodically to a TX or RX beacon. Since the beacons are generally AC powered devices which are not power-critical, we do not analyze the power consumption in them. The total power consumption in a SCUBA UE periodically synchronizing to an RX beacon is given by

$$P_{\text{sync, RXbeacon}} = \frac{N_{\text{A}}}{T_{\text{sync}}} \Big(P_{\text{TX}} \big(t_{\text{SS}} + t_{\text{req}} \big) + P_{\text{RX}} t_{\text{rsp}} \Big), \tag{3.9}$$

where T_{sync} is the sync interval of the UE, which corresponds to the maximum error allowed in the SCUBA system. On the other hand, the power consumption in a SCUBA UE for synchronizing with a SCUBA TX beacon is given by

$$P_{\text{sync, TXbeacon}} = \frac{N_{\text{A}}}{T_{\text{sync}}} \big(P_{\text{RX}} t_{\text{win}} \big). \tag{3.10}$$

3.5 Numerical Results

Since C-IoT applications are not latency critical, and SCUBA is designed to function as an S-RAT only when UEs are free from P-RAT, we consider power consumption as the metric to evaluate the performance of our proposed methods. Unless otherwise specified in the following sections, we use the values shown in Table 3.2 for the evaluations, which are obtained from [13, 102, 114].

Table 3.2: Evaluation settings for SCUBA synchronization.

Parameter	Value	Parameter	Value
P_{TX}	100 mW	T_{data}	2 hours
P_{RX}	80 mW	t_{win}	72 ms
t_{SS}	1 ms	T_{sync}	8.33 min
$t_{\text{req}} = t_{\text{rsp}}$	1 ms	N_{A}	1

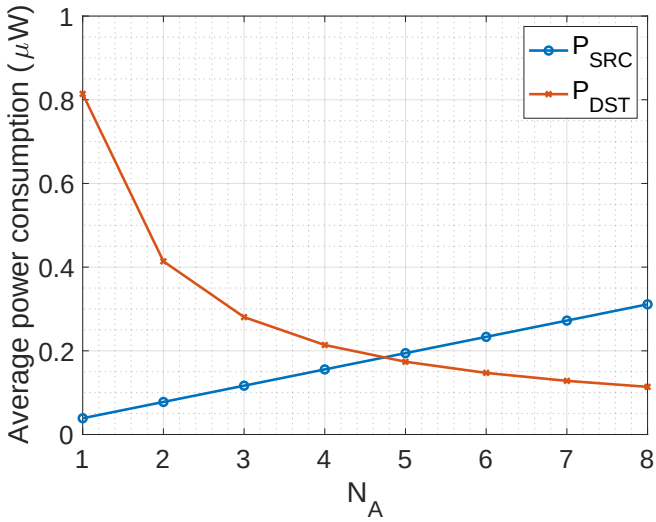


Figure 3.4: Average power consumption in SRC and DST for SCUBA resync using multiple attempts.

3.5.1 Flexi-Sync Method

We show the power consumption results of our flexi-sync solution in Figure. 3.4 for a range of values of N_{A}. We observe that while the power consumption in SRC increases linearly with increasing N_{A}, DST consumes lesser power with higher number of SRC TX attempts as it is required to listen for smaller durations of SSWs. Therefore, the results suggest that, if the SRCs in a SCUBA network are power-critical devices and the DSTs are not, for e.g., sensors reporting to a central controller node, lower number of sync attempts along with longer duration of SSW is preferable.

3.5.2 Beacon-Based Method

The variation of the total power consumption in a SCUBA UE for each of the beacon-based methods against a range of sync intervals is shown in Figure. 3.5 (a). For the evaluation, we choose an optimal t_{win} corresponding to each value of sync interval. For RX beacon based sync scheme, the power consumption in the SCUBA UE reduces with increasing values of sync interval. For the TX beacon based sync scheme, the power consumption in the UE remains constant regardless of the sync interval since t_{win} is chosen optimally. For $T_{\text{sync}} > 350$ s, SCUBA UE synchronizing to an RX beacon has lower power consumption compared to that of TX beacon. This suggests that RX beacon based synchronization is preferable when the sync interval in the SCUBA network is large.

Next, we present the evaluation results to show the impact of transmit power on the total power consumption in Figure. 3.5 (b). For this evaluation, we choose $T_{\text{sync}} = 8.33$ min which corresponds to the time during which the accumulated sync error reaches 5 ms, equivalent to the allowed SLSW in LTE-D2D. The power consumption in SCUBA UE for RX beacon based method increases with increasing values of transmission power. However, since there are no transmissions involved in the SCUBA UE when synchronizing to a TX beacon, the power consumption remains constant for the TX beacon based sync scheme. The power consumption traces intersect at $P_{\text{TX}} = 160$ mW, which corresponds to SCUBA ERP of 45 mW (16.53 dBm) for 45% power amplifier efficiency and 60 mW power consumption in the support circuitry [102]. We call the intersection point of power consumption traces of RX and TX beacon based methods as beacon power threshold (P_{BTh}), which plays a crucial role in choosing the appropriate sync scheme for a SCUBA network. The results suggest that RX beacon based synchronization scheme is preferable for SCUBA UEs belonging to those power classes transmitting at power lower than P_{BTh}. It should also be noted that the value of P_{BTh} will be higher for higher values of T_{sync}.

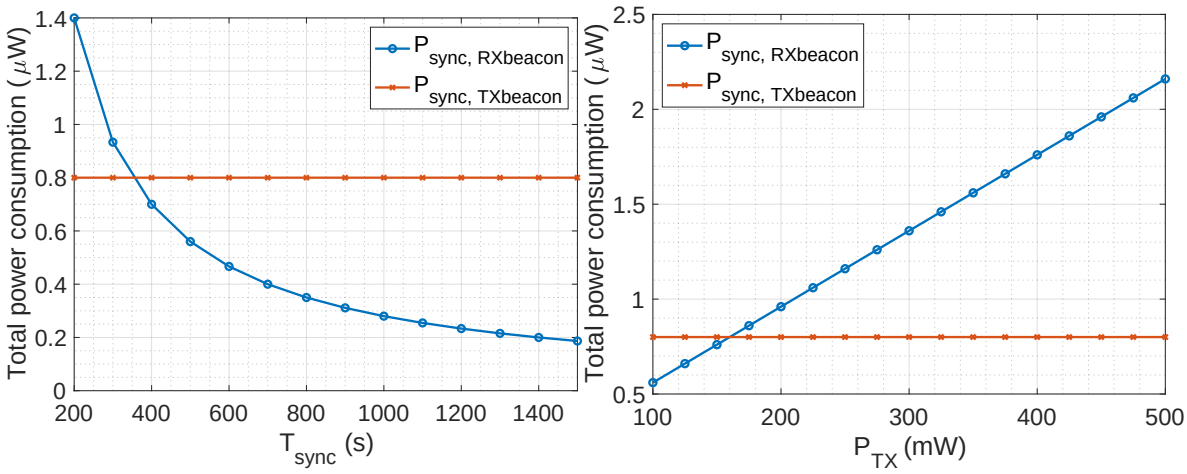


Figure 3.5: Variation of total power consumption for the beacon-based synchronization method as a function of (a) sync interval and (b) transmit power.

3.5.3 Battery Life

Finally, we present numerical values of the impact of our proposed solutions on the bottom-line metric for C-IoT device performance of battery life. To this end, we integrate our power consumption numbers with the operating power of native SCUBA protocol from Chapter 2. Since availability of non power-critical beacon devices is not always guaranteed in a SCUBA network, we use flexi-sync method for the battery life analysis. For an MTC traffic model [90], the battery life of a legacy SCUBA UE that coexists with LTE-M as P-RAT and utilizes the LTE-M network sync for SCUBA, is 328.3 days. For a SCUBA device that only uses the flexi-sync method for synchronization, the battery life is 328.1 days with $P_{\text{TX}} = 100$ mW and $N_{\text{A}} = 4$. The flexi-sync based SCUBA sync, which enables UE synchronization in all types of cellular coverage scenarios, thus results in less than 0.1% reduction in battery life compared to legacy SCUBA condition that uses the LTE-M network sync. Thus, our method extends the operation range and provides seamless operation in all coverage scenarios for mobile C-IoT devices only at a cost of negligible reduction in battery life.

3.6 Summary

In this chapter, we presented synchronization schemes for SCUBA devices to enable seamless D2D communication in mobile C-IoT UEs across all types of cellular coverage scenarios. Our solutions also ensure that time-of-flight does not limit the communication range of SCUBA devices. We provide flexible solutions that are adaptable based on UE hardware limitations and unlicensed band usage regulations. Numerical results showed that our proposed low-power solutions can achieve synchronization with a negligible impact on UE battery life. Comparison of other application-specific performance indicators for tailored network topologies is a straightforward extension of our work. Although our proposed methods are intended for SCUBA, they are also adaptable to other types of D2D communication technologies, such as NR sidelink. Synchronized SCUBA also strengthens the potential of being integrated into the MulteFire standard, which provides solutions for operating cellular communications on unlicensed bands.

Chapter 4

SPIN: Self-Positioning for IoT Non-Terrestrial Networks

4.1 Introduction

We discussed in Chapter 1 that our main research goal is to design enhancements for C-IoT networks for achieving contiguous universal coverage. In the previous chapters, we designed protocol SCUBA to enable low-cost C-IoT UEs to directly communicate via D2D links. SCUBA complements the conventional C-IoT network by closing the network coverage gaps while also providing other D2D-related benefits. Now, we focus on the problems related to NTN IoT which is primarily aimed at realizing a global coverage for C-IoT UEs. In this chapter, we design solutions for solving the UL synchronization problem (as explained in Section 1.1.3) caused mainly by the Doppler offset in NTN. To design a positioning solution which solves the UL synchronization problem, it is worthwhile to study the cellular positioning techniques developed for terrestrial networks to determine if there are any solutions adoptable for NTNs. The currently standardized positioning methods for the terrestrial cellular networks mainly include observed time difference of arrival (OTDOA), uplink time difference of arrival (UTDOA), and enhanced cell ID (E-CID) [41,115]. While OTDOA is based on TDOA measured by the UE using the DL reference signals, UTDOA is based on the TDOA measured by the BS using the UL sounding reference signals (SRS). On the other hand, E-CID uses the geographical location of the BS as a rough estimate for the UE location, along

with one or more additional measurements such as reference signal received power (RSRP), angle of arrival (AoA), TA, or RTT to get a finer estimate. However, both AoA and UTDOA require UE to transmit SRS resulting in huge battery drain, which is not feasible for an IoT UE. On the other hand, RSRP based methods are very sensitive to the characteristics of the channel and hence provide poor accuracy. Furthermore, their positioning accuracy does not benefit from the signal characteristics such as duration and bandwidth [116]. More importantly, in all the above cellular positioning methods, the location server (LS) estimates the UE location using the measurements reported by either the UE or the BS [41]. Such solutions can operate only after the UE establishes an RRC connection with the network, and hence they cannot be used to solve the NTN UL synchronization issues. However, if self-positioning is performed by the UE using the TDOA measurements on DL broadcast signals, it could be used to solve the UL synchronization problem. Similarly, the existing positioning solution for NR NTN [33] performs TDOA based self-positioning at the UE. To this end, inspired from both OTDOA and [33], we design SPIN, Synchronization signal-based Positioning in IoT Non-terrestrial networks, which performs TDOA measurements on the DL SSs, to enable self-positioning in an IoT UE. In contrast to the terrestrial network, the BSs in an NTN, i.e., the satellites, are moving, which results in varying Doppler shifts. Consequently, the UE experiences frequency differences between the DL signals received from one or more satellites at same or different time instants. Exploiting these additional measurements, we design SPIN such that the positioning algorithm includes both FDOA and TDOA measurements.

Thus, considering the shortcomings of the state-of-the-art solutions and the power constraints associated with IoT UEs, we design SPIN which

- solves the NTN UL synchronization problem by meeting the target accuracy requirements,
- utilizes both TDOA and FDOA measurements on PSS and SSS,

- achieves the theoretical bounds of position and velocity estimation accuracy,
- neither requires termination of the RRC connected mode nor interrupts the ongoing communication, thus resulting in minimal battery drain, and
- does not require any network modification, additional reference signals, or extra control signaling.

SPIN first estimates TOA and frequency of arrival (FOA) of SSs from one or more satellites. It then computes the differences between TOA and FOA measurements to get TDOA and FDOA measurements, respectively. Finally, SPIN estimates the UE's position and velocity from the joint set of TDOA and FDOA measurements, which it uses to compute and compensate residual TA and Doppler.

The rest of the chapter is organized as follows. We present the preliminaries including the system model and target requirements in Section 4.2, and the SPIN algorithm in Section 4.3. We discuss the theoretical bounds of positioning accuracy and the energy consumption in Section 4.4, and their numerical results and computational complexity in Section 4.5. In Section 4.6, we briefly discuss the behavior of our positioning algorithm under additional scenarios. Finally, a summary of the chapter is provided in Section 4.7.

4.2 Preliminaries

4.2.1 System Model

We consider an NTN which consists of one or more LEO satellites orbiting the earth, terrestrial UEs, gateway, radio access network (RAN) and core network (CN), as shown in Figure. 4.1. The NTN illustration shown in Figure. 4.1 depicts a transparent payload type architecture where the RAN is located on the earth [25, 26]. Conversely, there is regenerative payload type architecture where the RAN is located on-board the satellite [25, 26]. However,

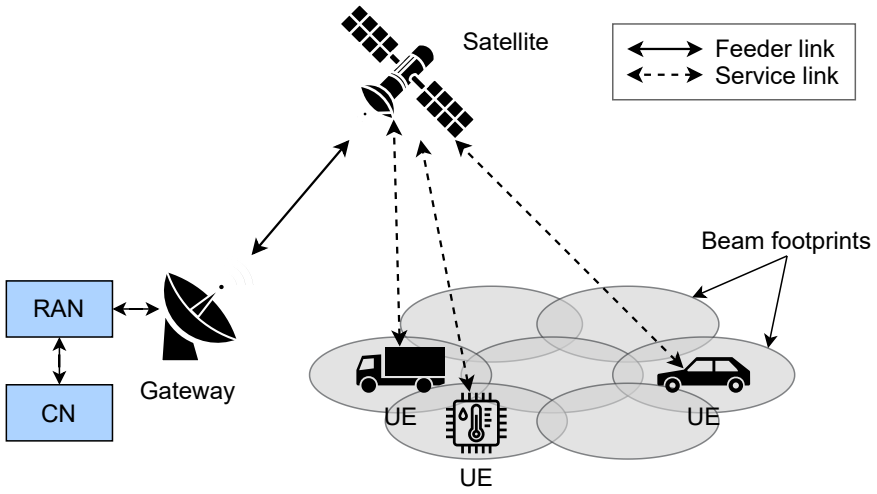


Figure 4.1: An illustration of a non-terrestrial network.

our design of SPIN is agnostic to the network architecture and hence works in both cases. In an NTN, the radio link which serves the UEs is referred to as the service link, whereas the one which connects the satellite with the gateway is called the feeder link. In the service link, the UEs are served by one or more beams from the satellite, where single or multiple beams can correspond to a cell depending on the network implementation [25]. When the NTN consists of a constellation of satellites, optionally, inter-satellite links (ISL) exist between them. The backhaul consists of the RAN, e.g., 5G RAN, which enables the cellular communication through the selected RAT, and the CN, which connects eventually to the data network. To serve the NTN IoT applications, 3GPP has chosen NB-IoT and LTE-M, the standards which are already implemented and currently in use in the terrestrial cellular networks [63]. Using NB-IoT and LTE-M standards for NTN IoT applications gives several advantages. We can reuse the existing infrastructure, which minimizes implementation cost, and also gives stable and predictable performance. Therefore, in our work, we consider NB-IoT and LTE-M standards as the baseline RATs to design and evaluate our solution.

Radio Access Technology

As we discussed above, we design and evaluate SPIN by considering NB-IoT and LTE-M as the underlying RATs. While 3GPP considers Ka-band (20 GHz) frequency for NR NTN, which enables high data rate applications, it chooses S-band (2 GHz) for NTN-IoT, which is ideal for IoT applications. Both NB-IoT and LTE-M standards use OFDM modulated signals for communication. The NB-IoT and LTE-M OFDM signals are structured as 1 ms time units called SFs, which consist of 14 OFDM symbols. In the frequency domain, the OFDM subcarriers are spaced 15 kHz apart. While the LTE Cat-M1 standard, the widely used LTE-M variant, occupies a bandwidth of 1.4 MHz, the NB-IoT standard uses a narrow bandwidth of 180 kHz. To enable synchronized communication between the BS and the UEs, LTE-M and NB-IoT standards periodically broadcast PSS, SSS and NPSS, NSSS, respectively. These SSs are implemented using standard sequences such as Zadoff-Chu (ZC) and/or m-sequences, which hold excellent correlation properties. In the following, unless the standard is specifically mentioned, we use the general terms SS, PSS, and SSS to indicate synchronization signals in both LTE-M and NB-IoT.

The PSS and SSS are known to the UE immediately after the DL sync. From the numerical values of Cramér-Rao Lower Bound (CRLB) discussed in Section 4.5, we observe that these SSs have sufficient bandwidth and duration to be utilized for positioning. Hence, we propose to use PSS, SSS and NPSS, NSSS in LTE-M and NB-IoT respectively, for the purpose of positioning in SPIN. The DL SSs received at the UE can be represented as

$$r_{\mathrm{s}}(t) = (s * h)(t) + w(t), \tag{4.1}$$

where s is the transmitted SS which includes both PSS and SSS, $w(t)$ are additive white Gaussian noise (AWGN) samples, which are assumed to be $w(t) \sim \mathcal{N}(0, \frac{N_0}{2})$, and h is the

impulse response of the NTN channel. The latter is modelled by

$$h(t) = \sum_{i=1}^{L} \alpha_i(t) \exp(\mathrm{j}2\pi f_{\mathrm{o},i} t)\delta(t - \tau_{\mathrm{o},i}), \tag{4.2}$$

where L is the number of paths in the channel and α_i, $\tau_{\mathrm{o},i}$, and $f_{\mathrm{o},i}$ are the attenuation, time offset, and frequency offset of ith channel path, respectively. In addition to small scale fading which is caused mainly due to the UE velocity and the obstacles close to the UE, $\tau_{\mathrm{o},i}$ and $f_{\mathrm{o},i}$ also include time of flight from the satellite to the UE and the Doppler shift due to satellite velocity, respectively [26]. It follows that, the signal $r_{\mathrm{s}}(t)$ is acquired by the UE at a sampling rate, $f_{\mathrm{s}} = 1/T_{\mathrm{s}}$, resulting in a discrete time signal, $r_{\mathrm{s}}(nT_{\mathrm{s}})$, where $n = 0, \pm 1, \pm 2, \ldots$ is the discrete time index. In the following, for simplicity, we use $r_{\mathrm{s}}(n)$ to represent $r_{\mathrm{s}}(nT_{\mathrm{s}})$.

Synchronization Process

During DL synchronization, an IoT UE performs a series of operations on the received DL signal to get time and frequency synchronization [117]. In an NTN, the DL SSs are already pre-compensated at the BS for the common Doppler with respect to an RP in the cell. Therefore, the received frequency, f_{RX}, suffers only from residual Doppler depending on the displacement of the UE from the RP. However, the frequency offset estimated by the UE in DL synchronization is the sum of the residual Doppler and the oscillator errors as shown in Figure. 4.2. Subsequently, the UE adjusts its frequency to f_{RX} and transmits UL signal at a frequency spaced at duplex distance away from f_{RX}. On the received signal, the satellite performs post-compensation of the common Doppler offset with respect to the RP to partially compensate for the Doppler effect in UL channel. The resultant signal experiences a frequency offset of twice the residual Doppler offset as shown in Figure. 4.2. In addition to the above method of UL synchronization, [37] discusses another option where the UE pre-compensates its UL signal by the total frequency offset estimated in the DL.

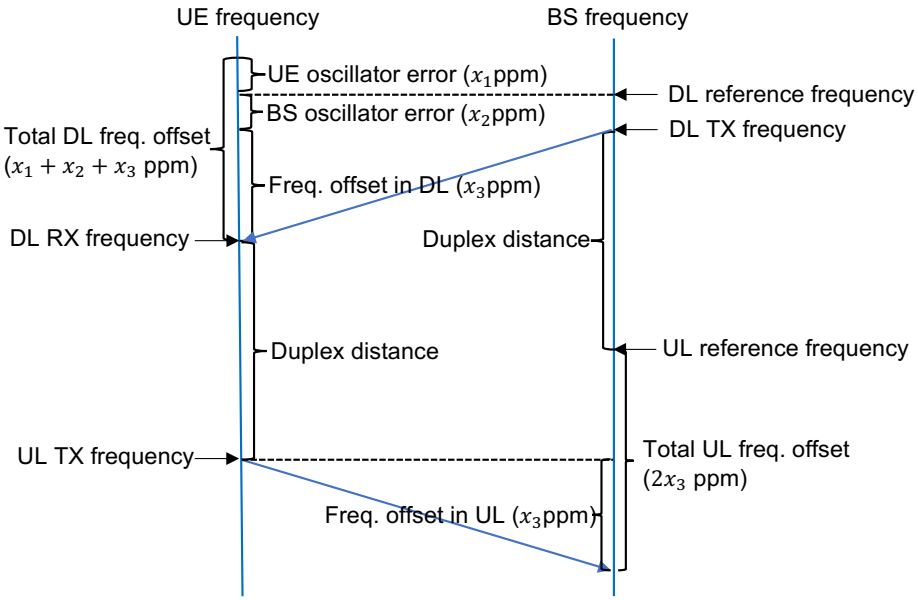


Figure 4.2: UL synchronization in NTN [37].

However, this method results in higher residual frequency offsets than the method discussed above. Therefore, in this work, we assume that the UE uses the method shown in Figure. 4.2 for UL synchronization, which results in lower residual frequency offsets. The values of residual frequency offsets for different LEO orbits and beam configurations[5] are given in Table 4.1 [37, 118]. It should be noted that the maximum allowed frequency offset in UL as per the standard is 0.1 ppm [119]. However, from Table 4.1, it is evident that the residual frequency offsets in NTN are significantly high compared to the allowed limit. Similarly, after the DL synchronization process, the UE time synchronization also suffers from residual TA with respect to the RP in the cell.

[5]The beam configuration refers to the satellite sets defined in [25, 63], which differ in terms of antenna parameters and the resulting beams.

Table 4.1: UL residual frequency offset in ppm.

Beam configuration	LEO 1200 km	LEO 600 km
Set 1	1.92	2.20
Set 2	3.92	3.86
Set 3	21.56	21.14
Set 4	-	39.90

Reference Coordinate System and Satellite Constellation

For SPIN, we consider the earth-centered earth-fixed (ECEF) coordinate system, which is a Cartesian coordinate system where the axes are named as X, Y, and Z in three-dimensional space [42, Ch. 2]. The ECEF system is termed earth-centered since the origin is the center of mass of the earth, i.e., the *geocenter*, and earth-fixed since the coordinate system is fixed with respect to the earth, i.e., the axes rotate along with the earth. The Z-axis aligns with the rotation axis of the earth and hence passes through the geocenter and the north pole. Furthermore, the X-axis intersects the equator of the earth at the *Greenwich meridian*. Finally, the Y axis is defined with reference to the X-axis such that it completes a right-handed orthogonal coordinate system. Compared to other coordinate systems, ECEF system is more convenient to calculate the line-of-sight (LOS) distance between two points [42, Ch. 2]. Therefore, it is ideal for SPIN that considers TDOA measurements which are functions of pseudo-range difference measurements. Moreover, global positioning system (GPS), the widely used GNSS solution which is owned and operated by the USA, adopts ECEF for positioning. This also makes ECEF a preferable coordinate system for SPIN, which effectively operates as an alternative to GNSS.

For the design and evaluation of SPIN, we consider a reference constellation which is represented in the standard format, $i : t/p/f$, where i is the inclination of the orbital plane, t is the total number of satellites, p is the number of orbital planes, and f is the phasing parameter which denotes the relative spacing between the adjacent orbital planes [120, 121].

In our evaluation, we consider a near-polar Walker-star LEO satellite constellation, which gives global coverage with simple constellation design [122]. Regardless, SPIN is agnostic to the type of constellation and works if the UE gets visibility to a sufficient number of satellites. In Figure. 4.3, we show an example of a LEO near-polar Walker star constellation which uses a pattern of $87.5^\circ : 100/10/5$, represented in an ECEF coordinate system.

In the design and evaluation of SPIN, we use the following notations to represent the position and velocity vector in the ECEF coordinate system. Let $[x_i\ y_i\ z_i]$ and $[v_{x_i} v_{y_i} v_{z_i}]$ be the position coordinates and the velocity vector of the satellite, respectively. Since SPIN uses a set of multiple measurements from the satellites, which we define in later sections, the subscript i denotes the index of the associated measurement in the set. Further, we represent the unknown location coordinates and the velocity vector of the UE as $\boldsymbol{X} = [x\ y\ z]$ and $\boldsymbol{V} = [v_x\ v_y\ v_z]$, respectively.

4.2.2 Target Requirements

In the following, we find the upper bound of the positioning error for solving the UL synchronization issues. To begin with, the Doppler shift, f_{Doppler}, experienced in an NTN relative to the carrier frequency, f_{c}, is given by

$$\frac{f_{\text{Doppler}}}{f_{\text{c}}} = \frac{\langle \boldsymbol{V}_{\text{sat-UE}}, \boldsymbol{U}_{\text{sat-UE}} \rangle}{c \left\| \boldsymbol{U}_{\text{sat-UE}} \right\|}, \tag{4.3}$$

where $\langle \cdot, \cdot \rangle$ represents the vector dot product operation, $\|\cdot\|$ represents the vector magnitude, $\boldsymbol{V}_{\text{sat-UE}}$ is the relative velocity between the satellite and the UE, $\boldsymbol{U}_{\text{sat-UE}}$ is the position vector from the satellite to the UE, and c is the speed of light. The first order Doppler error [124] is

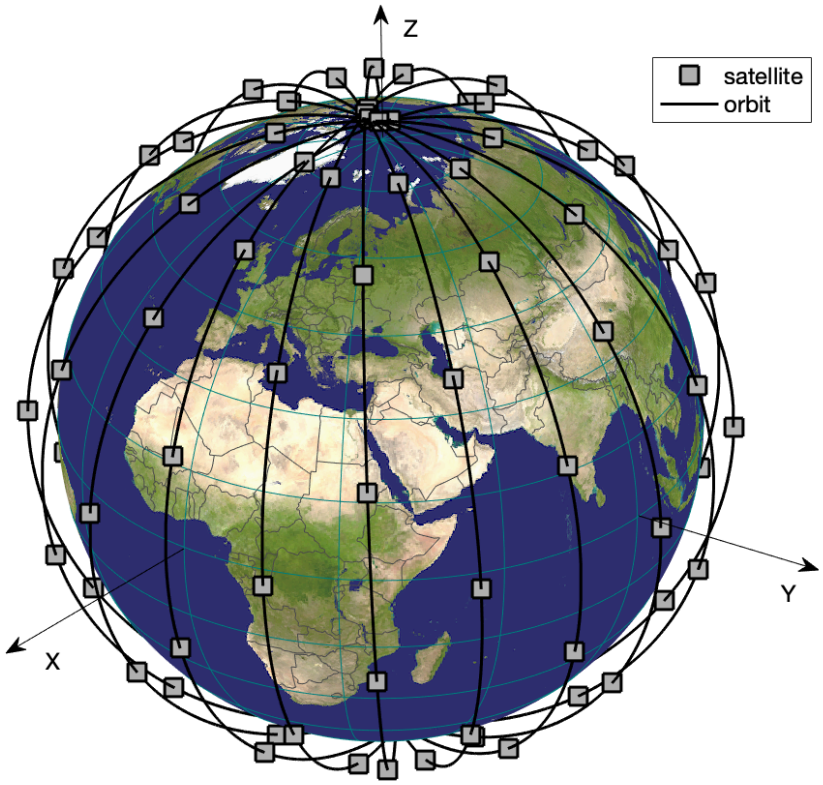


Figure 4.3: An illustration of LEO satellite constellation around the earth, represented in an ECEF coordinate system. The earth is plotted using [123].

given by

$$F_{\mathrm{e}} = \frac{\Delta f_{\text{Doppler}}}{f_{\mathrm{c}}} = \underbrace{\langle \boldsymbol{V}_{\text{sat-UE}}, \boldsymbol{P}_{\boldsymbol{U}_{\text{sat-UE}}} \frac{\Delta \boldsymbol{U}_{\text{sat-UE}}}{c \left\| \boldsymbol{U}_{\text{sat-UE}} \right\|} \rangle}_{P_{\mathrm{e}}} + \underbrace{\langle \Delta \boldsymbol{V}_{\text{sat-UE}}, \frac{\boldsymbol{U}_{\text{sat-UE}}}{c \left\| \boldsymbol{U}_{\text{sat-UE}} \right\|} \rangle}_{V_{\mathrm{e}}}, \tag{4.4}$$

where $\Delta f_{\text{Doppler}}$ is the Doppler error, $\Delta \boldsymbol{V}_{\text{sat-UE}}$ is the relative velocity error, $\Delta \boldsymbol{U}_{\text{sat-UE}}$ is the error in position vector $\boldsymbol{U}_{\text{sat-UE}}$, and $\boldsymbol{P}_{\boldsymbol{U}_{\text{sat-UE}}}$ is the orthogonal projection to $\boldsymbol{U}_{\text{sat-UE}}$. Hence, P_{e} is the portion of the Doppler error attributed to position error, and V_{e} is the portion attributed to relative velocity error. From (4.4), the position and relative velocity errors can be expressed as

$$\left\| \Delta \boldsymbol{U}_{\text{sat-UE}} \right\| \leq P_{\mathrm{e,max}} \frac{c}{\boldsymbol{V}_{\text{sat-UE}}} \left\| \boldsymbol{U}_{\text{sat-UE}} \right\| \tag{4.5}$$

and

$$\|\Delta \boldsymbol{V}_{\text{sat-UE}}\| \leq V_{\text{e,max}} c, \tag{4.6}$$

respectively. In (4.5) and (4.6), $P_{\text{e,max}}$ and $V_{\text{e,max}}$ represent the portion of the maximum Doppler error attributed to error in position and velocity respectively, and they are related to the maximum Doppler error, $F_{\text{e,max}}$, as,

$$F_{\text{e,max}} = P_{\text{e,max}} + V_{\text{e,max}}. \tag{4.7}$$

The achievable satellite position and velocity accuracy available in the public domain [125] are

$$\|\Delta \boldsymbol{U}_{\text{sat}}\| \leq 3 \text{ m}$$

and

$$\|\Delta \boldsymbol{V}_{\text{sat}}\| \leq 0.2 \text{ m/s}. \tag{4.8}$$

As per 3GPP standards, a maximum frequency error of 0.1 ppm is allowed in the UL to facilitate successful reception at the BS [119]. Since the Doppler shift due to UE velocity in NTN is same as that of a terrestrial network [37], we disregard the UE velocity factor in the following. Hence, we replace $\Delta \boldsymbol{V}_{\text{sat-UE}}$ with $\Delta \boldsymbol{V}_{\text{sat}}$, and we compute $V_{\text{e,max}}$ from (4.6) using (4.8). In the NR UL synchronization study conducted in [124], 80% of the frequency error is attributed to the crystal oscillator drift after DL synchronization and the remaining 20% to the residual Doppler shift. Since the residual Doppler error in the UL is twice the one-sided Doppler, the maximum Doppler pre-compensation error, $F_{\text{e,max}}$, is 0.01 ppm for the above error ratio. Now, we consider LEO satellite at an altitude of 600 km and a velocity of 7.69 km/s. We approximate $\|\boldsymbol{U}_{\text{sat-UE}}\|$ by the satellite orbit altitude, which is the minimum distance between the UE and the satellite. Using the above error ratio, $V_{\text{e,max}}$, (4.5), and (4.7),

we compute the allowed position error, $\|\Delta \boldsymbol{U}_{\text{sat-UE}}\|_{\text{max}} = 218.5$ m. To guarantee the overall positioning error requirement represented by $\|\Delta \boldsymbol{U}_{\text{sat-UE}}\|_{\text{max}}$, the satellite and UE position error should satisfy

$$\|\Delta \boldsymbol{U}_{\text{UE}}\| + \|\Delta \boldsymbol{U}_{\text{sat}}\| \leq \|\Delta \boldsymbol{U}_{\text{sat-UE}}\|_{\text{max}} . \tag{4.9}$$

Considering the worst-case satellite position error, represented by $\|\Delta \boldsymbol{U}_{\text{sat}}\|_{\text{max}}$, the UE position error should satisfy

$$\begin{aligned} \|\Delta \boldsymbol{U}_{\text{UE}}\| &\leq \|\Delta \boldsymbol{U}_{\text{sat-UE}}\|_{\text{max}} - \|\Delta \boldsymbol{U}_{\text{sat}}\|_{\text{max}} \\ &= 218.5 - 3 = 215.5 \text{ m}. \end{aligned} \tag{4.10}$$

If the above limit is satisfied by UE positioning, the overall positioning error requirement is guaranteed regardless of the satellite position error.

We performed the above analysis for a fixed error budget of $80\% - 20\%$ between crystal offset and Doppler shift errors. However, in practice, the oscillator errors in the UE can be different. Therefore, we have also plotted the allowed position errors for different error budget allocations in Figure. 4.4.

In the above, we discussed about the UL frequency synchronization problem and the target requirement for UE positioning to resolve the issue. As we already discussed in Section 1.1.3, UL time synchronization also faces issues. For successful detection at the BS, the timing error should be within $\pm\frac{\text{CP}}{2}$, where CP is the cyclic prefix. Correspondingly, the required positioning accuracy is given by $\|\Delta \boldsymbol{U}_{\text{sat-UE}}\| \leq \frac{\text{CP}}{4} c = \pm 7250$ m [124]. Compared to the UL frequency synchronization problem, residual TA problem requires more relaxed positioning accuracy. Therefore, we consider the accuracy needed for UL frequency sync as the target requirement for our positioning solution.

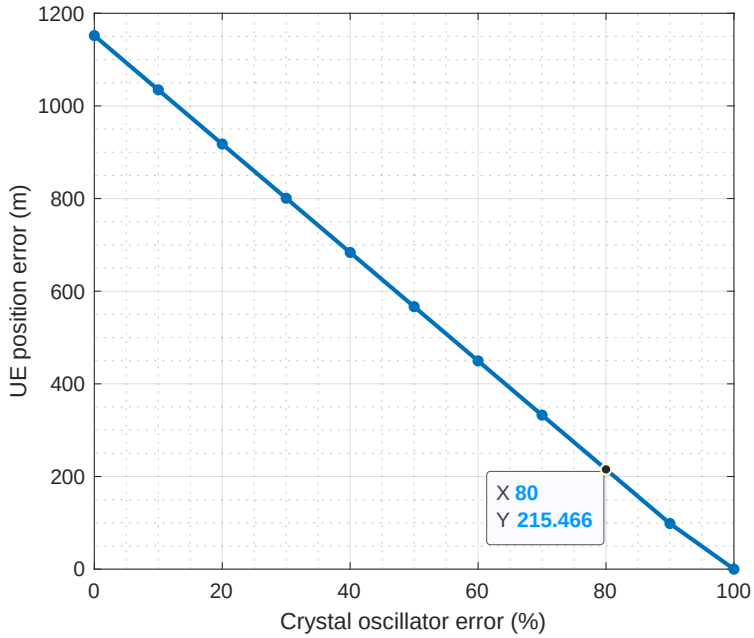


Figure 4.4: Maximum allowed position error for varying crystal oscillator error.

4.3 SPIN

We design SPIN by utilizing the varying TOA and FOA of the DL SSs transmitted by one or more satellites at same or different instants of time. Since TOA and FOA measurements suffer from UE clock time and frequency offsets, they are error-prone to be directly used for positioning. Therefore, SPIN takes differences between TOA and FOA measurements to obtain TDOA and FDOA measurements, respectively. Next, we need to solve UE position and velocity from the TDOA and FDOA measurements, which is a nonlinear and non-convex problem [126]. There are several state-of-the-art solutions available which can be broadly classified into two categories. The first category consists of two-step weighted least squares (2-WLS) solutions, which give closed-form position and velocity estimates [61, 127]. 2-WLS methods represent the nonlinear equations in linear form by adding additional unknown nuisance parameters. 2-WLS and modified 2-WLS methods discussed in the literature have

very low computational complexity [61, 127]. However, these methods suffer from poor accuracy due to high bias [128]. In the second category of methods, position and velocity are obtained using iterative algorithms [129–131]. Unfortunately, unless sufficiently accurate initial position and velocity are assumed, these methods converge to wrong local minima, resulting in high error. Foy's method, alternatively called Taylor series-based weighted least squares (TWLS) method, is one such iterative method [130]. TWLS uses the first order Taylor series to approximate the nonlinear equations as linear and solves them iteratively using weighted least squares. In [132], 2-WLS method is used to obtain rough estimate of position which then initializes the TWLS method. Finally, TWLS method delivers fine position estimates. Inspired from [132], we adopt a similar method for SPIN in the context of IoT NTN. We use the 2-WLS method to get rough estimates of position and velocity and use them to initialize TWLS, which finally delivers finer estimates. We provide a summary of the features supported in relevant prior-art positioning solutions and our solution in Table 4.2.

Like any other real-time positioning solutions, SPIN incorporates an acquisition block which acquires position for the first time during a cold-start. A UE usually performs acquisition when it is in RRC idle mode and gets a UL data request from the upper layers. Before initiating the connection request, the UE requires its own position to compute and compensate the residual Doppler during RA. In this case, the UE does not have valid measurements from the past to be used for positioning. However, after obtaining position and velocity fix by acquisition, SPIN maintains the accuracy within acceptable limits by performing tracking. In the following, we define SPIN by first introducing the SPIN block diagram, and then having a detailed discussion of the individual blocks and the operations involved in the positioning algorithm.

We show the block diagram of SPIN in Figure. 4.5, which mainly consists of three blocks, namely, TDOA and FDOA estimation, SPIN acquisition, and SPIN tracking. In the following,

Table 4.2: Summary of features in prior-art positioning solutions compared with SPIN.

Features \ Solutions	[33]	[133]	[134]	[135]	[61]	[130]	[62, Ch. 5]	[132]	SPIN
NTN	✓	✓	×	×	×	×	✓	×	✓
IoT	×	×	✓	×	×	×	×	×	✓
PSS	✓	×	×	✓	×	×	×	×	✓
SSS	×	×	×	✓	×	×	×	×	✓
Fine resolution/ curve fitting	×	×	×	×	×	×	×	×	✓
TDOA	✓	✓	✓	✓	✓	✓	✓	✓	✓
FDOA	×	✓	×	×	✓	×	✓	×	✓
Taylor series/ iterative	✓	✓	✓	×	×	✓	✓	✓	✓
Weighted	×	✓	✓	×	×	✓	×	✓	✓
2-WLS/ 2-WLS initialize	×	×	×	×	✓	×	✓	✓	✓
Time multiplexed tracking	×	×	×	×	×	×	×	×	✓
Self positioning	✓	✓	×	✓	✓	-	✓	✓	✓
Solves UL sync problem	*	*	×	*	*	*	*	*	✓

* represents that the solution has either insufficient accuracy or the accuracy has not been evaluated.

we explain each of these blocks and their functions.

4.3.1 TDOA and FDOA Estimation

TOA and FOA Estimation

SPIN first performs TOA and FOA measurements on SSs arriving from one or more satellites over a period called SPIN window (W_{SPIN}). For TOA and FOA estimation, SPIN reuses a part of the synchronization algorithm used in cellular DL and performs some additional operations. For instance, when SPIN operates in an NB-IoT UE, it can reuse the cellular NB-IoT DL synchronization algorithm to get time and frequency synchronization. An example of a low-complexity synchronization method for NB-IoT is given in [117]. If the NB-IoT UE uses

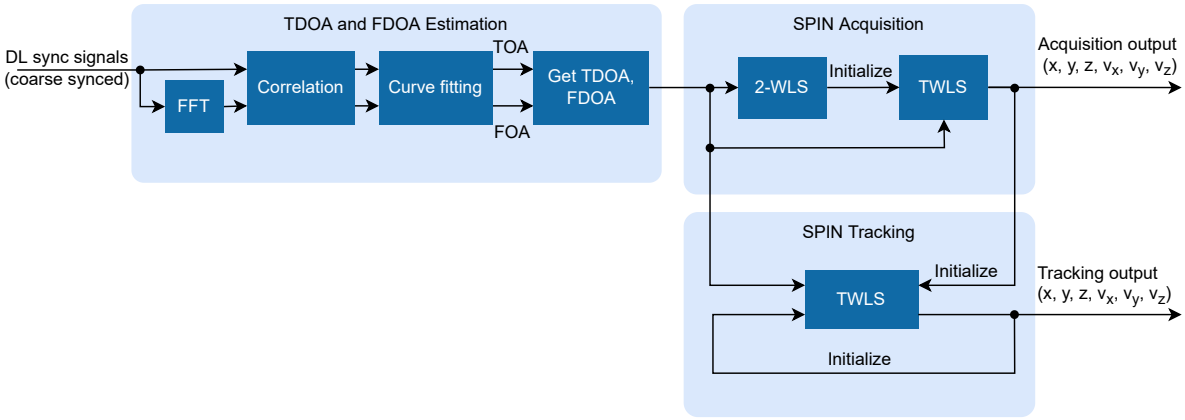


Figure 4.5: Block diagram of SPIN.

this method, SPIN can reuse the sliding auto-correlation based synchronization detection and coarse time-frequency synchronization as discussed in [117]. The coarse synchronization is followed by a fine synchronization step which involves cross-correlation of the received signal with the known templates of PSS. Since the channel SNR may vary depending on the coverage scenario, the correlation may require several synchronization signals to be combined. The UE may perform 1 ms coherent combining [136] followed by incoherent combining over longer duration to synchronize successfully. A successful synchronization detection delivers TOA and FOA estimates. However, the accuracy of these estimates are insufficient for SPIN mainly for two reasons. First, the cellular synchronization process in the UE does not necessarily use SSS to obtain the synchronization, especially during the initial cell access when the SSS sequence is unknown to the UE. Second, the TOA and FOA accuracies are largely affected by the finite sampling rate. Regardless, SPIN utilizes the time and frequency synchronization output of the DL synchronization process to refine the time-frequency search space. However, to achieve the target positioning accuracy, SPIN performs some additional operations.

First, SPIN performs cross-correlation of combined PSS and SSS templates with the received signal. Like the frequency offset estimation in cellular UEs [137], SPIN estimates

FOA by frequency-domain cross-correlation. To this end, the correlation block in Figure. 4.5 performs two types of cross-correlations, pre-FFT cross-correlation to estimate TOA and post-FFT cross-correlation to obtain FOA. The pre- and post-FFT signal correlations are represented by

$$\rho(\tau) = \sum_{n} r_{\mathrm{s}}(n) s^{*}(n+\tau) \tag{4.11}$$

and

$$\Lambda(\varphi) = \sum_{k} R_{\mathrm{s}}(k) S^{*}(k+\varphi), \tag{4.12}$$

respectively, where $s(n)$ is the SS template in discrete form sampled at f_{s}, $R_{\mathrm{s}}(k)$ and $S(k)$ are the FFTs of $r_{\mathrm{s}}(n)$ and $s(n)$, respectively. τ and φ are the pre- and post-FFT cross correlation lags, respectively. The lags at which the cross-correlation peaks occur in (4.11) and (4.12) are found by

$$\hat{\tau} = \arg \max_{\tau} |\rho(\tau)| \tag{4.13}$$

and

$$\hat{\varphi} = \arg \max_{\varphi} |\Lambda(\varphi)|, \tag{4.14}$$

respectively.

Second, SPIN incorporates fine-resolution curve fitting, which operates on the magnitude of pre- and post-FFT cross correlation outputs to get finer TOA and FOA estimates. To this end, we use 3-point parabolic curve fitting [138, 139] which uses the peak value of the cross-correlation magnitude and its adjacent two values. The fractional offsets from the estimated peak to the fine-resolution peak obtained using 3-point parabolic curve fitting on pre- and post-FFT correlation outputs are given by

$$\delta_{\tau} = \frac{|\rho(\hat{\tau}-1)| - |\rho(\hat{\tau}+1)|}{2\big(|\rho(\hat{\tau}-1)| - |\rho(\hat{\tau})| + |\rho(\hat{\tau}+1)|\big)} \tag{4.15}$$

and

$$\delta_{\varphi} = \frac{|\Lambda(\hat{\varphi}-1)| - |\Lambda(\hat{\varphi}+1)|}{2\big(|\Lambda(\hat{\varphi}-1)| - |\Lambda(\hat{\varphi})| + |\Lambda(\hat{\varphi}+1)|\big)}, \tag{4.16}$$

respectively. Note that curve-fitting based fine estimators are biased estimators [139]. However, from our evaluations, we observe that the bias in our case is negligible and hence the performance is not affected. The resulting TOA and FOA estimates achieve the CRLB, which is the theoretical bound for an unbiased estimator.

It should be noted that the FOA estimated using the above steps do not include the common Doppler shift with respect to the RP in the cell. This is due to the Doppler pre-compensation already performed by the BS on the DL SSs. Since the FOA measurements are possibly acquired from multiple satellites, the common Doppler may be different for different FOA measurements. Therefore, to get correct positioning, satellite-specific common Doppler needs to be added back to get the actual FOA estimates. For this purpose, SPIN computes the common Doppler, which can be easily done using the satellite ephemeris and the coordinates of the RP.

The fine resolution TOA and FOA estimates obtained after curve fitting are given by

$$t_{\text{TOA},i} = \frac{1}{f_{\text{s}}}(\hat{\tau}_i + \delta_{\tau,i}), \qquad i = 1, 2, \ldots, N, \tag{4.17}$$

and

$$f_{\text{FOA},i} = \frac{f_{\text{s}}}{N_{\text{FFT}}}\,(\hat{\varphi}_i + \delta_{\varphi,i}) + f_{\text{com},i}, \qquad i = 1, 2, \ldots, N, \tag{4.18}$$

respectively, where the subscript i denotes the measurement index, N is the total number of measurements, N_{FFT} is the length of the FFT, and $f_{\text{com},i}$ is the common Doppler with respect to the RP. Now, we compute range measurements, r_i, using the fine TOA measurements, $t_{\text{TOA},i}$, as

$$r_i = c \cdot t_{\text{TOA},i}, \qquad i = 1, 2, \ldots, N. \tag{4.19}$$

Similarly, we compute range rate [62, 133], $\dot{r}_i$, using fine resolution FOA measurements, $f_{\text{FOA},i}$, as

$$\dot{r}_i = -\frac{c}{f_c} f_{\text{FOA},i}, \qquad i = 1, 2, \ldots, N. \tag{4.20}$$

Next, we relate these range measurements and the UE and satellite position and velocity vectors by (4.21), where ΔT is the interval between consecutive TOA/FOA measurements, η_{TO} is the effective time offset due to satellite and UE clock deviations, and n_i is the estimation error in the ith range measurement which conforms to the Gaussian distribution $\mathcal{N}(0, \sigma^2_{n_i})$.

$$r_i = \underbrace{\sqrt{\big(x_i - (x + v_x i\Delta T)\big)^2 + \big(y_i - (y + v_y i\Delta T)\big)^2 + \big(z_i - (z + v_z i\Delta T)\big)^2}}_{f_i(\boldsymbol{X}, \boldsymbol{V})} + c \cdot \eta_{\text{TO}} + n_i,$$
$$i = 1, 2, \ldots, N. \tag{4.21}$$

$$\dot{r}_i = \underbrace{\frac{(v_{x_i} - v_x)\big(x_i - (x + v_x i\Delta T)\big) + (v_{y_i} - v_y)\big(y_i - (y + v_y i\Delta T)\big) + (v_{z_i} - v_z)\big(z_i - (z + v_z i\Delta T)\big)}{\sqrt{\big(x_i - (x + v_x i\Delta T)\big)^2 + \big(y_i - (y + v_y i\Delta T)\big)^2 + \big(z_i - (z + v_z i\Delta T)\big)^2}}}_{g_i(\boldsymbol{X}, \boldsymbol{V})}$$
$$+ \frac{c}{f_c}\eta_{\text{FO}} + n'_i, \qquad i = 1, 2, \ldots, N. \tag{4.22}$$

Similarly, the range rate measurements can be expressed as (4.22), where η_{FO} is the effective frequency offset due to satellite and UE crystal offsets, and n'_i is the estimation error in the ith range rate measurement which conforms to the Gaussian distribution $\mathcal{N}(0, \sigma^2_{n'_i})$. It should be noted that the mathematical model defined above is applicable to both acquisition and tracking and hence ΔT could be replaced by either acquisition interval or tracking interval, which are defined later in this section.

In the next step, to eliminate the UE clock and crystal offsets in (4.21) and (4.22), we take differences between TOA and FOA estimates to get TDOA and FDOA measurements,

respectively.

TDOA and FDOA Measurements

For N TOA estimates, there are $\frac{N(N-1)}{2}$ possible TDOA measurements. However, when the received signal is cross-correlated with the clean template of SSs, only $N-1$ TDOAs are sufficient. This is due to the fact that the remaining combinations result in redundant TDOA adding no extra information [140]. $N-1$ TDOAs can be generated either by taking the difference with respect to a single arbitrary reference TOA or by taking differences between the adjacent TOA measurements, which we refer to as adjacent TDOA [33]. Adjacent TDOA is particularly useful in continuous SPIN tracking which comprises of a sliding time window which includes new TOA measurements while dropping old measurements continuously over time. When adjacent TDOA is used in a continuous tracking scenario, a fixed number of stale TDOA get dropped when the same number of fresh TDOA get added, whereas the rest of the TDOA measurements remain unchanged between the tracking loops. The above reasoning is also applicable for getting FDOA values from the FOA measurements. Without loss of generality, we assume adjacent TDOA and FDOA in the following. Thus, we get range and range rate difference measurements, which are the scaled versions of TDOA and FDOA, as

$$r_{ij} = r_i - r_j = c(t_{\text{TOA},i} - t_{\text{TOA},j}) \tag{4.23}$$

and

$$\dot{r}_{ij} = \dot{r}_i - \dot{r}_j = -\frac{c}{f_{\text{c}}}(f_{\text{FOA},i} - f_{\text{FOA},j}), \tag{4.24}$$

respectively, where $i = 2, 3, \ldots, N$ and $j = i - 1$. Substituting (4.23) and (4.24) into (4.21) and (4.22), respectively, we obtain

$$r_{ij} = f_i(\boldsymbol{X}, \boldsymbol{V}) - f_j(\boldsymbol{X}, \boldsymbol{V}) + \xi_{ij} \tag{4.25}$$

and

$$\dot{r}_{ij} = g_i(\boldsymbol{X}, \boldsymbol{V}) - g_j(\boldsymbol{X}, \boldsymbol{V}) + \xi'_{ij}, \tag{4.26}$$

where $i = 2, 3, \ldots, N$, $j = i - 1$, $\xi_{ij} = n_i - n_j$, $\xi'_{ij} = n'_i - n'_j$ and $f_i(\cdot)$ and $g_i(\cdot)$ are the functions indicated in (4.21) and (4.22), respectively. We denote the set of $2(N-1)$ range and range rate difference measurements obtained in this step as

$$\boldsymbol{\phi} = [r_{21}, r_{32}, \ldots, r_{N\,N-1}, \dot{r}_{21}, \dot{r}_{32}, \ldots, \dot{r}_{N\,N-1}]^{\mathrm{T}}. \tag{4.27}$$

In the next step, we use (4.25) and (4.26) to estimate the UE position and velocity.

4.3.2 SPIN Acquisition

When a position fix is required before initiating an RRC connection, SPIN performs UE location acquisition. SPIN acquisition involves obtaining TDOA and FDOA measurements from the SSs over a time window called SPIN acquisition window ($W_{\text{SPIN,acq}}$). We call the time duration of incoherent combining required for each TOA/FOA measurement as SPIN acquisition duration ($T_{\text{SPIN,acq}}$). We decide $T_{\text{SPIN,acq}}$ based on the duration needed for successful synchronization for the MCL under consideration [63, 117]. The interval between successive TOA/FOA measurements in the acquisition is referred to as SPIN acquisition interval ($I_{\text{SPIN,acq}}$). We show the SPIN acquisition timing diagram in Figure. 4.6.

To perform positioning, in addition to TDOA and FDOA measurements, UE also requires the satellite position and velocity when the SSs were transmitted. As we mentioned in Section 1.1.3, the network broadcasts satellite ephemeris in NTN system information broadcast (SIB). The UE can extrapolate the ephemeris acquired at the beginning of an RRC connection until the expiry of a network defined timer called ephemeris validity timer [141].

From (4.25) and (4.26), we get a joint set of $2(N-1)$ nonlinear range and range rate

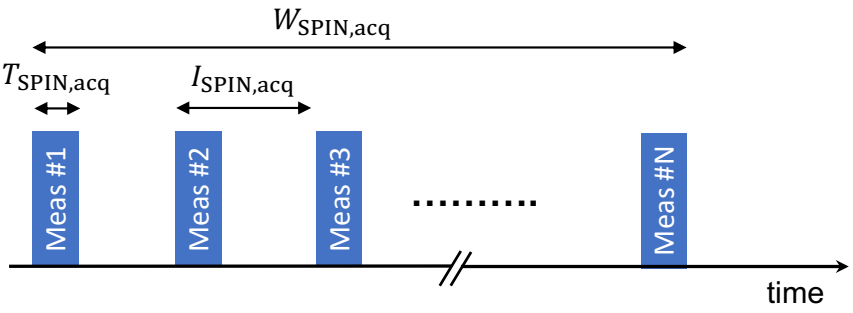


Figure 4.6: Timing diagram of SPIN acquisition.

difference equations, which are based on the TDOA and FDOA measurements taken during $W_{\text{SPIN,acq}}$. To begin with, SPIN uses the 2-WLS method to obtain initial rough estimates of UE position and velocity. In the first step of 2-WLS, the non-linear equations are represented in linear form by adding two extra estimation parameters. Thus, including the UE position and velocity vectors, 8 parameters are estimated using WLS. In the second step, using the relation between the extra parameters and the UE position and velocity, the estimation is refined to the original 6 parameters. To avoid repetition, we do not provide the details of 2-WLS method here but refer to [61].

Next, SPIN performs TWLS to get finer estimates of UE position and velocity. In TWLS, the $2(N-1)$ nonlinear equations given by (4.25) and (4.26) are first linearized using the first-order Taylor series based linear approximation. The resulting equations can be expressed in the form

$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{n}, \tag{4.28}$$

where $\boldsymbol{\theta} = [\boldsymbol{X}\ \boldsymbol{V}]^{\mathrm{T}}$ and $\boldsymbol{n} = [\xi_{21}, \xi_{32}, .., \xi'_{21}, .., \xi'_{N\ N-1}]^{\mathrm{T}}$. The WLS solution of (4.28) is given by

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{B}, \tag{4.29}$$

where $\boldsymbol{W}$ is a weight matrix of size $2(N-1) \times 2(N-1)$. The details of the Taylor series

approximation and the definitions of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{W}$ are given in Appendix A. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ contain elements which are functions of initial values of the estimation parameters. The matrix $\boldsymbol{W}$ is also unknown initially and hence we first compute that using the initial values of position and velocity. To get accurate estimates, further we perform several iterations of TWLS, every time updating $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{W}$ using the position and velocity estimates obtained in the previous iteration.

The SPIN acquisition is summarized in Algorithm 2. In the algorithm, for all the parameters and matrices which are updated iteratively, we have added superscripts inside parentheses to indicate the iteration number. Since the weight for 2-WLS is unknown initially, we use the identity matrix to get an initial position fix. Then we compute the weight matrix for the 2-WLS using the estimated UE location and velocity. We perform 2-WLS again to get more accurate estimates. We run three iterations of such reweighted 2-WLS which are sufficient to eliminate the effect of the uninformed initial estimate [142]. Then we compute the weight matrix for TWLS using the UE location and velocity estimated from 2-WLS. We also use the 2-WLS estimation results to initialize the TWLS algorithm. We run TWLS iteratively until the convergence criterion is met, i.e., the difference in estimation results between successive iterations is confined below a sufficiently low threshold, ϵ_{th}. To accommodate for the 2-WLS and TWLS computations associated with SPIN, we include an extra processing time, $T_{\text{SPIN,proc}}$, after completing the measurements in the acquisition window. One can vary the value of $T_{\text{SPIN,proc}}$ to adjust the SPIN computational complexity based on the capability of the UE.

4.3.3 SPIN Tracking

After acquisition, SPIN performs tracking to maintain the position accuracy within acceptable limits. SPIN tracking uses a similar technique for positioning as in acquisition except for

Algorithm 2: SPIN acquisition algorithm.

Input:
Range, range rate measurements: $\boldsymbol{\phi}$
Satellite positions: x_i y_i, z_i
Satellite velocities: v_{x_i}, v_{y_i}, v_{z_i}
Convergence threshold: ϵ_{th}
Maximum iteration: k_{max}

Output:
UE position estimate: $\boldsymbol{X}^{(k)} = [x^{(k)}, y^{(k)}, z^{(k)}]$
UE velocity estimate: $\boldsymbol{V}^{(k)} = [{v_x}^{(k)}, {v_y}^{(k)}, {v_z}^{(k)}]$

Initialization: 2-WLS on $\boldsymbol{\phi}$

Initialize weight to $\boldsymbol{I}_{2(N-1)\times 2(N-1)}$;
$k = 0$;
while $k \leq 3$ **do**
    Estimate $\boldsymbol{X}^{(k)}, \boldsymbol{V}^{(k)}$ using 2-WLS;
    Update weight using $\boldsymbol{X}^{(k)}, \boldsymbol{V}^{(k)}$;
    $k = k + 1$;

TWLS Positioning: TWLS on $\boldsymbol{\phi}$

Compute $\boldsymbol{W}^{(0)}$, $\boldsymbol{A}^{(0)}$, $\boldsymbol{B}^{(0)}$ using $\boldsymbol{X}^{(k-1)}, \boldsymbol{V}^{(k-1)}$;
TWLS initialize: $\boldsymbol{X}_0 = \boldsymbol{X}^{(k-1)}$, $\boldsymbol{V}_0 = \boldsymbol{V}^{(k-1)}$;
$k = 0$;
Iteration error, $\epsilon^{(k)} = \infty$;
while $\epsilon^{(k)} \geq \epsilon_{\text{th}}$ *and* $k \leq k_{\text{max}}$ **do**
    Estimate $\boldsymbol{X}^{(k)}, \boldsymbol{V}^{(k)}$ using TWLS;
    Compute $\boldsymbol{W}^{(k+1)}$, $\boldsymbol{A}^{(k+1)}$, $\boldsymbol{B}^{(k+1)}$ using $\boldsymbol{X}^{(k)}, \boldsymbol{V}^{(k)}$;
    TWLS initialize: $\boldsymbol{X}_0 = \boldsymbol{X}^{(k)}$, $\boldsymbol{V}_0 = \boldsymbol{V}^{(k)}$;
    $k = k + 1$;
    Compute iteration error $\epsilon^{(k)}$;

a few differences. SPIN tracking first initializes its TWLS using the UE location fix obtained by acquisition, and then tracks the location in loops by further initializing the TWLS using the position fixes obtained in the previous tracking loops. Like SPIN acquisition, tracking also involves a window called SPIN tracking window ($W_{\text{SPIN,track}}$), during which SPIN performs one full loop of tracking. During each $W_{\text{SPIN,track}}$, SPIN measures TOA and FOA by

combining SSs for a duration of SPIN tracking duration ($T_{\text{SPIN,track}}$). The tracking measurements are done at an interval called SPIN tracking interval ($I_{\text{SPIN,track}}$). We show the SPIN tracking timing diagram in Figure. 4.8(a). Like SPIN acquisition, tracking also includes extra processing time, $T_{\text{SPIN,proc}}$, after the tracking window.

Time Gaps for SPIN Tracking

Unlike SPIN acquisition which is usually performed before initiating an RRC connection, SPIN tracking does not get abundant time gaps. We summarize the time gaps we identified for tracking in Table 4.3. In connected active (ConA) mode, the UE gets sufficient time gaps to acquire DL SSs as the UL data SFs always precede with DL SFs which include UL grants. Along with the UL grants, UE can decode DL SSs from the serving cell for the purpose of positioning. A UE can also receive DL SSs during all other assigned and unassigned UL and DL data SFs. However, only a UE operating in full frequency division duplexing (FDD) mode can receive SSs during assigned UL data SFs. Conversely, an HD-FDD UE can only transmit UL data on an assigned UL data SF. However, a UE operating in HD-FDD additionally gets switch SFs between UL and DL SFs. As mentioned in Section 2.3.2, most of the radios require only a fraction of the switch SFs and hence the remaining time can be used for receiving DL SSs from the serving or non-serving satellite for positioning. During the DRX inactivity time, UE listens to the DL and hence can receive SSs from only the serving satellite. The round-trip wait times vary from 4 ms to 42 ms for LEO during which also the UE can potentially receive DL SSs from one or more satellites. In the current NB-IoT standard, the network provides the UEs with 40 ms gaps after every 256 ms of continuous communication for the purpose of resynchronization [56, Ch. 7]. During these gaps, UE decodes cellular reference signals (CRS) from the serving cell. For SPIN, UE can additionally decode DL SSs from the serving cell which also occur in the same SFs. The sleep times such

as CDRX-OFF and IDRX sleep are long gaps which can be certainly used for receiving SSs from both serving and non-serving satellites. In the context of GNSS based positioning in NTN, 3GPP has decided to enable the network to configure positioning gaps [141]. These time gaps can also be utilized for our positioning purpose. An illustration of position tracking operation which utilizes the switch SFs, RTT, DRX inactivity time, and DL SFs, multiplexed with NTN cellular communication is shown in Figure 4.7. In the figure, PSM stands for power saving mode where the UE stays in a dormant state by sleeping indefinitely.

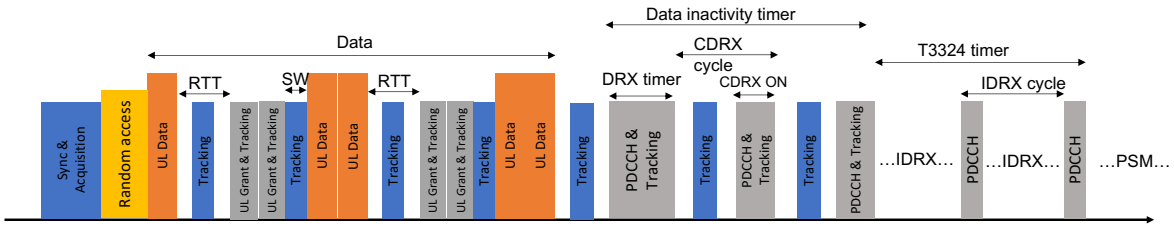


Figure 4.7: Timing diagram of SPIN tracking operation time-multiplexed with NTN cellular IoT communication

Based on the nature and availability of the time gaps discussed above, we propose two methods for SPIN tracking.

Periodic SPIN Tracking

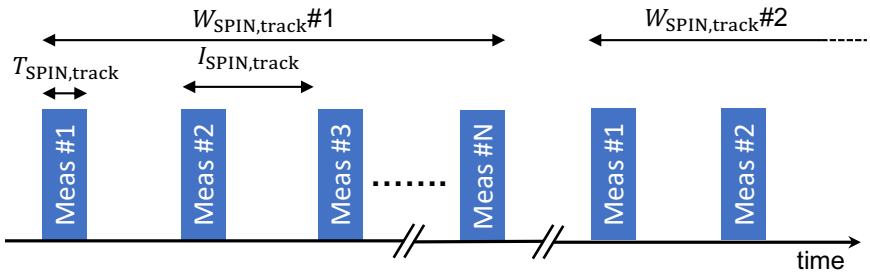
Periodic SPIN tracking involves tracking periodically at fixed intervals which provides the UE with high positioning accuracy that remains valid for a long time. To this end, we define a timer called SPIN position validity timer, $T_{\text{SPIN,val}}$, upon the expiry of which SPIN periodic tracking is mandatory. Thus, periodic SPIN tracking requires relatively long but infrequent time gaps. From Table 4.3, SPIN can potentially utilize the GNSS positioning measurement gaps, CDRX-OFF time, IDRX sleep time, and round-trip wait time for the purpose of periodic tracking. Furthermore, as shown in Figure. 4.8(a), the tracking windows in successive tracking loops are non-overlapping in this method.

Table 4.3: List of time gaps available for SPIN tracking.

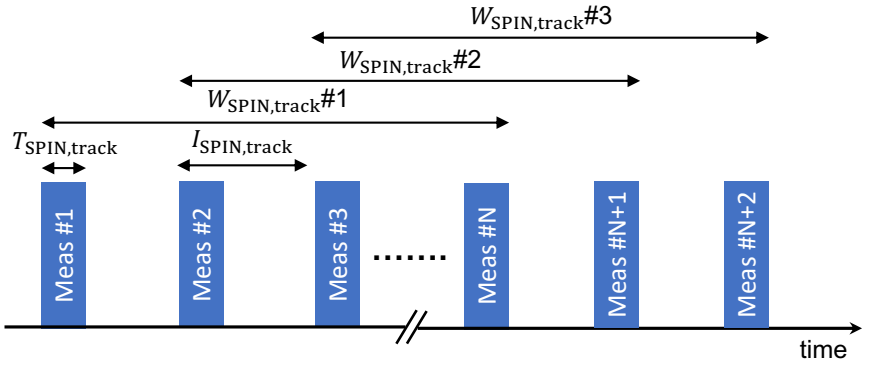
Mode	Opportunities	Usage scenario	
		Serving cell	Non-serving cell
ConA	Unassigned UL SFs	Yes	Yes
	Unassigned DL SFs	Yes	Yes
	Assigned UL SFs	Yes (FDD only)	No
	Assigned DL SFs	Yes	No
	Round-trip wait time	Yes	Yes
	Switch SFs	Yes	Yes
	Grant-to-data, data-to-ACK, and Control channel (CCH) to CCH processing gaps	Yes	Yes
	DRX inactivity time	Yes	No
	Resynchronization gaps	Yes	Yes
	Neighbor cell measurement gaps	Yes	Yes
	Positioning measurement gaps	Yes	Yes
CDRX	CDRX-ON	Yes	No
	CDRX-OFF	Yes	Yes
IDRX	IDRX-PO	Yes	No
	IDRX sleep	Yes	Yes

Continuous SPIN Tracking

Continuous SPIN tracking maintains the accuracy at a specific level by means of continuous tracking measurements. This method is ideal for the scenario when short time gaps are available frequently during the NTN cellular communication. For instance, when SPIN operates on an NB-IoT UE, resynchronization gaps can be used for tracking. Since the SFs containing the CRS for resynchronization measurements also include SSs, additionally acquiring SSs for the purpose of SPIN tracking does not affect the resynchronization operation. Similarly, as shown in Table 4.3, LTE-M UEs are usually provided with periodic time gaps for neighbor cell measurements [143]. The UE measures the reference signals (RS) in the neighbor cells during these gaps. Since both RS and SSs are present on the acquired signal, SPIN can additionally perform measurements on the SSs during these gaps. In addition to these time gaps,



(a)



(b)

Figure 4.8: Timing diagram of (a) Periodic SPIN tracking (b) Continuous SPIN tracking.

switch SFs in HD-FDD UEs, assigned and unassigned DL and UL SFs, processing gaps, and CDRX-ON time are potential time gaps which can be utilized for continuous tracking. Furthermore, in continuous SPIN tracking, the tracking window slides continuously by dropping old stale TOA/FOA measurements and by including fresh TOA/FOA measurements simultaneously. Therefore, the tracking windows in adjacent loops overlap in time, as shown in Figure. 4.8(b).

Throughout the positioning operation, we assume that the UE velocity does not change significantly during the short positioning windows, $W_{\text{SPIN,acq}}$ and $W_{\text{SPIN,track}}$. While long positioning window gives highly accurate positioning result for UEs with constant velocity, it results in poor accuracy for UEs with changing velocity. Therefore, while choosing $W_{\text{SPIN,acq}}$ and $W_{\text{SPIN,track}}$, one needs to carefully consider the UE mobility aspects and the target accu-

racy.

4.4 Performance Analysis

In the previous section, we defined the detailed steps of SPIN algorithm. In the following, we analyze the performance of SPIN using two key indicators, positioning accuracy and battery life.

4.4.1 Positioning Accuracy

We find the theoretical bounds for SPIN positioning accuracy by applying the CRLB [144, Ch. 3]. We do this in two steps: first we find the CRLB for TOA and FOA estimation, and then we use this result to obtain the CRLB for position and velocity estimation using TDOA and FDOA. For the TOA and FOA estimation in SPIN, where the received noisy signal is cross-correlated with a clean template, the CRLB inequalities [145] are given by

$$\sigma_{\mathrm{T}}^2 \geq \frac{1}{8\pi^2\beta^2\frac{E}{N_0}} \tag{4.30}$$

and

$$\sigma_{\mathrm{F}}^2 \geq \frac{1}{8\pi^2\kappa^2\frac{E}{N_0}}, \tag{4.31}$$

respectively, where σ_{T}^2 is the variance of the TOA estimation in (4.17), σ_{F}^2 is the variance of the FOA estimation in (4.18), E is the energy of $s(t)$ in the time window $T_{\text{SPIN,acq}}$ or $T_{\text{SPIN,track}}$, β is the root mean square (RMS) bandwidth, and κ is the RMS time of the signal. The RMS bandwidth and time of the signal are defined by

$$\beta = \sqrt{\frac{\int_{-\infty}^{\infty} f^2 \left|S(f)\right|^2 df}{\int_{-\infty}^{\infty} \left|S(f)\right|^2 df}} \tag{4.32}$$

and

$$\kappa = \sqrt{\frac{\int_{-\infty}^{\infty} t^2 \left|s(t)\right|^2 dt}{\int_{-\infty}^{\infty} \left|s(t)\right|^2 dt}}, \tag{4.33}$$

respectively, where $S(f)$ is the Fourier transform of $s(t)$.

Next, we use the CRLB inequalities in (4.30) and (4.31) to find the CRLB of joint TDOA-FDOA based SPIN position and velocity estimations. While closed-form expressions of TDOA CRLB are given in [146], [147, Ch. 4], we cannot use them since they are meant for 2D positioning and also do not include FDOA. Therefore, for SPIN, we apply the steps provided in the literature [148] for computing TDOA CRLB, and extend that to our scenario of joint TDOA-FDOA based positioning. We consider an estimation problem where we need to estimate $\boldsymbol{\theta}$ from the intermediate measurements $\boldsymbol{\phi}$. First, we obtain the Fisher information matrix (FIM), $\boldsymbol{J}_{\boldsymbol{\phi}}$, by assuming a Gaussian noise model for $\boldsymbol{\phi}$ using the variances given in (4.30) and (4.31). Since $\boldsymbol{\phi}$ are Gaussian distributed, they have a covariance matrix, $\mathbf{C}_{\boldsymbol{\phi}} = \boldsymbol{W}^{-1}$, where $\boldsymbol{W}$ is given in (A.19). It follows that, the FIM with respect to $\boldsymbol{\phi}$ is given by

$$\boldsymbol{J}_{\boldsymbol{\phi}} = \mathbf{C}_{\boldsymbol{\phi}}{}^{-1} = \boldsymbol{W} \tag{4.34}$$

The FIM with respect to the final estimation parameters $\boldsymbol{\theta}$ can be expressed using the chain rule,

$$\boldsymbol{J}_{\boldsymbol{\theta}} = \frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\theta}} \boldsymbol{J}_{\boldsymbol{\phi}} \left(\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\theta}}\right)^{\mathrm{T}} = \tilde{\boldsymbol{A}}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{\phi}} \tilde{\boldsymbol{A}}.$$

where $\tilde{\boldsymbol{A}}$ is given in (A.4). Finally, we find the CRLB inequality as

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\right]_{6\times 6}, \tag{4.35}$$

where $\mathbf{C}_{\hat{\boldsymbol{\theta}}}$ is the covariance matrix of the estimated parameters $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{J}_{\boldsymbol{\theta}}$ is the FIM of $\boldsymbol{\theta}$. The minimum mean squared error (MMSE) of the estimated position, $\hat{\boldsymbol{X}} = [\hat{x}\ \hat{y}\ \hat{z}]$, and velocity, $\hat{\boldsymbol{V}} = [\hat{v_x}\ \hat{v_y}\ \hat{v_z}]$, can be expressed as

$$E_{\boldsymbol{\theta}} \left\| \boldsymbol{X} - \hat{\boldsymbol{X}} \right\|^2 \geq \left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\right]_{11} + \left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\right]_{22} + \left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\right]_{33} \tag{4.36}$$

and

$$E_{\boldsymbol{\theta}} \left\| \boldsymbol{V} - \hat{\boldsymbol{V}} \right\|^2 \geq \left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\right]_{44} + \left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\right]_{55} + \left[\boldsymbol{J}_{\boldsymbol{\theta}}^{-1}\right]_{66}, \tag{4.37}$$

respectively.

4.4.2 Battery Life

In the following, we analyze the energy consumption of an NTN IoT UE which uses SPIN for the purpose of UL synchronization. We also provide an analysis for the energy consumption of an NTN IoT UE which uses a GNSS based solution. The energy consumption evaluation helps us to investigate the impact of SPIN on the UE's battery life and also to directly compare the battery life with that of a GNSS based solution. We provide the list of important notations in Table 4.4.

For the energy consumption analysis, we assume an IoT traffic model where the UE reports UL data periodically at an interval called data reporting interval. Considering the cellular communication protocol flow given in [63], the total energy consumption per data reporting interval of an NTN IoT UE which performs positioning through SPIN, is given by

$$E_{\text{SPIN}} = \underbrace{P_{\text{RX}}\left(W_{\text{SPIN,acq}} + T_{\text{SPIN,track}}\frac{T_{\text{con}}}{I_{\text{SPIN,track}}}\right)}_{\text{Sync + SPIN positioning}} + N_{\text{sat}}E_{\text{MIB-SIB}} + E_{\text{RA}} + E_{\text{UL}} + E_{\text{CDRX}} + E_{\text{IDRX}}, \tag{4.38}$$

Table 4.4: List of important notations in battery life analysis.

Notation	Meaning
P_{TX}	Transmission power
P_{RX}	Reception power
P_{LSL}	Power consumption during light sleep
P_{IDRX}	Power consumption during IDRX sleep
N_{sat}	Number of satellites visible to the UE
T_{sync}	Sync time
T_{con}	Total duration of RRC connection
$T_{\text{MIB,RX}}$	Master information block (MIB) decoding time
$T_{\text{MIB,idle}}$	Idle time in MIB
$T_{\text{SIB,RX}}$	SIB decoding time
T_{RTT}	RTT between UE and satellite
T_{msg1}	msg1 duration
$T_{\text{msg1,idle}}$	Idle time between msg1 and msg2
T_{msg2}	msg2 duration
T_{msg3}	msg3 duration
T_{msg4}	msg4 duration
$T_{\text{msg4,ACK}}$	msg4 ACK duration
$T_{\text{UL,data}}$	UL data duration
$T_{\text{UL,grant}}$	UL grant duration
$T_{\text{UL,RTT}}$	Total round trip wait time between UL data and grants
$T_{\text{DRX-INAT}}$	DRX inactivity timer
T_{CDRX}	CDRX cycle
$T_{\text{CDRX-ON}}$	CDRX-ON duration
T_{IDRX}	IDRX cycle
$T_{\text{IDRX-PO}}$	IDRX PO duration
$T_{\text{GNSS,acq}}$	GNSS acquisition duration
$T_{\text{GNSS,track}}$	GNSS tracking duration
$T_{\text{GNSS,val}}$	GNSS position validity timer

where $E_{\text{MIB-SIB}}$, E_{RA}, E_{UL}, E_{CDRX}, and E_{IDRX} are the energy consumption associated with MIB and SIB decoding, RA, UL data, CDRX, and IDRX, respectively. The above individual energy components are given by

$$E_{\text{MIB-SIB}} = P_{\text{RX}} T_{\text{MIB,RX}} + P_{\text{LSL}} T_{\text{MIB,idle}} + P_{\text{RX}} T_{\text{SIB,RX}},$$

$$E_{\text{RA}} = P_{\text{TX}}\big(T_{\text{msg1}} + T_{\text{msg3}} + T_{\text{msg4,ACK}}\big) + 2P_{\text{LSL}} T_{\text{RTT}} + P_{\text{RX}}\big(T_{\text{msg1,idle}} + T_{\text{msg2}} + T_{\text{msg4}}\big),$$

$$E_{\text{UL}} = P_{\text{TX}} T_{\text{UL,data}} + P_{\text{RX}} T_{\text{UL,grant}} + P_{\text{LSL}} T_{\text{UL,RTT}},$$

$$E_{\text{CDRX}} = P_{\text{RX}} T_{\text{DRX-INAT}} + P_{\text{RX}} \big(N_{\text{CDRX}} T_{\text{CDRX-ON}} \big) + P_{\text{LSL}} N_{\text{CDRX}} \big(T_{\text{CDRX}} - T_{\text{CDRX-ON}} \big),$$

and

$$E_{\text{IDRX}} = P_{\text{IDRX}} N_{\text{IDRX}} \big(T_{\text{IDRX}} - T_{\text{IDRX-PO}} \big) + P_{\text{RX}} N_{\text{IDRX}} T_{\text{IDRX-PO}},$$

where N_{CDRX} and N_{IDRX} represent the number of CDRX and IDRX cycles, respectively, which are decided by the values of data inactivity ($T_{\text{Data-INAT}}$) and T_{3324} timers, respectively [81, 149]. For periodic SPIN tracking, $I_{\text{SPIN,track}} = T_{\text{SPIN,val}}$ in (4.38). Note that when SPIN is used to get position fix for the purpose of UL sync, UE operations associated with establishing RRC connection are done only once per reporting interval. However, SPIN requires extra synchronization measurements from one or more satellites as it uses SSs for positioning purposes.

For the GNSS based solution, the total energy consumption per data reporting interval is given by

$$\begin{aligned} E_{\text{GNSS}} = & \underbrace{P_{\text{RX}} \big(T_{\text{GNSS,acq}} + N_{\text{GNSS}} T_{\text{GNSS,track}} \big)}_{\text{GNSS positioning}}, \\ & + N_{\text{GNSS}} (E_{\text{sync}} + E_{\text{MIB-SIB}} + E_{\text{RA}}) + E_{\text{UL}} + E_{\text{CDRX}} + E_{\text{IDRX}}, \end{aligned} \tag{4.39}$$

where E_{sync} is the energy consumption associated with cellular synchronization, given by

$$E_{\text{sync}} = P_{\text{RX}} T_{\text{sync}}, \tag{4.40}$$

and N_{GNSS} represents the number of times GNSS positioning is required in each reporting interval, given by

$$N_{\text{GNSS}} = \frac{T_{\text{con}}}{T_{\text{GNSS,val}}}. \tag{4.41}$$

We see from (4.39) that the UE operations associated with establishing cellular connection such as synchronization, RA, and MIB and SIB decoding are performed N_{GNSS} times. This is because the IoT UEs do not have the capability to simultaneously perform cellular and GNSS operations. The UE terminates RRC connection and enters RRC idle mode before GNSS positioning.

Based on the above energy consumption analysis, we now define the battery life of both SPIN and GNSS based solutions. Let the data reporting interval of the UE be I_{rep} hours and the UE battery capacity be E_{b} Wh. Then, the battery life, in years, of an NTN-IoT UE using SPIN and GNSS based solutions are given by

$$L_{\text{SPIN}} = \frac{E_{\text{b}} \times 60 \times 60}{E_{\text{SPIN}} \frac{24 \times 365}{I_{\text{rep}}}} \tag{4.42}$$

and

$$L_{\text{GNSS}} = \frac{E_{\text{b}} \times 60 \times 60}{E_{\text{GNSS}} \frac{24 \times 365}{I_{\text{rep}}}}, \tag{4.43}$$

respectively, where E_{SPIN} and E_{GNSS} are in units of J.

In the next section, we provide the numerical results for SPIN accuracy compared with the corresponding CRLB, the SPIN battery life compared with that of GNSS based solution, and the computational complexity of SPIN.

4.5 Numerical Results

In the above, we discussed the theoretical bounds of the SPIN positioning accuracy and provided the analytical equations for battery life comparison of a UE using SPIN with that of GNSS. In this section, we perform PHY and system level simulations of SPIN to obtain the achieved positioning accuracy and compare that with the corresponding CRLB. For the same evaluation settings, we also provide numerical results of battery life savings in an IoT

Table 4.5: Simulation settings for SPIN evaluation.

Parameter	Value	Parameter	Value
Positioning Evaluation Settings			
Constellation	87.5°:2400/40/20	Channel	NTN TDL-D
Coherent time	1 ms	Standard	NB-IoT
Beam configuration	Set-4	Environment	Suburban
Minimum elevation	30°	f_{s}	1.92 Msps
Satellite height	600 km	UE speed	120 km/h
k_{max} (acquisition)	10 for parallel, 30 for sequential	MCL	164 dB
		No. of UEs	1000
k_{max} (tracking)	20 for parallel, 40 for sequential	ϵ_{th}	0.0001
		$T_{\text{SPIN,proc}}$	100 ms
Battery Evaluation Settings			
E_{b}	5 Wh	T_{CDRX}	640 ms
P_{TX}	543 mW	$T_{\text{CDRX-ON}}$	100 ms
P_{RX}	37 mW	T_{IDRX}	1.28 s
P_{LSL}	8.75 mW	$T_{\text{IDRX-PO}}$	100 ms
P_{IDRX}	105 μW	T_{3324}	16 s
N_{sat}	8	$T_{\text{GNSS,acq}}$	1-30 s
T_{RTT}	26 ms	$T_{\text{GNSS,track}}$	1 s
$T_{\text{DRX-INAT}}$	100 ms	$T_{\text{GNSS,val}}$	6.4 s
$T_{\text{Data-INAT}}$	5 s	I_{rep}	2, 24 h
$T_{\text{SPIN,val}}$	5.9 s	Data size	200 bytes
Common Settings			
$W_{\text{SPIN,acq}}$	1.1 s for parallel, 5.2 s for sequential	$I_{\text{SPIN,acq}}$	215 ms
		$I_{\text{SPIN,track}}$	296 ms
$W_{\text{SPIN,track}}$	296 ms for parallel, 2.1 s for sequential	$T_{\text{SPIN,acq}}$	215 ms
		$T_{\text{SPIN,track}}$	40 ms

UE which uses SPIN instead of GNSS based solution. Finally, we show the computational complexity of SPIN in terms of the number of operations and memory requirements.

4.5.1 Simulation Settings

For both positioning accuracy and battery life evaluations, we choose the evaluation parameters and their values from the relevant 3GPP technical documents and reports [63, 81, 150]. We list the parameters and their values we use for our evaluations in Table 4.5. We adopt all the time parameters in the protocol flow and their values from [63], and hence we do

not repeat them in Table 4.5. To achieve global coverage, we select a satellite constellation of sufficient size which is comparable with that of currently planned and/or deployed LEO constellations such as Starlink [64, 65] and Amazon-Kuiper [65]. To evaluate the worst case which corresponds to deep coverage, we perform both SPIN accuracy and battery life evaluations at an MCL of 164 dB. Further, we choose set-4 satellite beam configuration for which the link budget matches with 164 dB MCL [63, 151]. For the sake of conciseness, we provide SPIN evaluation for only NB-IoT standard. Nevertheless, the evaluations can be extended in a similar way to LTE-M standard.

We apply the NTN channel models recommended by 3GPP [26]. We choose the parameters corresponding to a suburban environment, which is an ideal application scenario where NTN support is critical. The channel model includes small scale fading, which emulates the multipath effect mainly caused by the obstacles close to the UE, and the Doppler spread due to the UE motion [26]. The SNR is obtained from the NTN IoT link budget [25, 63] as

$$\gamma = \frac{P_{\text{EIRP}} G}{B k_{\text{B}} T \alpha_{\text{fs}} \alpha_{\text{atm}} \alpha_{\text{sh}} \alpha_{\text{sc}} \alpha_{\text{pol}} \alpha_{\text{add}}}, \tag{4.44}$$

where P_{EIRP} is the effective isotropic radiated power (EIRP) from the satellite, G/T is the antenna-gain-to-noise-temperature of the UE, B is the bandwidth of the synchronization signal, k_{B} is the Boltzmann constant, α_{atm} is the atmospheric path loss, α_{sh} is the shadowing margin, α_{sc} is the scintillation loss, α_{pol} is the polarization loss, α_{add} includes all other additional losses, and α_{fs} is the free space path loss [26] given by

$$\alpha_{\text{fs}} = (4\pi f_{\text{c}} d/c)^2, \tag{4.45}$$

where d is the distance between the satellite and the UE.

The number of synchronization measurements that a UE can perform in an acquisition

or tracking window depends heavily on the network implementation. This includes factors such as the partition of available network bandwidth into intra- and inter-frequency cells and the deterministic time offset between the cells. Therefore, we assume two extreme cases for our evaluations. First, we evaluate the worst case where the UE can only sequentially acquire synchronization measurements from one cell to another. The worst case corresponds to a scenario where all the available synchronization signals in the SPIN window belong to inter-frequency cells. However, it might be possible for a UE with adequate processing capability to measure synchronization signals from multiple intra-frequency cells in the acquisition bandwidth simultaneously, if they are time multiplexed. Therefore, we also evaluate the best case, i.e., with parallel measurements, where the UE can simultaneously acquire all available synchronization measurements in a SPIN window. SPIN takes longer time to get position fix when acquiring SSs sequentially. Therefore, for sequential method, we choose acquisition and tracking windows longer than that of parallel method.

For battery life comparison, we use the same values of time parameters which we used for SPIN accuracy evaluation. For estimating the battery life of a UE using GNSS based solution, we choose GNSS acquisition and tracking durations from 3GPP technical report [63]. Without loss of generality, we use the term GNSS throughout our evaluations to indicate GPS, which is the widely used GNSS service. Finally, to maintain low computational complexity along with reasonably high accuracy, we set $T_{\text{SPIN,proc}}$ to 100 ms.

4.5.2 Methodology

We perform the SPIN accuracy evaluation in two steps: evaluation of i) TOA and FOA estimation using PHY simulation, ii) position and velocity estimation using system level simulation, and compare with the corresponding CRLBs.

In the first step, using MATLAB LTE and 5G Toolbox, we generate SSs, pass them

through the NTN channel models, and estimate TOA and FOA. To account for the non-line-of-sight (NLOS) paths in a suburban environment, we apply a penalty by disregarding NLOS signals based on the LOS-NLOS probability given in [26]. In our evaluations, we assume that the UE filters out the NLOS signals, which can be done based on the signal strength as explained in [33].

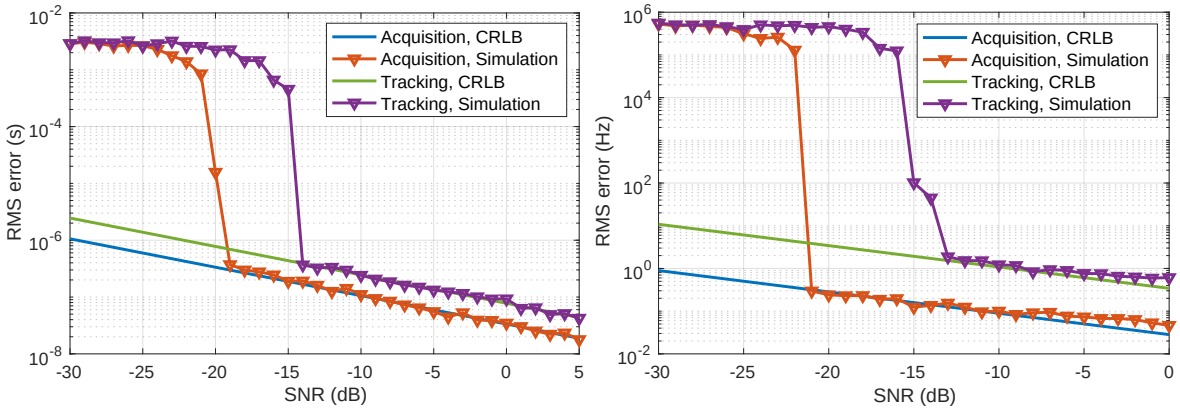
In the second step, we conduct system level simulations of SPIN acquisition and tracking, by considering a satellite constellation and a set of UEs located on the earth. Note that the evaluation models in the existing work on NTN NR positioning [33] do not consider the impact of true UE location on the achieved accuracy. In reality, the achievable accuracy varies significantly depending on the true UE location. To this end, to fully capture the impact of all locations, we consider uniformly spaced 1000 UEs on the earth. However, for a near-polar Walker-star constellation, over-crowding of satellites happens at the poles. In reality, some of the satellites turn off their beams in a coordinated manner to avoid high interference at those regions. Moreover, the probability that a UE might be located near the poles is negligible. Therefore, to prevent the results from skewing towards very high accuracy due to the above exceptional scenario, we exclude the regions close to both the poles. Furthermore, our simulation model accounts for the significant variation of SNR during the satellite fly-by, thus ensuring accurate simulation results. To this end, we calculate SNR using (4.44) for each measurement and apply noise variance separately, instead of setting a fixed variance for the noise in all measurements.

4.5.3 Positioning Evaluation

Fine Time and Frequency Offset Estimation

We first evaluate TOA and FOA estimation accuracy using PHY simulation of PSS and SSS in NTN channel for the range of SNR given by the NTN IoT link budget. We compare these

simulation results with the corresponding CRLB in Figure. 4.9. The SNR for 164 dB MCL ranges from -12 to -7 dB when the satellite moves from the minimum to the maximum elevation angles of $30°$ to $90°$, respectively. The results in Figure. 4.9 show that the TOA and FOA estimated using correlation over the chosen acquisition and tracking durations followed by fine resolution curve fitting approach the CRLB for the concerned range of SNR. Only under very low SNR, both the TOA and FOA estimations fail due to an unavoidable anomalous behavior [152]. Additionally, at very high SNR, the TOA and FOA estimations tend to saturate thus creating gaps with CRLB. This saturation is due to the estimation bias caused by the small scale fading in the channel.



(a) (b)

Figure 4.9: (a) TOA (b) FOA estimation errors compared with corresponding CRLBs.

Position and Velocity Estimation

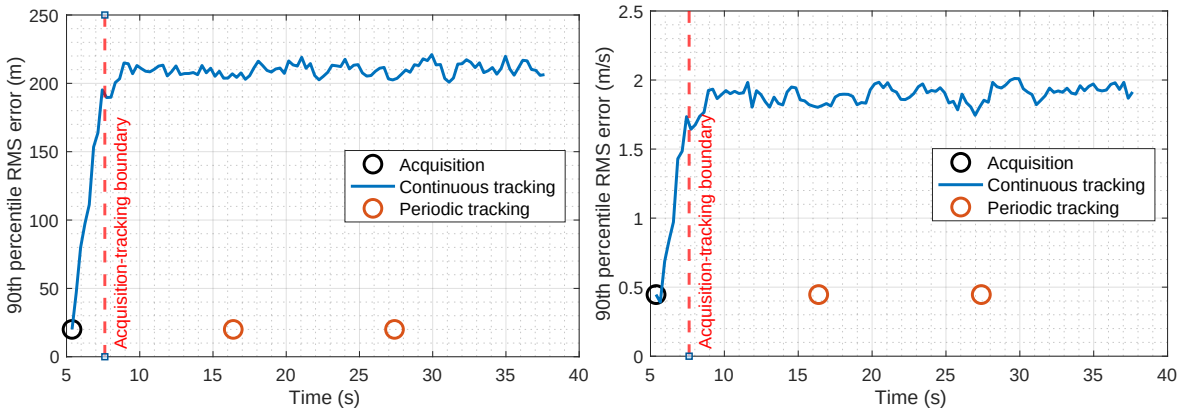
In the above, we showed that the TOA and FOA estimations in SPIN achieve the corresponding CRLBs for the concerned range of SNR. Therefore, for the following SPIN acquisition and tracking simulations, we assume that the TOA/FOA measurements have Gaussian distributions with mean and variance equal to the true TOA/FOA values and CRLB, respectively. For uniformly spaced 1000 UEs on the earth, we show the 90th percentile of RMS error in

SPIN position and velocity and their comparison with the corresponding CRLBs, in Table 4.6. We see that in all cases, SPIN results are close to the position and velocity CRLBs. As given in Table 4.5, we set the number of TWLS iterations to small values to have low computational complexity. However, if we run SPIN for more TWLS iterations, the position and velocity accuracy approach the corresponding CRLBs, but at the expense of a higher computational complexity. It should be noted that SPIN acquisition achieves the target requirements shown in Figure. 4.4 for a wide range of error budget covering low to very high oscillator errors. On the other hand, we choose reasonably short SPIN tracking window such that the results just meet the target requirements for $80 - 20\%$ error budget. However, to get higher accuracy, SPIN tracking window can be cautiously increased, provided that the assumption of UE velocity remaining constant during the entire window remains valid.

Table 4.6: 90th percentile RMS error of position and velocity.

Measurement Type	Positioning Step	Position (m)		Velocity (m/s)	
		CRLB	Simulation	CRLB	Simulation
Parallel	Acquisition	17.4	20.6	0.1	0.8
	Tracking	188.4	207.3	1.7	4.9
Sequential	Acquisition	15.9	18.2	0.1	0.4
	Tracking	193.9	212.6	2.0	3.1

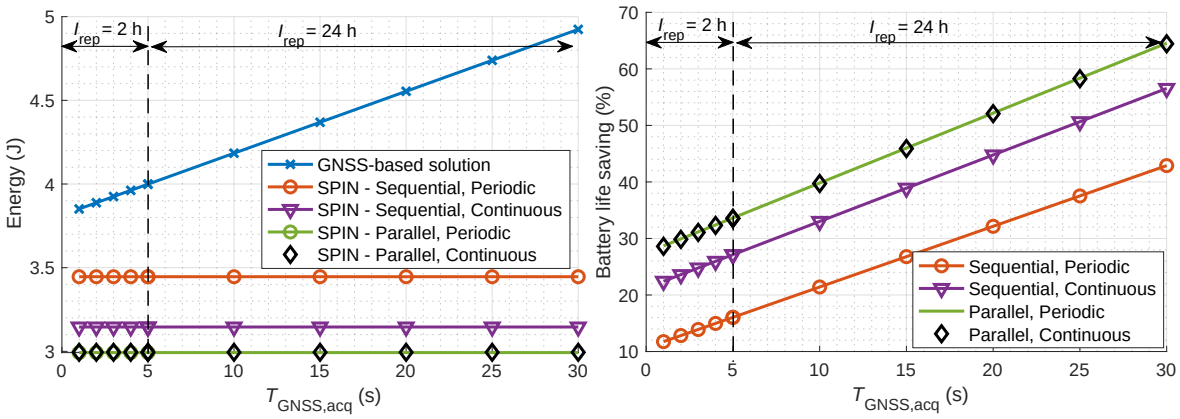
In Table 4.6, we showed SPIN accuracy results for stand-alone acquisition and tracking operations by configuring appropriate durations. To show the real-time accuracy variation starting from SPIN acquisition until a few loops of SPIN tracking, we have also plotted the 90th percentile of RMS error against time in Figure. 4.10. For the sake of conciseness, we show the results for only sequential measurements case here. However, we see a similar trend in the parallel measurements scenario too. For continuous tracking, the accuracy of both position and velocity initially reduce over time before finally saturating to a level. This is because SPIN maintains a sliding tracking window which discards old measurements while



(a) (b)

Figure 4.10: 90th percentile RMS error of (a) position and (b) velocity estimation vs. time, for sequential measurements case.

including new measurements. The TOA and FOA measurements performed during the initial acquisition are superior than the tracking measurements due to the longer acquisition duration and hence give higher accuracy. Until the acquisition-tracking boundary as shown in Figure. 4.10, the set of measurements include at least one measurement from the initial acquisition. However, afterwards, the accuracy remains approximately same since the following tracking measurements are performed with a fixed duration and interval. We observe that, SPIN maintains position and velocity RMS error below 215 m and 2 m/s, respectively, by performing measurements every 296 ms. On the other hand, in periodic tracking, SPIN gets a position fix every time when $T_{\text{SPIN,val}}$ expires. In our simulations, we set $T_{\text{SPIN,val}}$ based on the value of UE speed. Since the UE is moving, the error relative to the previous estimated position increases over time. In this regard, we set $T_{\text{SPIN,val}}$ equal to the time taken for the position error to reach 215 m, which is the allowed position error for $80 - 20\%$ error budget in Figure. 4.4. For continuous tracking, the values of tracking window, interval, and duration used in the simulations are specified in Table 4.5. However, periodic SPIN tracking measurement parameters need to be similar to that of acquisition measurement as they require high



(a) (b)

Figure 4.11: (a) Comparison of energy consumption per reporting interval of SPIN with GNSS based solution. (b) Battery life saving with SPIN.

accuracy. Therefore, for periodic tracking, we set $W_{\text{SPIN,track}}$, $I_{\text{SPIN,track}}$, and $T_{\text{SPIN,track}}$ equal to the values of $W_{\text{SPIN,acq}}$, $I_{\text{SPIN,acq}}$, and $T_{\text{SPIN,acq}}$ given in Table 4.5, respectively. The circles in Figure. 4.10 represent the accuracy achieved by SPIN acquisition followed by periodic tracking. Since we set the tracking window long enough to get highly accurate position, the periodic tracking accuracy remains valid until the expiry of $T_{\text{SPIN,val}}$.

As already mentioned in Section 4.2.2, UE velocity related Doppler offset in the UL is taken care by the network. However, given the high accuracy of UE velocity estimation in SPIN, the IoT UE can additionally pre-compensate the UL signal for the UE Doppler.

4.5.4 Battery Life Comparison

First, we compare the total energy consumption per reporting interval of an IoT UE which uses SPIN with that of a UE using GNSS, in Figure. 4.11(a). The time to get a GNSS position fix varies based on a range of factors including the type of chipset, the positioning interval, and the visibility to GNSS satellites. To this end, for $I_{\text{rep}} = 2$ h, we vary the GNSS acquisition duration, $T_{\text{GNSS,acq}}$, from 1 to 5 s where GNSS performs a hot-start positioning [63].

On the other hand, GNSS performs a cold-start positioning when $I_{\text{rep}} = 24$ h, and hence we vary $T_{\text{GNSS,acq}}$ from 5 to 30 s [63]. We see in Figure. 4.11(a) that, for the entire range of $T_{\text{GNSS,acq}}$, GNSS based solution consumes more energy than SPIN. This is mainly because of repeated termination and re-establishment of RRC connection associated with the GNSS positioning gaps. In addition, we see that the energy consumption in GNSS based solution increases with increase in $T_{\text{GNSS,acq}}$, since longer duration of GNSS reception is performed for higher $T_{\text{GNSS,acq}}$. On the other hand, the energy consumption in SPIN remains constant for all combinations of settings, since it is independent of $T_{\text{GNSS,acq}}$. SPIN consumes more energy when it performs sequential measurements than when it performs parallel measurements. This is because SPIN performs reception of SSs for longer duration when it performs measurements sequentially. For the sequential measurement case, periodic SPIN tracking consumes more energy than continuous SPIN tracking. The reason is that we set the periodic tracking duration to a large value to get high accuracy which should remain valid until the expiry of $T_{\text{SPIN,val}}$. However, continuous tracking maintains the accuracy within an acceptable limit by making use of short and frequent time gaps, thus consuming less energy. This changes for the parallel measurement case, in which the energy consumptions for periodic and continuous tracking are approximately the same. This is because of the same effective reception duration of SS measurements for the chosen settings.

Finally, we show the battery life saving associated with SPIN over the GNSS-based solution in Figure. 4.11(b). For the best case where SPIN performs parallel measurements, it saves around $29 - 64\%$ of the UE's battery life. Even in the worst case where sequential measurements are performed, SPIN offers a significant battery life saving which ranges from 12 to 43%

4.5.5 Computational Complexity

We show the number of basic arithmetic operations associated with SPIN acquisition and tracking for both parallel and sequential measurements scenarios in Table 4.7. The number of operations in TWLS includes all the computations performed during the entire iterative process until the convergence. Using the number of operations and $T_{\text{SPIN,proc}}$, we calculate the computational complexity in terms of million operations per second (MOPS). For SPIN acquisition, the MOPS associated with parallel measurements case is higher than that of sequential measurements. Therefore, for an NB-IoT UE with low complexity, SPIN acquisition with sequential measurements is more suitable. On the other hand, an LTE-M UE with higher computational capacity can perform SPIN acquisition with parallel measurements which can achieve sufficient accuracy in shorter measurement time. The evaluations also show that MOPS required for both sequential and parallel measurement cases in SPIN tracking are lower than that of acquisition since tracking is performed with less number of measurements. The most computationally demanding cases in NB-IoT are data channel processing and SS detection which require 18.5 and 30 MOPS, respectively [56]. We see that the MOPS in both parallel and sequential cases of SPIN are much lower than the MOPS required by the existing computationally expensive cases. For the above evaluation, we did not consider the operations associated with the coarse synchronization step since SPIN directly reuses the results from the cellular operation. As the cellular UE has the computational capacity to perform correlation on signals at same sampling rate and comparable lengths as part of fine synchronization, we also do not analyze the complexity associated with the subsequent SPIN correlation operations.

We also computed the memory requirement in the UE for performing SPIN, which is mainly decided by the size of the matrices involved in 2-WLS and TWLS steps. The highest memory requirement is for SPIN parallel acquisition which is around 28 kB. On the other

Table 4.7: Computational complexity of SPIN.

Positioning Step	Number of Operations			
	Parallel Acquisition	Sequential Acquisition	Parallel Tracking	Sequential Tracking
Curve Fitting	640	384	128	112
2-WLS	500702	153144	-	-
TWLS	1294950	1095030	77620	124280
MOPS	17.96	12.48	0.78	1.24

hand, sequential SPIN acquisition takes only 1 kB of memory. This is not significant for IoT UEs which usually have a random access memory (RAM) size of 256 kB [153, 154].

It should be noted that, in the battery life analysis, we did not consider the power consumption associated with the computations related to SPIN or cellular operation. For even a relatively less efficient processor with an efficiency of 144 MOPS/mW [155, Ch. 5], the power consumption associated with computations in SPIN parallel acquisition, the most computationally demanding case, is only 0.1 mW. However, for a superior processor efficiency of 970 MOPS/mW, the power consumption is only 18 μW. Therefore, it is meaningful to assume that the SPIN computations have negligible impact on the battery life of the UE.

4.6 Discussion

In this section, we provide brief qualitative analysis of SPIN in additional scenarios which are not evaluated in this chapter.

4.6.1 LTE-M

In the above, we evaluated SPIN for NB-IoT standard where it uses NPSS and NSSS for positioning. In a similar way, we can evaluate SPIN for the LTE-M standard by replacing NPSS and NSSS with PSS and SSS, respectively. In contrast to NB-IoT, LTE-M is meant to sup-

port lower coverage level and hence we can consider an MCL of 154 dB, which corresponds to set-3 or set-2 beam configurations [63, 151]. For the LTE-M standard, which uses PSS and SSS with a bandwidth higher than that of NPSS and NSSS, we observe that the SPIN accuracy is much higher than that of NB-IoT.

4.6.2 Reduced Constellation Size

The satellite constellation chosen for our evaluation gives visibility to more than one satellite at any instant of time to every UE. Nevertheless, SPIN also works in those scenarios where there is limited visibility, e.g., a constellation consisting of 510 LEO satellites at an altitude of 600 km, which is the minimum constellation size for global coverage [122]. However, in a limited visibility scenario, SPIN takes a longer duration to get an accurate position fix.

4.6.3 Indoor Scenario

We considered outdoor UEs for the SPIN evaluations in this chapter. However, SPIN also works in soft indoor cases, where an additional loss needs to be included in the link budget, e.g., 9 dB [151]. However, for deep indoor cases which suffer from severe outdoor-to-indoor (O2I) losses, the link budget as per the satellite reference parameters defined by 3GPP [25, 63] is insufficient. Alternatively, if LEO satellites with higher antenna gain are deployed, successful UL and DL communication are possible for deep indoor cases [39]. In such a scenario, SPIN can operate successfully in an NTN UE and can provide position fix to resolve UL synchronization issues and to serve LBS.

4.6.4 Other Environments

For the evaluations in this chapter, we considered an NTN channel in a suburban environment with LOS probability as defined in [26]. In an open-sky environment, e.g., when the UEs are

mounted on aircrafts or boats, we can assume an AWGN channel with 100% LOS possibility. On the other hand, in an urban or dense urban environment, the LOS probability is lower than that of a suburban scenario. Nevertheless, our evaluations can be directly extended to the above cases. We see that the accuracy in open-sky environment is better than the suburban scenario which we evaluated in this chapter. However, the accuracy will be lower in an urban environment due to the lower LOS probability. This is not a major concern since the IoT UEs located in urban environment are more likely to be served by the terrestrial cellular BSs.

4.7 Summary

In this chapter, we designed an algorithm called SPIN which enables self-positioning in NTN IoT UEs for the purpose of solving the UL synchronization problem and to provide LBS. SPIN performs positioning by utilizing TDOA and FDOA measurements on the existing DL SSs, thus requiring no network modification. SPIN adopts state-of-the-art signal processing techniques to ensure that the positioning accuracy achieves the theoretical bounds. The numerical results show that SPIN meets the target accuracy required for the UL sync problem while also achieving the CRLB. Our analysis also shows significant battery life savings when a GNSS-based solution is replaced with SPIN. Further, the computational complexity associated with SPIN is much lower than that of the existing cellular operations. The positioning accuracy, the battery life analyses, and the low complexity prove that SPIN is an appealing solution for the UL sync problem in NTN IoT UEs.

Chapter 5

Enhanced HARQ in NTN Cellular IoT

5.1 Introduction

As discussed in Chapter 1, one of our research goals is to solve the problems in NTN with an objective to achieve global network coverage for C-IoT UEs. In the previous chapter, we designed SPIN to solve the UL synchronization problem which results from the large Doppler offset in NTN. In this chapter, we focus on the impact of the extended RTT of bidirectional signals caused due to the increased propagation distance in NTNs on the network throughput. To this end, we propose methods for enhancing resource scheduling tailored for C-IoT and mMTC systems to operate in NTNs, which are collectively referred to as *IoT-NTN*. Typically, the achievable data rate may not be the primary target performance indicator to focus on while designing C-IoT and mMTC systems. One of the reasons for this is the greater resource consumption (e.g., extended bandwidth requirement) and higher transmit power associated with achieving increased throughput. However, we show in this work that the network throughput can be improved without demanding additional resources or increasing transmit power, but instead by efficiently scheduling the HARQ processes. As a result, the IoT-NTN can support a larger number of mMTC devices, which is critical in NTNs due to the significantly larger cell size compared to conventional terrestrial networks. Toward this end, we exploit the lengthy RTT in NTN links to overlap bidirectional signals in the air. We further utilize the frequency-division full-duplex nature of BSs to allow simultaneous transmission and reception at the satellites and BSs.

The rest of this chapter is organized as follows. We first provide a qualitative comparison of our solutions with the prior art in Section 5.2 and then introduce the system model in Section 5.3. We use the metrics and key performance indicators from Section 5.3 to highlight the shortcomings of relevant prior works in tackling the issue we addressed in Section 1.2.3. We propose our solutions in Section 5.4, which we evaluate and present their performance results in Section 5.5. We summarize the chapter in Section 5.6. We also include a list of important notations and their meanings in Table 5.1.

5.2 Qualitative Comparison with Prior Art

In this section, we provide qualitative comparisons to demonstrate how our proposed solutions are more suitable than those in the existing art discussed in Section 1.2.3.

Low Target iBLER

We discussed in Section 1.2.3 that when a large iBLER is used in NTNs, it does result in increased stop-and-wait gaps. Therefore, our solution, aimed precisely at reducing these stop-and-wait gaps with smart scheduling, presents a method to use a high iBLER for battery life optimization for IoT-NTN devices, while at the same time also increasing UE battery life by enabling a higher target iBLER. Nevertheless, note that our proposed solution is also compatible with designs that use a sub-optimal low iBLER transmission, but with reduced throughput gains.

Disabled HARQ

In Section 1.2.3, we discussed the importance of HARQ based communication, especially at the MAC layer, for lower latency, better battery life, and reduced jitters. We also pointed out the drawbacks associated with disabling HARQ. However, our solution does not demand

Table 5.1: List of important notations used in this chapter.

Notation	Meaning
η_{SUF}	SF utilization factor
N_{data}	Number of DL or UL data SFs
N_{rep}	Number of repetitions of TB
N_{HC}	Lengths of HARQ cycle in units of SFs
N_{A2G}	ACK processing delay in units of SFs
R	Data rate
n_{TB}	Transport block size in bits
t_{TB}	Transport block duration in seconds
n_{DD2A}	Delay between the DL data and the associated ACK in units of SFs
n_{UG2D}	Delay between the UL grant and the associated UL data in units of SFs
N_{DG2D}	Delay between the DL grant and the associated DL data in units of SFs
$n_{\text{DD2A, min}}$	Minimum mandatory DD2A in units of SFs
$n_{\text{UG2D, min}}$	Minimum mandatory UG2D in units of SFs
n_{bundle}	Number of ACKs bundled within one TTI
N_{HARQ}	Number of HARQ processes
T_{RTT}	RTT in seconds
$n_{\text{rep,data}}$	Number of data block repetitions
$n_{\text{rep,PDSCH}}$	Number of PDSCH repetitions
$n_{\text{rep,PDCCH}}$	Number of PDCCH repetitions
$n_{\text{rep,PUCCH}}$	Number of PUCCH repetitions
$n_{\text{rep,PUSCH}}$	Number of PUSCH repetitions
N_{switch}	Switching delay in units of SFs
N_{TBPHC}	Number of TBs scheduled in one HARQ cycle
γ	Signal-to-noise ratio
$P_{\text{EIRP,UE}}$	Effective isotropically radiated power of UE
k_{B}	Boltzmann constant
G/T	Satellite antenna-gain-to-noise-temperature
B	Signal bandwidth
f_{c}	Carrier frequency
α_{fs}	Free space path loss
α_{atm}	Atmospheric loss
α_{sh}	Shadow fading margin
α_{sc}	Scintillation loss
α_{pol}	Polarization loss
α_{add}	Additional loss

disabling the HARQ process. Therefore, our solutions are superior to those methods which require HARQ disabling.

Increasing the Number of HARQ Processes

In Section 1.2.3, we explained that increasing the N_{HARQ} to fill the stop-and-wait gaps is not ideal for IoT-NTN systems. Nevertheless, in this chapter, we show that a cautious increase without the use of MTBG and bundled ACK may be acceptable, especially in the uplink for IoT-NTN applications, to obtain tangible increase in throughput. However, we demonstrate that this increase must be accompanied with network modifications that we propose in this work to allow flexible HARQ scheduling to extract meaningful throughput gain.

With this backdrop, we propose HARQ scheduling solutions in the following section, that do not require disabling HARQ processes or demand any mandatory increase in N_{HARQ}.

5.3 System Model

We use the same system model as illustrated in Figure. 4.1. In this work, we consider the satellite to be in a LEO revolving around the earth in a circular orbit. But our study is directly applicable to higher satellite altitudes of MEO and GEO satellites and also for lower altitudes of high-altitude platform stations (HAPS) and unmanned aerial vehicles (UAVs). Our study is agnostic to the data processing/forwarding architecture, and thus supports both transparent or bent-pipe payloads and regenerative ones. In the bent-pipe architecture, the satellite acts primarily as a relay node that forwards packets to a gateway that eventually connects to the core network. On the other hand, BSs that support regenerative payloads contain an on-board processor to process the payload before connecting to a gateway either directly or via another satellite over an ISL. For the purposes of our study, the architecture only impacts the RTT of the message, and further insights on the architectures are available at [25, 156].

Table 5.2: RTTs for LEO NTNs.

Satellite Altitude	Payload Type	Min. RTT (ms)	Max. RTT (ms)
600 km	Regenerative	4	13
600 km	Transparent	8	26
1200 km	Regenerative	8	21
1200 km	Transparent	16	42

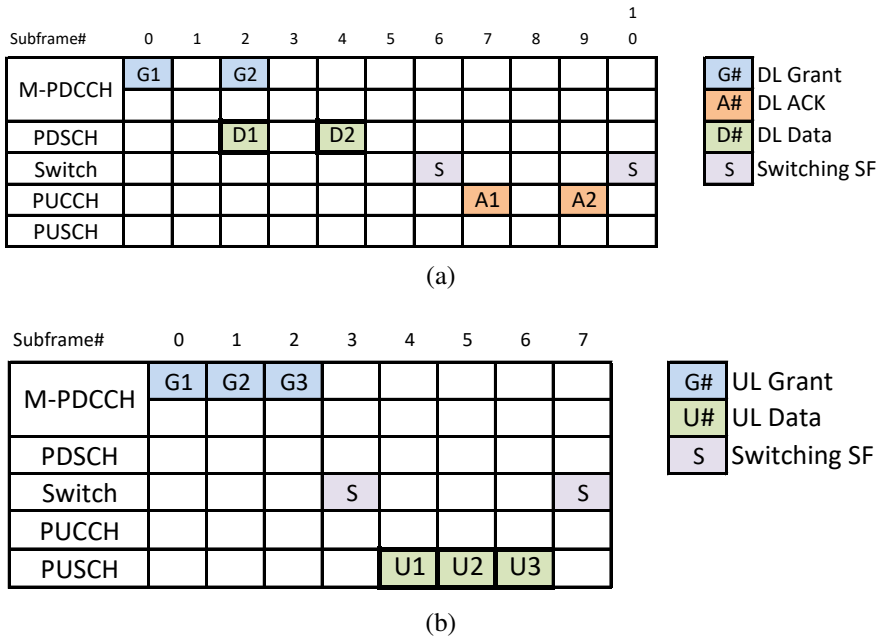


Figure 5.1: Timing diagrams for LTE-M operation in (a) the DL and (b) the UL.

Every satellite steers multiple beams toward the earth with multiple beam footprints on earth as shown in Figure. 4.1. The beams can either be earth-fixed, i.e., steerable beams, or moving, i.e., non-steerable beams. While our work is applicable for both types of beam steering configurations, the satellite configuration aspect that impacts our solution is the RTT of the satellite link. The beam footprint of each beam determines the RTT range of the satellite link [25]. We present the RTTs of IoT-NTN links for different LEO altitudes for both types of payload architectures in Table 5.2. The minimum and the maximum RTTs correspond to the maximum ($90°$) and minimum ($10°$) beam elevation angles, respectively.

We borrow the UL and DL communication mechanism between the UE and BS for IoT-

NTN directly from the LTE-M and NB-IoT specifications [11,93]. Our reasoning behind this is two-fold. First, LTE-M and NB-IoT are the industry chosen standards for enabling low-power wide area networks. Several works have been presented in the past to demonstrate the benefits of using these 3GPP standards, e.g., [157,158]. The use of LTE-M and NB-IoT provides the advantage of reusing existing infrastructure and operating on licensed bands. These help in providing low-cost, stable, reliable, and predictable performance across application scenarios. Second, similar to using the NR standard as the starting point for eMBB over NTN [25, Sec. 6], IoT-NTN standardization activities in 3GPP have agreed to build on the existing LTE-M and NB-IoT standards to expand them into the NTN realm [63]. By showing the effectiveness of our proposed solution for LTE-M and NB-IoT, we aim to demonstrate that integrating our cost-efficient method into legacy systems is practically feasible. According to both the LTE-M and NB-IoT standards, UL and DL data bits are grouped into TBs of varying sizes for transmission [93]. The TBS is dependent on the adaptive MCS chosen based on the operating conditions and target block error rate (BLER). The transmission time interval (TTI), which is the time spanned by one unit of transmission corresponds to one SF of 1 ms duration, during which one or more TBs are transmitted. A timing diagram of the DL and UL transmissions for LTE-M are shown in Figs. 5.1(a) and (b), respectively. Figs. 5.1(a) and (b) demonstrate the HARQ-based DL and UL communication [159, Ch. 10], which is used in both LTE-M and NB-IoT. The need for using a HARQ-based design in cellular low-power wide area network technologies, such as LTE-M and NB-IoT, has been extensively shown in the past, e.g., [160].

A DL data TB on the physical downlink shared channel (PDSCH) is preceded by a corresponding single TB grant (STBG) sent on the MTC physical data control channel (M-PDCCH). For the case of NB-IoT, the grants are sent on the NB-IoT PDCCH (N-PDCCH). Henceforth, we drop the prefix for brevity and refer to it only as PDCCH. One grant may

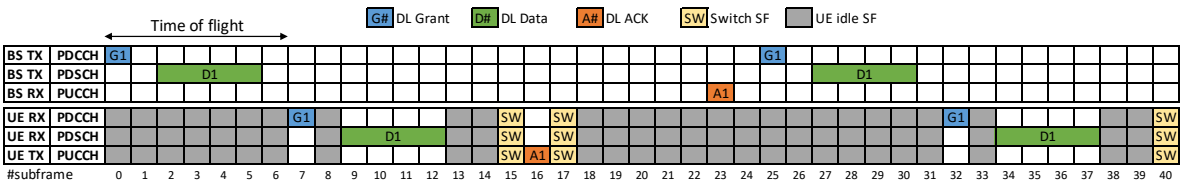


Figure 5.2: Timing diagram of legacy HARQ operation for HD-FDD in NTN DL.

also configure multiple TBs, and such grants are referred to as multiple TB grants (MTBGs). For every data TB received, the UE responds with an ACK TB on the physical uplink control channel (PUCCH). The ACK TB can either acknowledge one or more TBs using unbundled or bundled ACKs, respectively. UL transmissions are analogous to DL, where UL data TBs that are configured by UL grants (UGs) in an STBG or MTBG fashion are transmitted on the physical uplink shared channel (PUSCH) by the UE. For this work, we consider low-cost IoT UEs that operate in HD-FDD manner, which is the industry preferred design for cost and complexity reduction [57, 82]. As a result, the UE uses one or more SFs to switch between transmission and reception modes.

The timing diagrams shown in Figure. 5.1 are identical at both the BS and UE for terrestrial networks due to the negligible signal propagation delay. However, in NTNs, the timing shown in Figure. 5.1 is valid at the UE, while the DL transmissions are sent in advance from the BS by half the RTT. Similarly, the UL packets reach the BS also after half the RTT. We show an example of this in the timing diagram of Figure. 5.2. We see that the overhead caused by the processing delays and the time of flight between the UE and BS results in several idle SFs at the UE-end, which significantly reduce the achievable throughput.

To quantify the data rates obtainable, we define the concepts of HARQ cycle and SF utilization factor (SUF). We define one HARQ cycle as the total duration of time during which the UE receives on the PDCCH and/or PDSCH and the duration of time it transmits on PUCCH or PUSCH. A HARQ cycle also includes the switching SF(s). For example, Figure. 5.1 shows one HARQ cycle for both DL and UL, where the lengths of a HARQ

cycle, N_{HC}, in units of SFs are $N_{\text{HC}} = 11$ and $N_{\text{HC}} = 8$, respectively.

Next, we define the SUF, η_{SUF}, as

$$\eta_{\text{SUF}} = \frac{N_{\text{data}}}{n_{\text{rep}} N_{\text{HC}}}, \tag{5.1}$$

where N_{data} is the total number of SFs occupied by UL or DL data TBs and n_{rep} is the number of repetitions used for each TB, i.e., the number of times each TB is repeated including its first transmission. Note that the TB repetitions introduce redundancy to improve the error rates, and are not the same as HARQ re-transmissions. The value of n_{rep} is chosen based on the adaptive MCS used by the UE. HARQ re-transmissions, on the other hand, are repeated transmissions of one or more repetitions of a TB when the TB is not successfully acknowledged. A detailed explanation of the HARQ operation can be found extensively in the literature, e.g., [159, Ch. 10].

Eq. (5.1) shows that SUF indicates the proportion of time spent on transmitting the payload as opposed to transmission overheads. We therefore compute the useful data rate, R, as

$$R = \eta_{\text{SUF}} \frac{n_{\text{TB}}}{t_{\text{TB}}}, \tag{5.2}$$

where n_{TB} and t_{TB} are the TBS in bits and time spanned by one TB, respectively. Therefore, for a given TBS corresponding to the adaptive MCS chosen, we can maximize the data rate by maximizing the SUF, i.e., by reducing the transmission overheads. Since higher values of RTT can increase N_{HC}, it can be seen from (5.1) that it thus deteriorates SUF and the useful data rate.

5.4 Enhanced HARQ for IoT NTN

The value of N_{HARQ} required to fill up all the stop-and-wait gaps in its entirety is

$$N_{\text{HARQ}} \geq \frac{T_{\text{RTT}}}{t_{\text{TB}}}, \tag{5.3}$$

where T_{RTT} is the RTT in seconds. From Table 5.2, we can observe that N_{HARQ} must be increased to up to 42 under typical operating conditions where one TB spans 1 SF. The current LTE-M and NB-IoT specifications only allow a maximum of 8 and 2 HARQs, respectively [11, 93]. This increase in N_{HARQ} is impractical for low cost mMTC and C-IoT UEs, where a high N_{HARQ} introduces increased complexity due to the required size of the soft-buffer. Furthermore, it also demands corresponding increments in the size of the DCI bits.

For the case of IoT-NTN, which uses few PRBs and operates under high path loss environments, multiple repetitions of the TB are often required to achieve reliable communication. Therefore, the required N_{HARQ} is

$$N_{\text{HARQ}} \geq \frac{T_{\text{RTT}}}{n_{\text{rep,data}} t_{\text{TB}}}, \tag{5.4}$$

where $n_{\text{rep,data}}$ is the number of data block repetitions, i.e., $n_{\text{rep,data}} = n_{\text{rep,PDSCH}}$ or $n_{\text{rep,data}} = n_{\text{rep,PUSCH}}$, for indicating repetitions on the PDSCH or the PUSCH, respectively. In Section 5.5, we present the results from a comprehensive link-level simulation evaluation to show that IoT-NTN does not require any further increase in N_{HARQ} for MTC applications, and up to $N_{\text{HARQ}} = 4$ for NB-IoT applications to achieve reliable communications while also filling the stop-and-wait gaps.

The issue with the current HARQ implementation is that the delay between the DL data and the corresponding ACK, n_{DD2A}, and similarly in the uplink, the delay between the UL

grant and UL data, n_{UG2D}, are fixed. Therefore,

$$i_{\text{ACK},j} = i_{\text{DLdata},j} + n_{\text{DD2A}} + 1, \text{ for DL} \tag{5.5}$$

$$i_{\text{ULdata},j} = i_{\text{UG},j} + n_{\text{UG2D}} + 1, \text{ for UL}, \tag{5.6}$$

where $i_{\text{ACK},j}$, $i_{\text{DLdata},j}$, $i_{\text{ULdata},j}$, and $i_{\text{UG},j}$ are the positions in SFs of the ACK for the jth TB, the last SF of the jth DL TB, the first SF of the jth UL TB, and the last SF of the jth UG, respectively. When $n_{\text{rep}} > n_{\text{DD2A}}$, $i_{\text{ACK},j}$ overlaps with $i_{\text{DLdata},k}$ for $k > j$ or the switching SF. Similarly, when $n_{\text{rep}} > n_{\text{UG2D}}$, $i_{\text{DLdata},k}$ for $k > j$ overlaps with $i_{\text{ULdata},j}$. Thus, the current HARQ configuration can only support one TB per HARQ cycle when $n_{\text{rep}} > n_{\text{DD2A}}, n_{\text{UG2D}}$. We can re-write (5.1) under such an operation as

$$\eta_{\text{SUF,DL}} = (n_{\text{rep,PDCCH}} + n_{\text{rep,PDSCH}} + n_{\text{rep,PUCCH}} + N_{\text{DG2D}} + n_{\text{DD2A}} + N_{\text{switch}})^{-1} \tag{5.7}$$

and

$$\eta_{\text{SUF,UL}} = (n_{\text{rep,PDCCH}} + n_{\text{rep,PUSCH}} + n_{\text{UG2D}} + N_{\text{switch}})^{-1}, \tag{5.8}$$

for DL and UL, respectively, where $n_{\text{rep,PDCCH}}$ and $n_{\text{rep,PUCCH}}$ represent the number of repetitions of grant on PDCCH and ACK on PUCCH, respectively, N_{DG2D}, n_{DD2A}, and n_{UG2D} and the processing delays between DL grant and DL data, DL data and ACK, and UL grant and UL data, respectively, and N_{switch} is the switching delay, all in units of SFs.

We show in Section 5.5 that $n_{\text{rep,PUSCH}} > n_{\text{DD2A}}, n_{\text{UG2D}}$ always holds true, especially with the delay values specified in the legacy LTE-M protocol. Therefore, irrespective of the N_{HARQ} that can be used, only one TB can be scheduled per HARQ cycle. We notice that the processing delays and the time of flight between the UE and BS result in several idle SFs in the UE due to the single TB in one HARQ cycle. When multiple TBs are scheduled on one HARQ cycle, the burden of additional delays can be spread over those multiple TBs. This

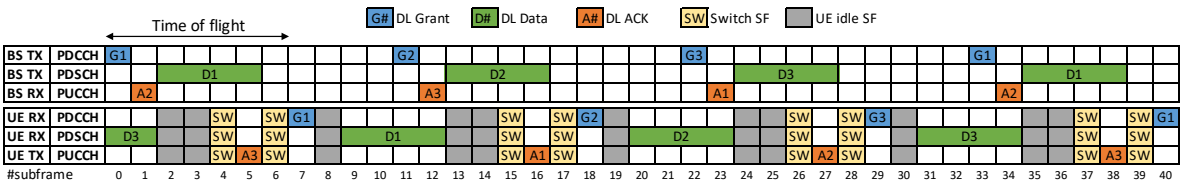


Figure 5.3: Timing diagram of modified HARQ operation for HD-FDD in NTN DL.

reduces the overall overheads per TB and improves the SUF. It is even more beneficial in the context of NTN, since the over-the-air (OTA) travel time can be exploited to overlap the bidirectional UL and DL signals in air. To this end, we propose using flexible values of n_{DD2A} and n_{UG2D} to accommodate multiple TBs per HARQ cycle.

As the first step, we modify the HARQ scheme to utilize the bidirectional overlap of signals in the air. An example for this is shown in the timing diagram for downlink in Figure. 5.3 where $n_{\text{rep,PDSCH}} = 4$ and $n_{\text{DD2A}} = 3$. In contrast with the timing diagram shown in Figure. 5.2 where $N_{\text{HARQ}} = 1$, we consider $N_{\text{HARQ}} = 3$ in Figure. 5.3, to better utilize the possible bidirectional overlap in the air due to the large propagation delay. However, we still see several idle subframes at the UE which is mainly due to the fixed n_{DD2A} that does not allow the base station to transmit more than 1 TB per HARQ cycle or more number of HARQ processes even if the UE supports.

In the next step, we set the processing delays, DD2A and UG2D variable such that some more idle subframes can be utilized. We show an example of our proposed method in Figure. 5.4 for $N_{\text{HARQ}} = 4$, DD2A of 3 SFs, and 4 SFs of data repetitions. In contrast with the timing diagram shown in Figure. 5.3, we could use higher N_{HARQ} along with reducing the number of idle SFs at the UE which results in an improved SUF. The better SUF in Figure. 5.4 is achieved due to the use of flexible n_{DD2A} which allows more than 1 TB per HARQ cycle. Our proposal of using TB-specific downlink data to acknowledgment (DD2A) and uplink grant to data (UG2D) delays[6], i.e., $n_{\text{DD2A},j}$ and $n_{\text{UG2D},j}$, eliminates the overlap between

[6]DD2A delay can also be considered as a physical downlink shared channel (PDSCH) to physical uplink

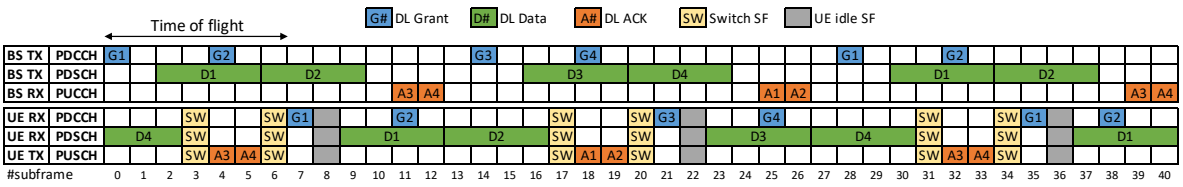


Figure 5.4: Timing diagram of proposed HARQ scheme for HD-FDD in NTN DL.

the data TBs and/or the switching SFs. Thereby, we ensure that an arbitrary number of TBs can be scheduled within one HARQ cycle. The SUF in this case can be expressed as in (5.9) and (5.11), respectively, where N_{TBPHC} is the number of TBs scheduled in one HARQ cycle and $n_{\text{DD2A, min}}$ and $n_{\text{UG2D, min}}$ are the minimum mandatory DD2A and UG2D delays that must be used to ensure that sufficient processing time is available at the UE to process DL data and UL grants, respectively.

$$\eta_{\text{SUF,DL}} = \frac{N_{\text{TBPHC}}}{N_{\text{SF}}}, \tag{5.9}$$

where

$$\begin{aligned} N_{\text{SF}} = {} & n_{\text{rep,PDCCH}} + N_{\text{DG2D}} + N_{\text{TBPHC}} n_{\text{rep,PDSCH}} + n_{\text{rep,PUCCH}} \\ & + \max(n_{\text{DD2A, min}}, (N_{\text{TBPHC}} - 1) n_{\text{rep,PUCCH}}) + 2N_{\text{switch}}. \end{aligned} \tag{5.10}$$

$$\eta_{\text{SUF,UL}} = \frac{N_{\text{TBPHC}}}{n_{\text{rep,PDCCH}} + \max(n_{\text{UG2D, min}}, (N_{\text{TBPHC}} - 1) n_{\text{rep,PDCCH}}) + N_{\text{TBPHC}} n_{\text{rep,PDSCH}} + 2N_{\text{switch}}}. \tag{5.11}$$

While any $N_{\text{TBPHC}} > 1$ can be chosen to obtain higher $\eta_{\text{SUF,DL}}$ and $\eta_{\text{SUF,UL}}$, the limit on the maximum N_{TBPHC} that can be scheduled is set by the maximum N_{HARQ} supported by the specification. The relation between N_{TBPHC} and N_{HARQ} is dependent on the RTT of the network and can be expressed as in (5.12), where N_{A2G} is the ACK processing delay at the BS before scheduling the HARQ process whose TB is acknowledged.

control channel (PUCCH) delay, and UG2D delay can also be considered as a physical downlink control channel (PDCCH) to physical uplink shared channel (PUSCH) delay.

$$N_{\text{HARQ}} = \left\lceil N_{\text{TBPHC}} \left(1 + \frac{T_{\text{RTT}} + N_{\text{A2G}}}{t_{\text{TB}}(n_{\text{rep,PDCCH}} + N_{\text{DG2D}} + N_{\text{TBPHC}}(n_{\text{rep,PDSCH}} + n_{\text{rep,PUCCH}}) + 2N_{\text{switch}})} \right) \right\rceil \tag{5.12}$$

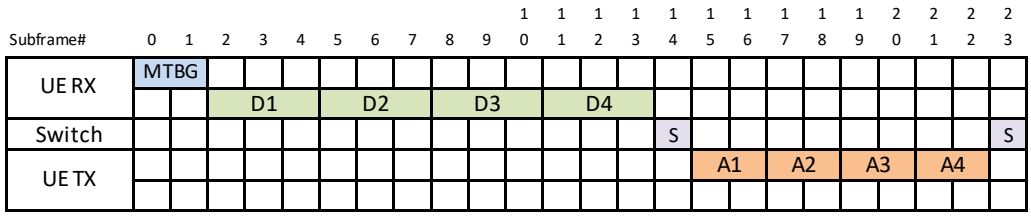


Figure 5.5: Timing diagram for DL operation with MTBG.

5.4.1 Computing and Signaling Variable DD2A and UG2D

After establishing that a variable n_{DD2A} and n_{UG2D} can increase data rates, we now present the methods to compute, choose, and signal these values from the BS to the UE. Toward this end, we consider three different scenarios of HARQ scheduling. The use of MTBG and ACK bundling may not be considered together due to their complexity, and has also already been investigated for delay flexibility in terrestrial networks [161] (and can therefore be extended to NTNs if needed). Therefore, we consider the three other cases of *grouping* grants and ACKs, namely, MTBG without ACK bundling, STBG without ACK bundling, and STBG with ACK bundling.

No ACK Bundling

An example of MTBG without ACK bundling for DL transmission is shown in Figure. 5.5. When STBG is used instead, the grants are split to schedule each TB individually. The variable DD2A of the jth TB is the sum of the time left for the remaining TBs scheduled in the MTBG, the ACKs corresponding to all TBs from $1, 2, ...j-1$, and the switching SFs.

Therefore,

$$n_{\text{DD2A},j} = (N_{\text{TBPHC}} - j)n_{\text{rep,PDSCH}} + (j-1)n_{\text{rep,PUCCH}} + N_{\text{switch}}. \tag{5.13}$$

All the parameters used in (5.13) are already available to the UE to compute $n_{\text{DD2A},j}$ since N_{TBPHC}, $n_{\text{rep,PDSCH}}$, and j are extracted from the MTBG and $n_{\text{rep,PUCCH}}$ is a RRC configured parameter. Therefore, the UE requires no additional signaling from the BS to compute $n_{\text{DD2A},j}$. However, with the use of STBG, $n_{\text{rep,PDSCH}}$ is conveyed by legacy grants, while the identifier j is to be signaled by the BS when scheduling the jth TB. Similarly, N_{TBPHC} can also be explicitly signaled to the UE if the BS chooses to configure an N_{TBPHC} corresponding to an N_{HARQ} that is lower than the supported maximum value. The same principle can also be applied in the UL to obtain the variable UG2D delay as

$$n_{\text{UG2D},j} = (N_{\text{TBPHC}} - j)n_{\text{rep,PDCCH}} + (j-1)n_{\text{rep,PUSCH}} + N_{\text{switch}}. \tag{5.14}$$

Similar to the case of DL, all parameters required to compute (5.14) is acquired by the UE as in the case of DL transmission.

With ACK Bundling

The condition of ACK bundling is only applicable in the DL since there is no notion of acknowledgment in the UL. When the ACKs are bundled, (5.13) can be modified as

$$n_{\text{DD2A},j} = (N_{\text{TBPHC}} - j)n_{\text{rep,PDSCH}} + \left\lfloor \frac{j-1}{n_{\text{bundle}}} \right\rfloor n_{\text{rep,PUCCH}} + N_{\text{switch}}, \tag{5.15}$$

where n_{bundle} is the number of ACKs that are bundled within one TTI. n_{bundle} can be learned by the UE during RRC configuration and hence requires no signaling overhead.

The above analyses assume that the number of repetitions used by all TBs within a HARQ

cycle is the same, as is the case in current LTE-M and NB-IoT operations. However, the BS may choose to use different MCS for each TB based on the type of data being transmitted. While this method is currently not supported in LTE-M or NB-IoT specifications, we address this condition to make our solution future-proof when such an adaptive transmission technique may be implemented in the future. In this case, $n_{\text{rep,PDSCH}}$ and $n_{\text{rep,PUSCH}}$ of each of the remaining or previous TBs are required to compute the DD2A and UG2D delays, respectively. Therefore, (5.13), (5.14), and (5.15) can be modified as

$$n_{\text{DD2A},j} = \sum_{k=j}^{N_{\text{TBPHC}}} n_{\text{rep,PDSCH},k} + (j-1)n_{\text{rep,PUCCH}} + N_{\text{switch}}, \tag{5.16}$$

$$n_{\text{UG2D},j} = (N_{\text{TBPHC}} - j)n_{\text{rep,PDCCH}} + \sum_{k=1}^{j-1} n_{\text{rep,PUSCH},k} + N_{\text{switch}}, \tag{5.17}$$

$$n_{\text{DD2A},j} = \sum_{k=j}^{N_{\text{TBPHC}}} n_{\text{rep,PDSCH},k} + \left\lfloor \frac{j-1}{n_{\text{bundle}}} \right\rfloor n_{\text{rep,PUCCH}} + N_{\text{switch}}, \tag{5.18}$$

respectively, where $n_{\text{rep,PDSCH},k}$ and $n_{\text{rep,PUSCH},k}$ are the number of repeats on the PDSCH and PUSCH for the kth TB, respectively.

5.5 Evaluation Results

In this section, we present numerical results of the increase in data rates achieved with the use of our proposed solutions. The primary reason for the throughput gains obtained by our method can be attributed to the higher number of TBs that can be supported in one HARQ cycle by our variable delay design. However, we recognize that increasing the number of TBs in one HARQ cycle and the consequent possible use of an increased total number of HARQs, especially in the case of NB-IoT UEs, results in a higher complexity at the receiver due to a larger size of the soft-buffer. This condition is typically not preferable when the receiver

is the UE (i.e., for DL communications). On the other hand, a higher receiver complexity introduced by the increased buffer size is negligible in the uplink where the BS receives the TBs. Hence, considering practical implementation scenarios, we focus on the uplink to evaluate the throughput gains achievable using our proposed methods. From (5.9)−(5.18), we notice that the throughput achievable with the use of our method relies on the number of data TB repetitions. To this end, we first begin our evaluation campaign with a link-level simulation of a point-to-point IoT-NTN uplink path to determine a suitable value of $n_{\text{rep,PUSCH}}$ required to achieve a target BLER.

5.5.1 Simulation Settings

We derive a majority of our simulation settings from the relevant 3GPP technical reports and technical documents related to NTN [25, 26, 162]. As suggested by 3GPP, we consider the NTN UEs to be enabled with a GNSS ability such that it can perform pre- and post-compensation of the frequency offset [163]. Therefore, we apply a maximum residual frequency offset of 34 Hz after post-compensation at the satellite and pre-compensation along with continuous frequency tracking at the UE [67]. We use the 3GPP recommended NTN tapped delay line (TDL) non-line-of-sight (NLOS) channel model to investigate the performance of our solution. We scale the power delay parameters of the reference TDL-A model [26] according to the desired value of delay spread specified for suburban environment [25,164]. We choose two different values of $n_{\text{TB}} = \{144, 504\}$ to investigate the impact of TBS on BLER and spectral efficiency. We list all simulation parameters in Table 5.3.

5.5.2 Numerical Results

We begin by presenting our link-level simulation results.

Table 5.3: Evaluation settings for NTN HARQ.

Parameter	Value
n_{TB}	144, 504 bits
Modulation	QPSK
No. of PRBs	1
SFs for channel estimation	5
No. of HARQ processes	1
Channel model	NTN TDL-A
UE environment	Suburban NLOS
No. of transmit antenna	1
No. of receive antenna	2
UE Speed	15 km/h
UE-satellite elevation angle	30 degrees
Feeder-satellite elevation angle	10 degrees
Residual frequency offset	Uniformly distributed in $[-34$ Hz, 34 Hz]
f_{c}	2 GHz
Target BLER	10%
$P_{\text{EIRP,UE}}$	23 dBm
G/T	-4.9 dB/T
B	180 kHz
α_{atm}	0.1 dB
α_{sh}	3 dB
α_{sc}	2.2 dB
α_{pol}	3 dB
α_{add}	3 dB

Number of Repetitions

In Figs. 5.6 and 5.7, we present the results of the variation of BLER for different possible SNR values for TDL-A channel. As expected, the BLER achieved with increased number of repetitions is lower due to the increased redundancy. Based on our desired operating conditions (i.e., SNRs), we use the results in Figs. 5.6 and 5.7 to choose the required number of repetitions to achieve a target BLER. We then use these numbers to present the throughput gains of our proposed solutions in the following.

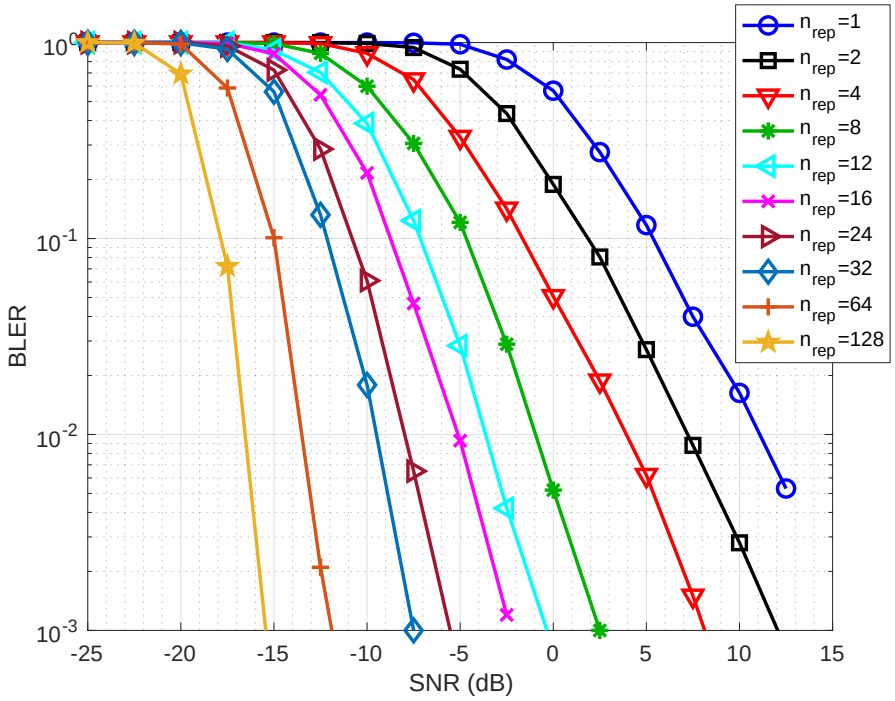


Figure 5.6: BLER vs SNR for different repetitions (n_{rep}) of PUSCH with TBS of 144 bits in NTN TDL-A channel.

Spectral Efficiency and Throughput Gain

We focus on the UL throughput gain provided by the use of our variable delay methods using both LTE-M and NB-IoT based UEs. We consider the two LEO satellite access types of LEO600 and LEO1200 that consist of satellites at altitudes of 600 km and 1200 km, respectively. We use an elevation angle of 30 degrees, which provides $T_{\text{RTT}} = 20$ ms and $T_{\text{RTT}} = 34$ ms for LEO600 and LEO 1200 scenarios, respectively. Next, we compute the operating SNR condition to determine a suitable value of n_{rep} to use for evaluating our proposed solutions. We compute the SNR, γ, as

$$\gamma = \frac{P_{\text{EIRP,UE}} G}{B k_{\text{B}} T \alpha_{\text{fs}} \alpha_{\text{atm}} \alpha_{\text{sh}} \alpha_{\text{sc}} \alpha_{\text{pol}} \alpha_{\text{add}}}, \tag{5.19}$$

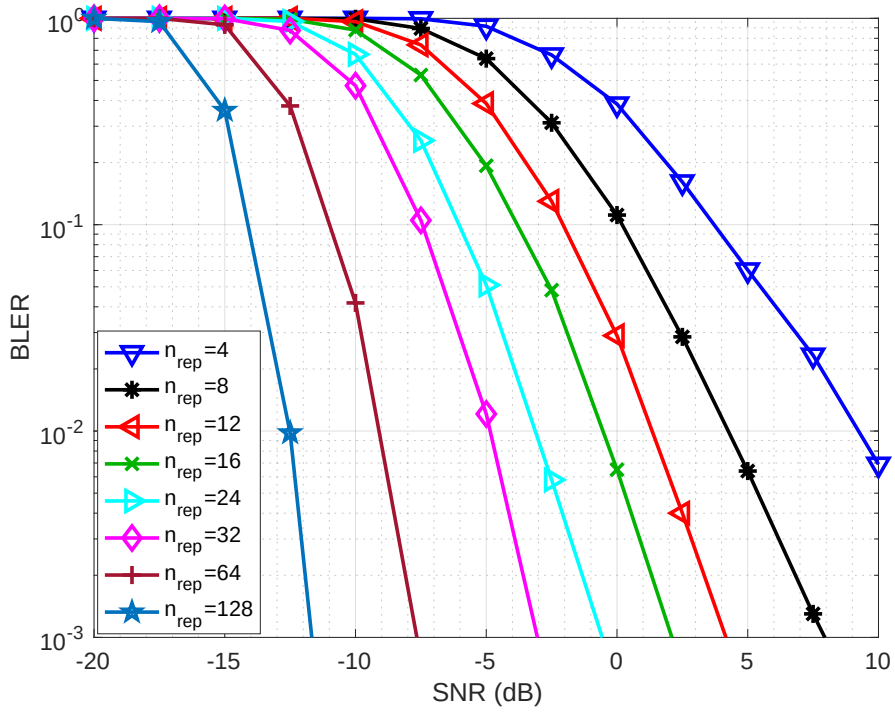


Figure 5.7: BLER vs SNR for different repetitions (n_{rep}) of PUSCH with TBS of 504 bits in NTN TDL-A channel.

where α_{fs} is the free space path loss given by [26]

$$\alpha_{\text{fs}} = (4\pi f_{\text{c}} d/c)^2, \tag{5.20}$$

B is the signal bandwidth, $P_{\text{EIRP,UE}}$ is the effective isotropically radiated power from the NTN UE, G/T is the antenna-gain-to-noise-temperature value of the satellite antenna, k_{B} is the Boltzmann constant, f_{c} is the carrier frequency in GHz, d is the distance between the UE and the satellite, and α_{atm}, α_{sh}, α_{sc}, α_{pol}, α_{add} are the atmospheric loss, shadow fading margin, scintillation loss, polarization loss, and additional loss, respectively. We compute d based on the satellite altitude and the elevation angle listed in Table 5.3. Using the values from Table 5.3, we obtain $\gamma = -0.2$ dB and $\gamma = -5.6$ dB for LEO600 and LEO1200 scenarios, respectively. These conditions are also consistent with the suggested link budget evaluations

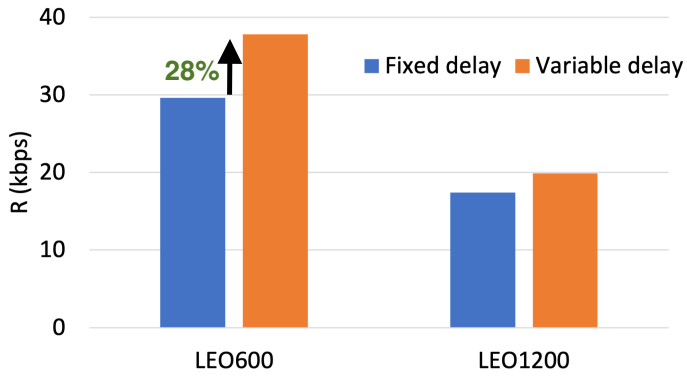


Figure 5.8: Throughput for LEO600 and LEO1200 for the conventional fixed delay method and the proposed variable delay design in an LTE-M UE.

presented for NR-NTN in [25].

We find from Figs. 5.6 and 5.7 that for a fixed target BLER, choosing a higher value of n_{TB} provides better spectral efficiency. For example, for a target BLER of 10% at the operating value of $\gamma = -5.6$ dB for LEO1200, $n_{\text{rep,PUSCH}} = 12$ and $n_{\text{rep,PUSCH}} = 24$ for $n_{\text{TB}} = 144$ and $n_{\text{TB}} = 504$, respectively. This results in a spectral efficiency of 12 and 21 bits/PRB for $n_{\text{TB}} = 144$ and $n_{\text{TB}} = 504$, respectively. This phenomenon of higher spectral efficiency for larger TBS is also true across satellite access types and target BLERs. Therefore, we perform our throughput gain evaluation for our proposed method with $n_{\text{TB}} = 504$. We extract the corresponding numbers for n_{rep} for both satellite access types at their operating SNR values from Figure. 5.7 as $n_{\text{rep}} = 12$ and $n_{\text{rep}} = 24$ for LEO600 and LEO1200 scenarios, respectively. This also clearly shows that $n_{\text{rep,PUSCH}}$ for both LEO600 and LEO1200 cases is greater than n_{DD2A} and n_{UG2D}. As demonstrated in Section 5.4, this condition results in the PUCCH SFs overlapping the PDSCH time slots in the DL and PDCCH SFs overlapping with PUSCH in the UL, respectively, for the case of fixed DD2A and UG2D methods of the state-of-the-art.

Next, we compute the throughput using (5.11) and (5.2). We begin with LTE-M systems, where the state-of-the-art method uses a fixed $n_{\text{UG2D}} = 3$ and one switching SF. We compare

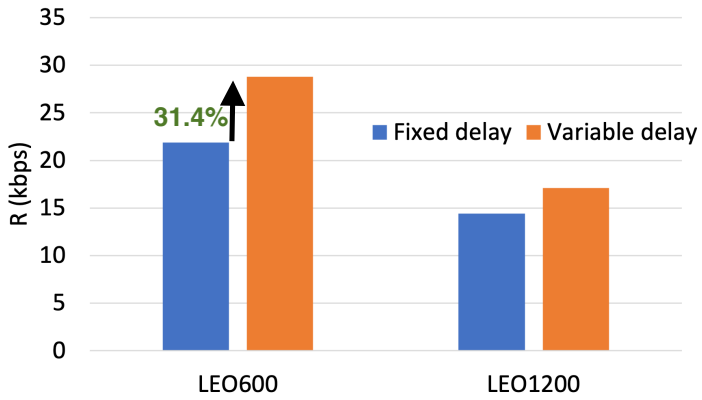


Figure 5.9: Throughput for LEO600 and LEO1200 for the conventional fixed delay method and the proposed variable delay design in an NB-IoT UE.

the achievable data rates using the state-of-the-art technique and our proposed solution with variable n_{UG2D} in Figure. 5.8. The results demonstrate that we achieve a 28% increase in data rate using our proposed method for LEO600 satellite constellation altitude. For the case of NB-IoT UEs, whose results are shown in Figure. 5.9, we observe an even higher increase in throughput of over 31% for LEO600 with the use of $n_{\text{UG2D}} = 8$ for the fixed delay design and two SFs allotted for switching the UE between transmission and reception. Note that the results for NB-IoT were with an increased $N_{\text{HARQ}} = 4$, whereas the LTE-M system evaluations were with the current allowed maximum of $N_{\text{HARQ}} = 8$.

We further observe in both Figure. 5.8 and Figure. 5.9 that the throughput gains obtained with the use of our method is higher for lower satellite altitudes. This is because the value of n_{rep} required to achieve a target BLER increases with higher satellite altitudes due to the increase in pathloss that demands lower code rates. As a result, the increase in throughput obtained with greater number of TBs per HARQ cycle crosses a point of diminishing returns. This result is also intuitive, since a higher number of repetitions can cover a larger portion of the propagation time delay and therefore presents smaller stop-and-wait gaps to be filled with the use of a larger number of TBs per HARQ cycle. Consequently, this reduces the amount of throughput gain that is achievable from our proposed solution. Nevertheless, for

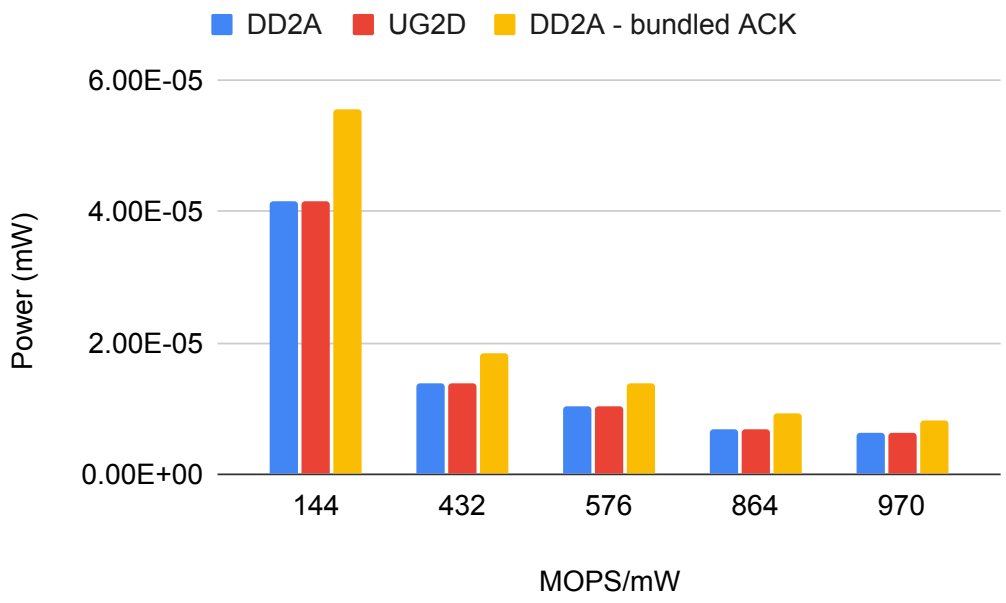


Figure 5.10: Power consumption for implementing our solution in a UE with different processor architectures.

each of these cases, our method provides a superior value of R by reducing the transmission overheads associated with HARQ-based communications.

Run-time Complexity and Power Consumption

In our final evaluation portion, we present computational complexity results associated with the use of our solution. In particular, we show the run-time complexity and the power consumption resulting from our solution. To this end, we only focus on the computations performed at the battery-powered UE side.

The additional computations at the UE-end associated with the deployment of our solution include calculating the variable delays shown in Section 5.4.1. We first determine the number of operations involved with computing DD2A and UG2D. We then use a worst-case assumption that the computations are performed at every SF level, i.e., once every 1 ms. Note that this is an exaggerated case. In practice, computations at the UE-side need to be

performed only when there is a noticeable change in the RTT, for example, when the BS uses a different N_{TBPHC} in response to the RTT variation. Previous investigations have shown that the RTT varies at a rate of less than 100 μs/s for a LEO satellite at an altitude of 600 km and a UE that is moving at a speed of 1200 km/h in the opposite direction of the satellite movement [165]. This results in an RTT variation of less than a nanosecond per SF.

Despite using these numbers, the power consumption results in Figure. 5.10 show that our method introduces less than 60 nW of additional power, even with a relatively less efficient processor that provides an efficiency of 144 million operations per second (MOPS) per mW. With a superior processor efficiency, e.g., 970 MOPS/mW, our method introduces less than 7 nW of additional power consumption to determine the variable delays. Such a low power consumption introduces a truly negligible impact on the battery life of an IoT device.

5.6 Summary

In this chapter, we presented a dynamic HARQ scheduling design targeted at IoT-NTN UEs to exploit the extended signal propagation time encountered in satellite communication links. We presented an analysis to determine a suitable number of HARQs to be supported to extract superior throughput under any propagation condition. Our detailed simulation evaluation demonstrated noticeable gains in the achieved throughput with the use of our proposed methods, by considering suitable coding rates for different transmission link conditions and satellite altitudes. Our solution enables NTNs to serve an increased number of UEs, which is critical given the extended cell-size in NTN and the increasing number of connected devices.

Chapter 6

Concluding Remarks and Future Directions

In this final chapter of the thesis, we provide a summary, conclusions drawn from our research, and the potential directions for future work.

6.1 Summary and Conclusions

With the primary goal to achieve seamless global network coverage for all C-IoT UEs, we developed and evaluated solutions for problems related to C-IoT communication in D2D and satellite links. First, we designed an S-RAT D2D protocol called SCUBA which operates in a TDM manner with the P-RAT in the cellular UE. Then, we designed a positioning algorithm called SPIN for the C-IoT devices in NTN with the main objective to achieve UL synchronization in the presence of high residual Doppler and TA. Finally, we designed dynamic HARQ scheduling techniques to improve the throughput of C-IoT devices in NTN which is degraded by the stop-and-wait gaps due to large RTT.

D2D communication is a critical constituent of IoT which benefits the network and the UEs in numerous ways. While D2D offers potential reduction in network latency, UE power consumption, and usage of network resources, it also helps in extending the network coverage. Further exploiting the unlicensed bands of frequency for D2D communication offloads network traffic from the limited licensed spectral resources and also reduces spectrum licens-

ing costs. Our feasibility analysis provided insights about the availability of free time to perform D2D S-RAT communication. It also gave us preliminary indications about how the P-RAT traffic impacts the S-RAT communication latency. The duty cycle analysis for different traffic models also helped us to decide whether any special considerations are required for SCUBA to meet the regulations. This approach to design a new protocol by starting with a feasibility analysis can also be adopted to any other protocol design. Such a methodology ensures that the prospective protocol can function successfully in the application scenario by coexisting with the pre-existing conditions while also meeting its own target requirements.

The protocol SCUBA, designed based on the insights from the feasibility study, brings all the benefits of D2D-U while also supporting sidelink communication in low-cost HD-FDD UEs. The TDM operation of SCUBA as an S-RAT along with any underlying P-RAT makes it ideally fit into a C-IoT UE with HD-FDD and single radio architecture. SCUBA provides options for configuring battery-latency tradeoff thereby making itself appealing to be used in a wide variety of applications. The flexibility to allow long SL-DRX cycles makes SCUBA ideal for power-critical applications that include sensors, logistics, and tracking. Conversely, the LLM feature of SCUBA makes it a favourable choice for low latency applications such as traffic safety and control, smart grid, and industrial applications and control. The latency and power analyses and simulations help us in deciding the battery-latency tradeoff for a chosen application scenario. From the evaluations, we could also identify the use-cases where SAM feature of SCUBA could be beneficial. The power consumption and latency comparison with BLE and ZigBee showed that SCUBA consumes less power than other standards while also maintaining a comparable latency. The synchronization methods that we designed for the SCUBA protocol extend its applicability to all sorts of terrestrial network coverage scenarios including HC, PC, and OOC. Synchronized SCUBA allows seamless D2D communication thus making it a good choice for mobile IoT applications. The power consumption analysis of

synchronization methods acts as a guiding tool to select the most suitable method for the specific application under consideration. From the analysis, we also learned that an RX beacon based synchronization method is more battery friendly than the conventional TX beacons, when UEs belonging to lower power classes are considered. The evaluation also highlights a beacon power threshold which is vital in deciding the appropriate sync method for a chosen SCUBA network. These findings are independent of the underlying communication protocol and hence are applicable to the synchronization operations for any D2D communication technology. Moreover, the SCUBA synchronization methods proposed in the thesis can be directly adopted to any D2D communication network to provide synchronization. Finally, since SCUBA operates on unlicensed band of frequencies, it is also an ideal candidate for the LTE D2D counterpart under the MulteFire framework.

NTN integration is crucial for C-IoT devices as it extends network coverage to UEs located at unserved and underserved areas. The UE requires the knowledge of its own location to compute and compensate the residual TA and Doppler to have its UL signals synchronized at the BS. However, for mobile C-IoT devices with long connected mode, using GNSS for positioning requires repeated termination and re-establishment of RRC connection. This results in huge battery drain which is not ideal for an IoT UE. To this end, we designed SPIN, an algorithm for the C-IoT UEs to perform self-positioning using the DL synchronization signals. SPIN uses fine resolution techniques to accurately measure TOA and FOA values and then performs joint TDOA-FDOA based positioning to estimate the UE location and velocity. The UE uses its estimated location, the coordinates of RP, and the location and velocity of satellite to compute and pre-compensate residual TA and Doppler offset in the UL signal. The Doppler offset due to UE mobility is usually compensated by the BS. Since SPIN also estimates the UE velocity with high accuracy, the UE can additionally compute and compensate the Doppler caused by its own mobility and hence relieve the burden off the network.

The CRLB for TOA, FOA, position, and velocity estimations helped us to determine if the target accuracy requirements are achievable using the available signals and measurements. The CRLBs also guided us to identify the room for improvement at every step of SPIN design and thus helped us to add necessary operations to attain better accuracy. The numerical results that we obtained through PHY and system level simulations of SPIN showed that the positioning accuracy attains the CRLB and meets the target requirement for UL synchronization, without demanding much computational complexity. The computational complexity and positioning accuracy analyses help us to decide on the appropriate accuracy-complexity tradeoff for a selected use-case. For instance, one can choose to either increase the SPIN processing time or reduce the number of TWLS iterations to relax the computational complexity. The positioning operation takes longer if SPIN processing time is increased, whereas the accuracy gets degraded if the number of TWLS iterations is reduced. More importantly, our numerical results showed that using SPIN offers significant battery life saving over the GNSS based solution. Thus, in contrast with the state-of-the-art solutions, SPIN solves the UL synchronization problem with less power consumption, no additional signaling, and low computational complexity. To the best of our knowledge, the positioning algorithm for IoT NTN using DL broadcast signals proposed in this thesis is the first of its kind. Therefore, the analyzes and evaluation methodology and results presented in this thesis are useful guiding tools for future research in IoT NTN positioning. Although the primary goal of SPIN is to provide UL synchronization in NTN, it also finds applications in LBS, such as emergency services, e.g., E911, infotainment services, and local advertising.

The second problem we identified in IoT NTN is the throughput degradation caused by the extended RTT. The prior art solutions addressing this issue include increasing the number of HARQ processes, targeting low iBLER, and disabling HARQ. However, we designed improved HARQ scheduling techniques which do not require any mandatory increase in the

number of HARQ processes or disabling the HARQ. Therefore, our solution retains the low complexity, low power consumption, and the benefits of HARQ combining in PHY/MAC layer which avoids relying on RLC layer retransmissions. We proposed flexible DL data to acknowledgment delays and UL grant to data delays for more efficient bidirectional signaling. Our dynamic HARQ scheduling design lets the UE utilize the lengthy RTT in NTN more efficiently. Our design also includes the methods for calculating and signaling the flexible delay for dynamic transmission adaptation conditions. The numerical results showed that our proposed solution results in noticeable throughput gain over the conventional HARQ scheduling method. Our HARQ scheduling solution with flexible delay provides throughput gain of 28% and 31% for LTE-M and NB-IoT, respectively, over the conventional method. Consequently, our solution helps the network to serve more UEs, which is particularly beneficial in NTN considering its large cell size.

In conclusion, we designed and evaluated solutions which address the problems associated with C-IoT networks and enhance their performance. Synchronized SCUBA is an ideal solution for realizing the D2D links in a C-IoT network with an objective to fill the coverage gaps and/or to extend the coverage. SPIN algorithm and flexible HARQ scheduling techniques proposed in the thesis are highly beneficial to IoT NTN which eventually help us to accomplish global C-IoT coverage. Our solutions are appealing to low-cost C-IoT devices since they are cost-effective and power-efficient. Our comprehensive analyses, detailed evaluations, and comparison with the state-of-the-art methods prove the effectiveness and substantial superiority of our solutions over the prior art. Altogether, the solutions proposed in the thesis take us one step closer to the ultimate goal of seamless universal cellular coverage which connects every object together to create a smarter world.

Research Impact on Standards and Future C-IoT Devices

The SCUBA protocol designed in this thesis has potential to be defined as an industry standard as the unlicensed D2D counterpart for the LTE/NR standard. The standardization may occur within an existing standards group (e.g., ETSI, MulteFire) or within a new standards group. Sierra Wireless Inc. also considers SPIN as a viable solution for UL synchronization which could be implemented in future NTN IoT products. Finally, our contribution on NTN HARQ has been presented to RAN-Layer 1 (RAN-1) meetings of 3GPP in the context of NTN IoT HARQ enhancements. As the 3GPP IoT is constantly evolving, our variable HARQ delay technique is a candidate to be adopted to the 3GPP specification in future releases.

6.2 Future Work

In this final section of the thesis, we briefly discuss the potential directions for future research work.

6.2.1 Self-Organized SCUBA Networks

The SCUBA protocol designed in this thesis includes detailed specification of MAC layer and expects to adopt the PHY same as that of the P-RAT. While the decentralized operation of SCUBA offers the advantage of not requiring assistance from a BS, it has some limitations. For example, consider an application where there is a group of UEs rather than a single pair of UEs. The UEs need to report data to a central server which may not be necessarily accessible by all the UEs due to different coverage scenarios. In this case, SCUBA can be potentially used to establish D2D links thereby realizing a mesh. In this SCUBA mesh network, some UEs can act as gateways to aggregate data from other UEs and upload them to the server. The

SCUBA mesh network has potential benefits of reduction in overall cellular subscription cost and power consumption. However, the current SCUBA specifications do not define the roles and responsibilities of SCUBA UEs in a mesh network. Designing algorithms and defining specifications for a self-organized SCUBA network (SOSN) is a potential area for future research. The prospective SOSN should optimize the SCUBA network organization to save cost and energy with guaranteed end-to-end delay and reliability requirements.

6.2.2 SCUBA Prototype Implementation

Prototype implementation is naturally the next step after designing a product. For example, we could create a working demo of SCUBA using RFmx LTE/LTE-Advanced LabVIEW tool by National Instruments (NI) along with their vector signal transceiver (VST), e.g., NI PXIe-5840. We can implement the PHY of both P-RAT and SCUBA using the RFmx LTE/LTE-Advanced tool. However, we need to write a customized LabVIEW program for the assumed traffic model, the associated packet generation trigger, and the switching between licensed and unlicensed frequency bands. Another step which is closer towards realizing the end product is to create a minimum viable product (MVP) with only essential features to validate the SCUBA design and to attract early adoption customers.

6.2.3 Positioning Using SCUBA

To complement and/or enhance the state-of-the-art positioning techniques in terrestrial networks, we can use D2D communications to achieve a crowd-sourced location estimation strategy. In this positioning method, one or more tracker or finder devices, e.g., smart phones, need to search and find one or more targets, e.g., a biking group, family members, children, the elderly, pets, or equipment at a construction site. For such use-cases, the existing C-IoT positioning can achieve macro-positioning up to 100 m accuracy. As explained in Sec-

tion 1.1.3, GNSS may not be necessarily available due to cost and/or power constraints or since the UE is located indoor. Therefore, for sub-100 m micro-positioning, we can potentially develop a cooperative positioning-based approach where targets, trackers, and other nodes available to cooperate can exchange positioning messages to each other via sidelink. Since we mainly target low-cost C-IoT devices, we may additionally impose single radio and HD-FDD constraints. Therefore, we need to devise a positioning protocol which uses a cellular sidelink radio interface for positioning message transfer. SCUBA is an ideal candidate for the sidelink RAT to be used for this cooperative positioning approach. However, when there are no devices available to cooperate, we also consider a prospective solution where the user can deploy a set of temporary tracking devices, which we refer to as agent nodes. Here, the agent nodes and targets can communicate to each other using SCUBA and can do minimal communication to the cellular network when it is required, for e.g., to send the positioning measurements to the server or a smart phone to compute the location. For this solution, we can borrow techniques from SCUBA which we have already defined in Chapter 2. For the positioning methodology, we need to investigate the suitability of using state-of-the-art positioning techniques and expand on it by designing novel positioning signals and associated communication schemes, e.g., discontinuous transmission (DTX) and DRX. As either a complementary or a standalone use-case where the tracker device can move and is required to find the target device, we also consider a warm-cold approach based positioning technique. This will be similar to conventional Bluetooth received signal strength indicator (RSSI)-based tracking solution, but constrained by low-cost single radio device. Therefore, SCUBA is an ideal choice for this solution too. In the same way as above, we can perform macro-positioning using conventional cellular-IoT positioning. However, the final step of physically finding the target needs to be done based on the RSSI of the positioning reference signal received at the tracker. SCUBA can be used to transmit this measurement signal.

6.2.4 NR Reduced Capability Devices

NR reduced capability (RedCap) devices are recently introduced category of NR devices which are targeted for mid-tier use cases [166]. Unlike the high bandwidth NR UEs used in eMBB applications, very low complexity NR UEs used in mMTC applications, and low latency UEs used in ultra-reliable low-latency communications (URLLC), RedCap UEs close the gap by targeting those applications which are not covered by the former categories. The RedCap devices find applications in video surveillance, industrial wireless sensors, and wearables. As 3GPP recently completed the study phase for creating a sub-category of RedCap devices with further reduced complexity, a related work item to investigate the performance degradation due to the complexity reduction gains attention [167]. Specifically, the reduction in bandwidth aimed at complexity reduction impacts the positioning accuracy of a RedCap UE. Investigation of the performance degradation and design of appropriate solutions to close the gap with the target accuracy requirements are potential areas for future research.

6.2.5 Ambient IoT

3GPP has recently identified a new study item which targets IoT applications which cannot be served using the existing standards such as NB-IoT or LTE-M [168]. This study item targets those UEs with ultra-low complexity and ultra-low power consumption such that they are battery-less either with no energy storage or with very limited energy storage. The targeted ambient IoT applications include automated warehouse inventory, medical inventory, smart grid and substation safety monitoring, and automobile manufacturing inventory [169]. The study item also finds that the radio frequency identification (RFID) which is the most common technology currently used in inventory applications, is not feasible as it is labour-intensive, has short reading range, and also leads to severe interference. Bringing ambient IoT under 3GPP standardization requires further research on various areas such as designing a proper

communication technology which satisfies the target requirements [169], devising energy harvesting processes to utilize the ambient sources of energy, and bulk management of IoT devices.

Bibliography

[1] R. Borralho, A. Mohamed, A. U. Quddus, P. Vieira, and R. Tafazolli, "A survey on coverage enhancement in cellular networks: Challenges and solutions for future deployments," *IEEE Commun. Surveys & Tuts.*, vol. 23, no. 2, pp. 1302–1341, 2021.

[2] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys & Tuts.*, vol. 16, no. 4, pp. 1801–1819, 2014.

[3] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in LTE-advanced networks: A survey," *IEEE Commun. Surveys & Tuts.*, vol. 17, no. 4, pp. 1923–1940, 2014.

[4] H. A. U. Mustafa, M. A. Imran, M. Z. Shakir, A. Imran, and R. Tafazolli, "Separation framework: An enabler for cooperative and D2D communication for future 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 419–445, 2015.

[5] IEEE Spectrum. Applications of device-to-device communication in 5G networks. Accessed: Sept. 3, 2022. [Online]. Available: https://bit.ly/2PMG43u

[6] L. Babun, A. I. Yürekli, and I. Güvenç, "Multi-hop and D2D communications for extending coverage in public safety scenarios," in *IEEE 40th Local Computer Netw. Conf. Workshops (LCN Workshops)*, 2015.

[7] F. S. Shaikh and R. Wismüller, "Routing in multi-hop cellular device-to-device (D2D) networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2622–2657, 2018.

[8] M. Ahmed, Y. Li, M. Waqas, M. Sheraz, D. Jin, and Z. Han, "A survey on socially aware device-to-device communications," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2169–2197, 2018.

[9] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in LTE-Advanced networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1923–1940, 2015.

[10] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, 2014.

[11] TS 36.321 V15.3.0, "Evolved universal terrestrial radio access (E-UTRA) medium access control (MAC) protocol specification (Release 15)," 3rd Generation Partnership Project (3GPP), Tech. Spec., Sept. 2018.

[12] TS 36.211 V16.5.0, "Evolved universal terrestrial radio access (E-UTRA) physical channels and modulation (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Spec., Mar. 2021.

[13] J. Schlienz and A. Roessler, "Device to device communication in LTE," White Paper, Rhode & Schwarz, Sept. 2015.

[14] S. Lien, D. Deng, C. Lin, H. Tsai, T. Chen, C. Guo, and S. Cheng, "3GPP NR sidelink transmissions toward 5G V2X," *IEEE Access*, vol. 8, pp. 35 368–35 382, 2020.

[15] TS 38.300 V16.4.0, "NR and NG-RAN overall description (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Spec., Dec. 2020.

[16] P. Keshavamurthy, E. Pateromichelakis, D. Dahlhaus, and C. Zhou, "Edge cloud-enabled radio resource management for co-operative automated driving," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 7, pp. 1515–1530, 2020.

[17] A. Moubayed, A. Shami, P. Heidari, A. Larabi, and R. Brunner, "Edge-enabled V2X service placement for intelligent transportation systems," *IEEE Trans. Mobile Comput.*, vol. 20, no. 4, pp. 1380–1392, 2021.

[18] R. Srinivasa, N. K. S. Naidu, S. Maheshwari, C. Bharathi, and A. R. Hemanth Kumar, "Minimizing latency for 5G multimedia and V2X applications using mobile edge computing," in *Int. Conf. Intell. Commun. Comput. Tech.*, 2019, pp. 213–217.

[19] Y. Qi, L. Tian, Y. Zhou, and J. Yuan, "Mobile edge computing-assisted admission control in vehicular networks: The convergence of communication and computation," *IEEE Veh. Technol. Mag.*, vol. 14, no. 1, pp. 37–44, 2019.

[20] Y. Zhou, L. Tian, L. Liu, and Y. Qi, "Fog computing enabled future mobile communication networks: A convergence of communication and computing," *IEEE Commun. Mag.*, 2019.

[21] H. T. Mouftah, M. Erol-Kantarci, and M. H. Rehmani, *Transportation and Power Grid in Smart Cities: Communication Networks and Services*. Wiley, 2018.

[22] S. Andreev, O. Galinina, A. Pyattaev, M. Gerasimenko, T. Tirronen, J. Torsner, J. Sachs, M. Dohler, and Y. Koucheryavy, "Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 32–40, 2015.

[23] L. Liang, L. Xu, B. Cao, and Y. Jia, "A cluster-based congestion-mitigating access scheme for massive M2M communications in Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2200–2211, 2018.

[24] M. Bouzouita, Y. Hadjadj-Aoul, N. Zangar, and G. Rubino, "Estimating the number of contending IoT devices in 5G networks: Revealing the invisible," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 4, pp. 1–16, 2018.

[25] TR 38.821 V1.0.0, "Solutions for NR to support non-terrestrial networks (NTN) (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Dec. 2019.

[26] TR 38.811 V15.2.0, "Study on New Radio (NR) to support non-terrestrial networks (Release 15)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Sept. 2019.

[27] A. Guidotti, A. Vanelli-Coralli, A. Mengali, and S. Cioni, "Non-terrestrial networks: Link budget analysis," in *IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–7.

[28] F. Rinaldi, H. Määttänen, J. Torsner, S. Pizzi, S. Andreev, A. Iera, Y. Koucheryavy, and G. Araniti, "Broadcasting services over 5G NR enabled multi-beam non-terrestrial networks," *IEEE Trans. Broadcast.*, pp. 1–13, 2020.

[29] X. Lin, B. Hofström, Y.-P. E. Wang, G. Masini, H.-L. Maattanen, H. Rydén, J. Sedin, M. Stattin, O. Liberg, S. Euler, S. Muruganathan, S. E. Löwenmark, and T. Khan, *5G New Radio Evolution Meets Satellite Communications: Opportunities, Challenges, and Solutions*. Springer International Publishing, 2021, pp. 517–531. [Online]. Available: https://doi.org/10.1007/978-3-030-58197-8_18

[30] N. H. Mahmood, S. Böcker, A. Munari, F. Clazzer, I. Moerman, K. Mikhaylov, O. Lopez, O.-S. Park, E. Mercier, H. Bartz *et al.*, "White paper on critical and massive machine type communication towards 6G," *arXiv preprint arXiv:2004.14146*, 2020.

[31] M. Bacco, F. Davoli, G. Giambene, A. Gotta, M. Luglio, M. Marchese, F. Patrone, and C. Roseti, "Networking challenges for non-terrestrial networks exploitation in 5G," in *IEEE 2nd 5G World Forum (5GWF)*, 2019, pp. 623–628.

[32] Eutelsat, "R1-2008868: Summary #4 of 8.15.1 IoT NTN scenarios," 3GPP TSG RAN WG1 Meeting #103e, Tech. Doc., Nov. 2020.

[33] F. Zhu, T. Ba, Y. Zhang, X. Gao, and J. Wang, "Terminal location method with NLOS exclusion based on unsupervised learning in 5G-LEO satellite communication systems," *Int. J. Satell. Commun. Netw.*, vol. 38, no. 5, pp. 425–436, Feb. 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sat.1346

[34] X. Lin, Z. Lin, S. E. Löwenmark, J. Rune, and R. Karlsson, "Doppler shift estimation in 5G new radio non-terrestrial networks," in *IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.

[35] O. Kodheli, A. Astro, J. Querol, M. Gholamian, S. Kumar, N. Maturo, and S. Chatzinotas, "Random access procedure over non-terrestrial networks: From theory to practice," *IEEE Access*, vol. 9, pp. 109 130–109 143, 2021.

[36] H. Chougrani, S. Kisseleff, W. A. Martins, and S. Chatzinotas, "NB-IoT random access for non-terrestrial networks: Preamble detection and uplink synchronization," *IEEE Internet Things J.*, 2021.

[37] Huawei, HiSilicon, and CAICT, "R1-1908049 Discussion on Doppler compensation, timing advance and RACH for NTN," 3GPP TSG RAN WG1 Meeting #98, Tech. Doc., Aug. 2019.

[38] AT&T, "RP-212729 New WID on IoT NTN enhancements," 3GPP TSG RAN WG1 Meeting #94-e, Tech. Doc., Dec. 2021.

[39] S. Cobb, D. Lawrence, M. O'Connor, G. Gutt, F. Tremblay, and P. Laplante. Test results from an operationally deployed Iridium-based secure timing solution, Satelles, Inc. Accessed: July 9, 2022. [Online]. Available: https://wsts.atis.org/wp-content/uploads/sites/9/2018/11/5-03_Satelles_Cobb_Test-Results-Iridium.pdf

[40] M. Karaim, M. Elsheikh, and A. Noureldin, "GNSS error sources," in *Multifunctional Operation and Application of GPS*, R. B. Rustamov and A. M. Hashimov, Eds. Rijeka: IntechOpen, 2018, ch. 4. [Online]. Available: https://doi.org/10.5772/intechopen.75493

[41] M. Thorpe and E. Zelmer, "LTE location based services technology introduction," White Paper, Rhode & Schwarz, Sept. 2013.

[42] A. K̈upper, *Location-based Services Fundamentals and Operation*. John Wiley & Sons Ltd, 2014.

[43] F. S. Prol, R. M. Ferre, Z. Saleem, P. Välisuo, C. Pinell, E. S. Lohan, M. Elsanhoury, M. Elmusrati, S. Islam, K. Çelikbilek, K. Selvan, J. Yliaho, K. Rutledge, A. Ojala, L. Ferranti, J. Praks, M. Z. H. Bhuiyan, S. Kaasalainen, and H. Kuusniemi, "Position, navigation, and timing (PNT) through low earth orbit (LEO) satellites: A survey on current status, challenges, and opportunities," *IEEE Access*, vol. 10, pp. 83 971–84 002, 2022.

[44] *Bluetooth core specification version 5.1*, Bluetooth SIG Std., Jan. 2019.

[45] *ZigBee specification*, ZigBee Alliance Std., Sept. 2012.

[46] D. Camps-Mur, A. Garcia-Saavedra, and P. Serrano, "Device-to-device communications with Wi-Fi Direct: overview and experimentation," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 96–104, 2013.

[47] G. Ding, J. Wang, Q. Wu, Y. Yao, F. Song, and T. A. Tsiftsis, "Cellular-base-station-assisted device-to-device communications in TV white space," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 107–121, Jan. 2016.

[48] F. Wu, H. Zhang, B. Di, J. Wu, and L. Song, "Device-to-device communications underlaying cellular networks: To use unlicensed spectrum or not?" *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6598–6611, 2019.

[49] B. Kang, S. Choi, S. Jung, and S. Bahk, "D2D communications underlaying cellular networks on licensed and unlicensed bands with QoS constraints," *J. Commun. Netw.*, vol. 21, no. 4, pp. 416–428, 2019.

[50] R. Liu, G. Yu, F. Qu, and Z. Zhang, "Device-to-device communications in unlicensed spectrum: Mode selection and resource allocation," *IEEE Access*, vol. 4, pp. 4720–4729, 2016.

[51] H. Zhang, Y. Liao, and L. Song, "D2D-U: Device-to-Device communications in unlicensed bands for 5G system," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3507–3519, 2017.

[52] E. M. Mohamed, M. A. Abdelghany, and M. Zareei, "An efficient paradigm for multi-band WiGig D2D networks," *IEEE Access*, vol. 7, pp. 70 032–70 045, 2019.

[53] B. Ismaiel, M. Abolhasan, W. Ni, D. Smith, D. Franklin, E. Dutkiewicz, M. M. Krunz, and A. Jamalipour, "PCF-based LTE Wi-Fi aggregation for coordinating and offloading the cellular traffic to D2D network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12 193–12 203, Dec. 2018.

[54] A. Asadi and V. Mancuso, "Network-assisted outband D2D-clustering in 5G cellular networks: Theory and practice," *IEEE Trans. Mobile Comput.*, vol. 16, no. 8, pp. 2246–2259, Aug. 2017.

[55] Y. Niu, L. Su, C. Gao, Y. Li, D. Jin, and Z. Han, "Exploiting device-to-device communications to enhance spatial reuse for popular content downloading in directional mmWave small cells," *IEEE Trans. Veh. Tech.*, vol. 65, no. 7, pp. 5538–5550, July 2016.

[56] O. Liberg, M. Sundberg, E. Wang, J. Bergman, and J. Sachs, *Cellular Internet of Things Technologies, Standards, and Performance*. Academic Press, Sept. 2017.

[57] S. R. Borkar, "Long-term evolution for machines (LTE-M)," in *LPWAN Technologies for IoT and M2M Applications*, B. S. Chaudhari and M. Zennaro, Eds. Academic Press, 2020, pp. 145 – 166. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128188804000077

[58] S. R. Hussain, S. Mehnaz, S. Nirjon, and E. Bertino, "Secure seamless Bluetooth Low Energy connection migration for unmodified IoT devices," *IEEE Trans. Mobile Comput.*, vol. 17, no. 4, pp. 927–944, 2018.

[59] P. Mach, Z. Becvar, and T. Vanek, "In-band device-to-device communication in OFDMA cellular networks: A survey and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1885–1922, 2015.

[60] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, 2009.

[61] K. Ho and W. Xu, "An accurate algebraic solution for moving source location using TDOA and FDOA measurements," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2453–2463, 2004.

[62] F. Guo, Y. Fan, Y. Zhou, C. Xhou, and Q. Li, *Space Electronic Reconnaissance: Localization Theories and Methods*. Wiley Telecom, 2014.

[63] TR 36.763, "Study on narrow-band Internet of Things (NB-IoT) / enhanced machine type communication (eMTC) support for non-terrestrial networks (NTN)," 3GPP, Tech. Rep., Jan. 2021.

[64] Starlink. Accessed: Nov. 25, 2020. [Online]. Available: https://www.starlink.com/

[65] O. B. Osoro and E. J. Oughton, "A techno-economic framework for satellite networks applied to low earth orbit constellations: Assessing Starlink, OneWeb and Kuiper," *IEEE Access*, vol. 9, pp. 141 611–141 625, 2021.

[66] Ericsson, "R1-2005502 On UL time and frequency synchronization enhancements for NTN," 3GPP TSG RAN WG1 Meeting #102-e, Tech. Doc., Aug. 2020.

[67] ——, "R1-1909107 On frequency compensation, uplink timing and random access in NTN," 3GPP TSG RAN WG1 Meeting #98, Tech. Doc., Aug. 2019.

[68] M. Conti, A. Guidotti, C. Amatetti, and A. Vanelli-Coralli, "NB-IoT over non-terrestrial networks: Link budget analysis," in *IEEE Global Commun. Conf.*, 2020, pp. 1–6.

[69] M. Gineste, T. Deleu, M. Cohen, N. Chuberre, V. Saravanan, V. Frascolla, M. Mueck, E. C. Strinati, and E. Dutkiewicz, "Narrowband IoT service provision to 5G user equipment via a satellite component," in *IEEE Globecom Workshops (GC Workshops)*, 2017, pp. 1–4.

[70] G. Charbit, D. Lin, K. Medles, L. Li, and I.-K. Fu, "Space-terrestrial radio network integration for IoT," in *6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[71] M. Conti, S. Andrenacci, N. Maturo, S. Chatzinotas, and A. Vanelli-Coralli, "Doppler impact analysis for NB-IoT and satellite systems integration," in *Int. Conf. Commun. (ICC)*, 2020, pp. 1–7.

[72] O. Kodheli, N. Maturo, S. Chatzinotas, S. Andrenacci, and F. Zimmer, "On the random access procedure of NB-IoT non-terrestrial networks," in *Adv. Satellite Multimedia Syst. Conf. and Sig. Proc. Space Commun. Workshop (ASMS/SPSC)*. IEEE, 2020, pp. 1–8.

[73] H. Chougrani, S. Kisseleff, W. A. Martins, and S. Chatzinotas, "NB-IoT random access for non-terrestrial networks," *arXiv preprint arXiv:2101.08079*, 2021.

[74] O. Kodheli, A. Astro, J. Querol, M. Gholamian, S. Kumar, N. Maturo, and S. Chatzinotas, "Random access procedure over non-terrestrial networks: From theory to practice," *IEEE Access*, vol. 9, pp. 109 130–109 143, 2021.

[75] O. Kodheli, S. Andrenacci, N. Maturo, S. Chatzinotas, and F. Zimmer, "Resource allocation approach for differential doppler reduction in NB-IoT over LEO satellite," in *Adv. Satellite Multimedia Syst. Conf. and the Signal Proc. Space Commun. Workshop (ASMS/SPSC)*. IEEE, 2018, pp. 1–8.

[76] ——, "An uplink UE group-based scheduling technique for 5G mMTC systems over LEO satellite," *IEEE Access*, vol. 7, pp. 67 413–67 427, 2019.

[77] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, 2020.

[78] R. Barbau, V. Deslandes, G. Jakllari, J. Tronc, J.-F. Chouteau, and A.-L. Beylot, "NB-IoT over GEO satellite: Performance analysis," in *Adv. Satellite Multimedia Syst. Conf. and Sig. Proc. Space Commun. Workshop (ASMS/SPSC)*. IEEE, 2020, pp. 1–8.

[79] Qualcomm Incorporated, "R1-1705013: Uplink HARQ-ACK feedback," 3GPP TSG RAN WG1 Meeting #88bis, Tech. Doc., Apr. 2017.

[80] X. Lin, S. Rommer, S. Euler, E. A. Yavuz, and R. S. Karlsson, "5G from space: An overview of 3GPP non-terrestrial networks," *arXiv preprint arXiv:2103.09156*, 2021.

[81] TS 36.331 V15.4.0, "Evolved universal terrestrial radio access (E-UTRA) radio resource control (RRC) protocol specification (Release 15)," 3rd Generation Partnership Project (3GPP), Tech. Spec., Dec. 2018.

[82] A. Hoglund, J. Bergman, X. Lin, O. Liberg, A. Ratilainen, H. S. Razaghi, T. Tirronen, and E. A. Yavuz, "Overview of 3GPP Release 14 further enhanced MTC," *IEEE Commun. Standards Mag.*, vol. 2, no. 2, pp. 84–89, 2018.

[83] "Enabling global deployment of LTE IoT devices with SAW-less radio design architecture," White Paper, Sequans Communications, 2017.

[84] A. Hoglund, G. A. Medina-Acosta, S. N. K. Veedu, O. Liberg, T. Tirronen, E. A. Yavuz, and J. Bergman, "3GPP Release-16 preconfigured uplink resources for LTE-M and NB-IoT," *IEEE Commun. Standards Mag.*, vol. 4, no. 2, pp. 50–56, 2020.

[85] "MulteFire Release 1.1 technical overview," White Paper, MulteFire Alliance, Dec. 2018.

[86] "US Code of Federal Regulations," *Part 15, Section 15.247*, Sept. 2005.

[87] TR 36.889 V13.0.0, "Study on licensed-assisted access to unlicensed spectrum (Release 13)," 3rd Generation Partnership Project (3GPP), Tech. Rep., June 2015.

[88] *Short range devices (SRD) operating in the frequency range 25 MHz to 1000 MHz; Part 1: Technical characteristics and methods of measurement*, European Standard Std. ETSI EN 300 220-1 V3.1.1, Feb. 2017.

[89] *Short range devices (SRD) operating in the frequency range 25 MHz to 1 000 MHz; Part 2: Harmonised standard for access to radio spectrum for non specific radio equipment*, European Standard Std. ETSI EN 300 220-2 V3.2.1, June 2018.

[90] TR 36.888, "Study on provision of low-cost machine-type communications (MTC) user equipments (UEs) based on LTE (Release 12)," 3rd Generation Partnership Project (3GPP), Tech. Rep., June 2013.

[91] TR 45.820 V13.1.0, "Cellular system support for ultra-low complexity and low throughput internet of things (CIoT) (Release 13)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Nov. 2015.

[92] V. Krishnan, *Probability and random processes*. Wiley, 2015.

[93] TS 36.213 V15.3.0, "Evolved universal terrestrial radio access (E-UTRA); Physical layer procedures (Release 15)," 3rd Generation Partnership Project (3GPP), Tech. Spec., Sept. 2018.

[94] Ericsson, "R2-1803077 Report on MTC/NB-IoT padding issue in msg3," 3GPP TSG RAN WG2 Meeting #101, Tech. Doc., 2018.

[95] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," *in Proc. ACM Int. Conf. Mobile Sys., Appl., and Services (MobiSys)*, pp. 225–238, June 2012.

[96] TS 36.300 V16.2.0, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); Overall description; Stage 2 (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Spec., July 2020.

[97] A. T. Koc, S. C. Jha, R. Vannithamby, and M. Torlak, "Device power saving and latency optimization in LTE-A networks through DRX configuration," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2614–2625, 2014.

[98] C. Tseng, H. Wang, F. Kuo, K. Ting, H. Chen, and G. Chen, "Delay and power consumption in LTE/LTE-A DRX mechanism with mixed short and long cycles," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1721–1734, 2016.

[99] TS 36.304 V16.0.0, "Evolved universal terrestrial radio access (E-UTRA); User equipment (UE) procedures in idle mode," 3rd Generation Partnership Project (3GPP), Tech. Spec., Mar. 2020.

[100] TS 36.322 V15.1.0, "Evolved universal terrestrial radio access (E-UTRA) radio link control (RLC) protocol specification (Release 15)," 3rd Generation Partnership Project (3GPP), Tech. Spec., July 2018.

[101] G. C. Madueño, J. J. Nielsen, D. M. Kim, N. K. Pratas, Č. Stefanović, and P. Popovski, "Assessment of LTE wireless access for monitoring of energy distribution in the smart grid," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 675–688, 2016.

[102] Ericsson, "R1-1706161 Early data transmission for MTC," 3GPP TSG RAN WG1 Meeting #88bis, Tech. Doc., Apr. 2017.

[103] ——, "R1-1703112 On mMTC, NB-IoT and eMTC battery life evaluation," 3GPP TSG RAN WG1 Meeting #88, Tech. Doc., Apr. 2017.

[104] A. Dementyev, S. Hodges, S. Taylor, and J. Smith, "Power consumption analysis of Bluetooth Low Energy, ZigBee and ANT sensor nodes in a cyclic sleep scenario," in *IEEE Int. Wireless Symp. (IWS)*, 2013, pp. 1–4.

[105] J. Liu, C. Chen, and Y. Ma, "Modeling and performance analysis of device discovery in Bluetooth Low Energy networks," in *IEEE Global Commun. Conf. (GLOBECOM)*, 2012, pp. 1538–1543.

[106] R. Rondón, K. Landernäs, and M. Gidlund, "An analytical model of the effective delay performance for Bluetooth low energy," in *IEEE 27th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, 2016, pp. 1–6.

[107] J. Xu, J. Yao, L. Wang, Z. Ming, K. Wu, and L. Chen, "Narrowband Internet of Things: Evolutions, technologies, and open issues," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1449–1462, 2018.

[108] TR 37.985 V1.0.0, "Overall description of radio access network (RAN) aspects for vehicle-to-everything (V2X) based on LTE and NR (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Nov. 2019.

[109] LG Electronics, "RP-193231 New WID on NR sidelink enhancement," 3GPP TSG RAN Meeting #86, Tech. Doc., Dec. 2019.

[110] TR 36.814, "Evolved universal terrestrial radio access (E-UTRA) further advancements for E-UTRA physical layer aspects (Release 9)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Mar. 2017.

[111] *IEEE standard for information technology-telecommunications and information exchange between systems local and metropolitan area networks-specific requirements - Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, IEEE Standard Std. IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012), 2016.

[112] TR 38.855, "Study on NR positioning support (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Mar. 2019.

[113] Department of Defense, USA. (2008, Sept.) Global positioning system standard positioning service performance standard. Accessed: Sept. 12, 2022. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/16930

[114] TS 36.101 V16.4.0, "Evolved universal terrestrial radio access (E-UTRA) user equipment (UE) radio transmission and reception (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Spec., Dec. 2019.

[115] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: from 1G to 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1124–1148, 2018.

[116] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 70–84, 2005.

[117] Qualcomm Incorporated, "R1-161936 NB-PSS and NB-SSS design," 3GPP TSG RAN WG1 NB-IoT Ad-Hoc Meeting, Tech. Doc., Mar. 2016.

[118] Huawei and HiSilicon, "R1-2102344 Discussion on time and frequency synchronization enhancement for IoT in NTN," 3GPP TSG RAN WG1 Meeting #104bis-e, Tech. Doc., Apr. 2021.

[119] TS 38.101-1 V15.2.0, "Evolved universal terrestrial radio access (E-UTRA) user equipment (UE) radio transmission and reception part 1: Range 1 standalone (Release 15)," 3rd Generation Partnership Project (3GPP), Tech. Spec., June 2018.

[120] J. G. Walker, "Satellite constellations," *J. Brit. Interplanetary Soc.*, vol. 37, p. 559, Dec. 1984.

[121] S. W. Paek, S. Kim, and O. de Weck, "Optimization of reconfigurable satellite constellations using simulated annealing and genetic algorithm," *Sensors*, Feb. 2019.

[122] Huawei and HiSilicon, "R1-1911858 Discussion on performance evaluation for NTN," 3GPP TSG RAN WG1 Meeting #99, Tech. Doc., Nov. 2019.

[123] M. Kleder. WGS84 ellipsoidal earth plotted at 4 pixels per degree, MATLAB central file exchange. Retrieved: July 4, 2022. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/8590-wgs84-ellipsoidal-earth-plotted-at-4-pixels-per-degree

[124] MediaTek and Eutelsat, "R1-2005496 UL time and frequency synchronisation for NR-NTN," 3GPP TSG RAN WG1 Meeting #102-e, Tech. Doc., Aug. 2020.

[125] NORAD GP element sets current data. Accessed: Apr. 19, 2022. [Online]. Available: https://www.celestrak.com/norad/elements/

[126] G. Wang, Y. Li, and N. Ansari, "A semidefinite relaxation method for source localization using TDOA and FDOA measurements," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 853–862, 2013.

[127] R. Amiri, F. Behnia, and A. Noroozi, "An efficient estimator for TDOA-based source localization with minimum number of sensors," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2499–2502, 2018.

[128] L. Zhang, T. Zhang, and H.-S. Shin, "An efficient constrained weighted least squares method with bias reduction for TDOA-based localization," *IEEE Sensors J.*, vol. 21, no. 8, pp. 10 122–10 131, 2021.

[129] H. Yu, G. Huang, J. Gao, and B. Liu, "An efficient constrained weighted least squares algorithm for moving source location using TDOA and FDOA measurements," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 44–47, 2012.

[130] W. H. Foy, "Position-location solutions by Taylor-series estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-12, no. 2, pp. 187–194, 1976.

[131] B. Jin, X. Xu, and T. Zhang, "Robust time-difference-of-arrival (TDOA) localization using weighted least squares with cone tangent plane constraint," *Sensors*, vol. 18, no. 3, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/3/778

[132] X. Jin-yu, W. Wei, and Z. Zhong-liang, "A new TDOA location technique based on Taylor series expansion in cellular networks," in *Proc. 4th Int. Conf. Parallel Distrib. Comput. Appl. Technol.*, 2003, pp. 378–381.

[133] N. Levanon, "Quick position determination using 1 or 2 LEO satellites," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 3, pp. 736–754, 1998.

[134] R. Keating, M. Säily, J. Hulkkonen, and J. Karjalainen, "Overview of positioning in 5G new radio," in *16th Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2019, pp. 320–324.

[135] K. Shamaei, J. Khalife, and Z. M. Kassas, "Exploiting LTE signals for navigation: Theory to implementation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2173–2189, 2018.

[136] Qualcomm Incorporated, "R1-1802324 Reduced system acquisition time," 3GPP TSG RAN WG1 Meeting #92, Tech. Doc., Mar. 2018.

[137] A. Golnari, M. Shabany, A. Nezamalhosseini, and G. Gulak, "Design and implementation of time and frequency synchronization in LTE," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 12, pp. 2970–2982, 2015.

[138] X. Lai and H. Torp, "Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 46, no. 2, pp. 277–290, 1999.

[139] I. Céspedes, Y. Huang, J. Ophir, and S. Spratt, "Methods for estimation of subsample time delays of digitized echo signals," *Ultrason. Imag.*, Apr. 1995.

[140] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 12, pp. 1661–1669, 1987.

[141] MCC Support, "Final report of 3GPP TSG RAN WG1 #107-e v1.0.0," 3GPP TSG RAN WG1 Meeting #107-e, Tech. Doc., Jan. 2022.

[142] R. J. Carroll, C. F. J. Wu, and D. Ruppert, "The effect of estimating weights in weighted least squares," *J. Amer. Statist. Assoc.*, vol. 83, no. 404, pp. 1045–1054, 1988. [Online]. Available: http://www.jstor.org/stable/2290134

[143] TS 36.133 V17.6.0, "Evolved universal terrestrial radio access (E-UTRA) requirements for support of radio resource management (Release 17)," 3rd Generation Partnership Project (3GPP), Tech. Spec., June 2022.

[144] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Prentice Hall, 1993.

[145] S. Stein, "Algorithms for ambiguity function processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 3, pp. 588–599, June 1981.

[146] Y. Qi, H. Kobayashi, and H. Suda, "Analysis of wireless geolocation in a non-line-of-sight environment," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 672–681, Mar. 2006.

[147] Z. Şahinoğlu, S. Gezici, and I. Güvenç, *Ultra-wideband Positioning Systems: Theoretical Limits, Ranging Algorithms, and Protocols*. Cambridge University Press, 2008.

[148] Y. Qi and H. Kobayashi, "An optimum positioning receiver for nonsynchronized mobile systems," in *Proc. 37th Annu. Conf. Inf. Sci. Syst.*, Mar. 2003.

[149] Y. Shi, "Power saving methods for LTE-M and NB-IoT devices," White Paper, Rhode & Schwarz, June 2019.

[150] TR 38.913 V15.0.0, "Study on scenarios and requirements for next generation access technologies (Release 15)," 3rd Generation Partnership Project (3GPP), Tech. Rep., June 2018.

[151] MediaTek, "R1-2103962 Summary #3 of AI 8.15.1 Scenarios applicable to NB-IoT/eMTC," 3GPP TSG RAN WG1 Meeting #104bis-e, Tech. Doc., Apr. 2021.

[152] K. Scarbrough, R. J. Tremblay, and G. C. Carter, "Performance predictions for coherent and incoherent processing techniques of time delay estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 5, pp. 1191–1196, Oct. 1983.

[153] *Product Specification*, nRF9160, Nordic Semiconductor, Oct. 2021, v2.1.

[154] *Cellular IoT Module with BLE*, AES-CELLIOT-AVT9152MOD, Avnet, July 2021, rev 1.1.

[155] R. Fasthuber, F. Catthoor, P. Raghavan, and F. Naessens, *Energy-Efficient Communication Processors: Design and Implementation for Emerging Wireless Systems*. Springer, 2013.

[156] F. Rinaldi, H.-L. Maattanen, J. Torsner, S. Pizzi, S. Andreev, A. Iera, Y. Koucheryavy, and G. Araniti, "Non-terrestrial networks in 5G & beyond: A survey," *IEEE Access*, vol. 8, pp. 165 178–165 200, 2020.

[157] A. Sørensen, H. Wang, M. J. Remy, N. Kjettrup, R. B. Sørensen, J. J. Nielsen, P. Popovski, and G. C. Madueño, "Modelling and experimental validation for battery lifetime estimation in NB-IoT and LTE-M," *IEEE Internet Things J.*, 2022.

[158] K. A. Aldahdouh, K. A. Darabkh, W. Al-Sit *et al.*, "A survey of 5G emerging wireless technologies featuring LoRaWAN, Sigfox, NB-IoT and LTE-M," in *2019 IEEE Int. Conf. Wireless Commun., Sig. Proc. and Netw. (WiSPNET)*, 2019, pp. 561–566.

[159] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.

[160] Y.-P. 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 Commun. Mag.*, vol. 55, no. 3, pp. 117–123, 2017.

[161] G. Vos and S. Dost, "Method and apparatus for resource allocation for half duplex frequency division duplexing in a wireless communication system," Jan., 2021, United States Patent Application 17/066002.

[162] ZTE, "R1-1913238 Discussion on the delay-tolerant HARQ operation for NTN," 3GPP TSG RAN WG1 Meeting #99, Tech. Doc., Nov. 2019.

[163] MCC Support, "Final report of 3GPP TSG RAN WG1 #103-e v1.0.0," 3GPP TSG RAN WG1 Meeting #103-e, Tech. Doc., Nov. 2020.

[164] TR 38.901 V16.1.0, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Dec. 2019.

[165] Nokia, "R1-2103719 Correction of round trip delay drift rate for NTN scenarios," 3GPP TSG RAN1 #104bis-e, Tech. Doc., Apr. 2021.

[166] S. Moloudi, M. Mozaffari, S. N. K. Veedu, K. Kittichokechai, Y.-P. E. Wang, J. Bergman, and A. Höglund, "Coverage evaluation for 5G reduced capability new radio (NR-RedCap)," *IEEE Access*, vol. 9, pp. 45 055–45 067, 2021.

[167] "RAN1 chair's notes," 3GPP TSG RAN WG1 Meeting #110, Aug. 2022.

[168] Huawei, HiSilicon, "RP-222685 New SID on ambient IoT," 3GPP TSG RAN Meeting #97e, Tech. Doc., Sept. 2022.

[169] TR 22.840 V0.2.0, "Study on ambient power-enabled internet of things (Release 19)," 3rd Generation Partnership Project (3GPP), Tech. Rep., Sept. 2022.

Appendix A

First Order Taylor Series Approximation

Using Taylor series first order approximation, any nonlinear multivariable function $q(x, y, \ldots)$, which is differentiable at $[x_0, y_0, \ldots]$ can be expressed in linear form as

$$q(x, y, \ldots) = q(x_0, y_0, \ldots) + \left[\left.\frac{\partial q}{\partial x}\right|_{(x_0, y_0, \ldots)} \left.\frac{\partial q}{\partial y}\right|_{(x_0, y_0, \ldots)} \ldots \right] \times \left[x - x_0 \; y - y_0 \ldots\right]^{\mathrm{T}}. \quad \text{(A.1)}$$

Using (A.1), $f_i(\boldsymbol{X}, \boldsymbol{V})$ and $g_i(\boldsymbol{X}, \boldsymbol{V})$ in (4.25) and (4.26) can be expressed in linear form, which results in a set of $2(N-1)$ linear equations that can be represented in the form of (4.28). The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in (4.28) and (4.29) are given by

$$\boldsymbol{A} = \left.\tilde{\boldsymbol{A}}\right|_{\boldsymbol{X}=\boldsymbol{X}_0, \boldsymbol{V}=\boldsymbol{V}_0} \quad \text{(A.2)}$$

and

$$\boldsymbol{B} = \begin{bmatrix} r_{21} - (R_2 - R_1) \\ \vdots \\ r_{N\,N-1} - (R_N - R_{N-1}) \\ \dot{r}_{21} - (\dot{R}_2 - \dot{R}_1) \\ \vdots \\ \dot{r}_{N\,N-1} - (\dot{R}_N - \dot{R}_{N-1}) \end{bmatrix} + \boldsymbol{A}[\boldsymbol{X}_0 \; \boldsymbol{V}_0]^{\mathrm{T}}, \quad \text{(A.3)}$$

respectively, where

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \frac{\partial(f_2-f_1)}{\partial x} & \frac{\partial(f_2-f_1)}{\partial y} & \cdots & \frac{\partial(f_2-f_1)}{\partial v_z} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(f_N-f_{N-1})}{\partial x} & \frac{\partial(f_N-f_{N-1})}{\partial y} & \cdots & \frac{\partial(f_N-f_{N-1})}{\partial v_z} \\ \frac{\partial(g_2-g_1)}{\partial x} & \frac{\partial(g_2-g_1)}{\partial y} & \cdots & \frac{\partial(g_2-g_1)}{\partial v_z} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(g_N-g_{N-1})}{\partial x} & \frac{\partial(g_N-g_{N-1})}{\partial y} & \cdots & \frac{\partial(g_N-g_{N-1})}{\partial v_z} \end{bmatrix}, \tag{A.4}$$

$$R_i = f_i(\boldsymbol{X} = \boldsymbol{X}_0, \boldsymbol{V} = \boldsymbol{V}_0), \tag{A.5}$$

and

$$\dot{R}_i = g_i(\boldsymbol{X} = \boldsymbol{X}_0, \boldsymbol{V} = \boldsymbol{V}_0), \tag{A.6}$$

respectively. The differential terms in (A.4) are given by (A.7) to (A.18).

$$\frac{\partial(f_i - f_j)}{\partial x} = \left(\frac{x + v_x T_i - x_i}{R_i} - \frac{x + v_x T_j - x_j}{R_j}\right), \tag{A.7}$$

$$\frac{\partial(f_i - f_j)}{\partial y} = \left(\frac{y + v_y T_i - y_i}{R_i} - \frac{y + v_y T_j - y_j}{R_j}\right), \tag{A.8}$$

$$\frac{\partial(f_i - f_j)}{\partial z} = \left(\frac{z + v_z T_i - z_i}{R_i} - \frac{z + v_z T_j - z_j}{R_j}\right), \tag{A.9}$$

$$\frac{\partial(f_i - f_j)}{\partial v_x} = \left(\frac{T_i(x + v_x T_i - x_i)}{R_i} - \frac{x + v_x T_j - x_j}{R_j}\right), \tag{A.10}$$

$$\frac{\partial(f_i - f_j)}{\partial v_y} = \left(\frac{T_i(y + v_y T_i - y_i)}{R_i} - \frac{y + v_y T_j - y_j}{R_j}\right), \tag{A.11}$$

$$\frac{\partial(f_i - f_j)}{\partial v_z} = \left(\frac{T_i(z + v_z T_i - z_i)}{R_i} - \frac{z + v_z T_j - z_j}{R_j}\right), \tag{A.12}$$

$$\frac{\partial(g_i - g_j)}{\partial x} = \left(\frac{(v_{x,0} - v_{x_i})R_i^2 - R_i\dot{R}_i(x_0 + v_{x,0}T_i - x_i)}{R_i^3} + \frac{(v_{x,0} - v_{x_j})R_j^2 - R_j\dot{R}_j(x_0 + v_{x,0}T_j - x_j)}{R_j^3}\right), \tag{A.13}$$

$$\frac{\partial(g_i - g_j)}{\partial y} = \left(\frac{(v_{y,0} - v_{y_i})R_i^2 - R_i\dot{R}_i(y_0 + v_{y,0}T_i - y_i)}{R_i^3} + \frac{(v_{y,0} - v_{y_j})R_j^2 - R_j\dot{R}_j(y_0 + v_{y,0}T_j - y_j)}{R_j^3}\right), \tag{A.14}$$

$$\frac{\partial(g_i - g_j)}{\partial z} = \left(\frac{(v_{z,0} - v_{z_i})R_i^2 - R_i\dot{R}_i(z_0 + v_{z,0}T_i - z_i)}{R_i^3} + \frac{(v_{z,0} - v_{z_j})R_j^2 - R_j\dot{R}_j(z_0 + v_{z,0}T_j - z_j)}{R_j^3}\right), \tag{A.15}$$

$$\frac{\partial(g_i - g_j)}{\partial v_x} = \left(\frac{Q_{1,i}R_i^2 - R_i\dot{R}_iT_i(x_0 + T_iv_{x,0} - x_i)}{R_i^3} + \frac{Q_{1,j}R_j^2 - R_j\dot{R}_jT_j(x_0 + T_jv_{x,0} - x_j)}{R_j^3}\right), \tag{A.16}$$

$$\frac{\partial(g_i - g_j)}{\partial v_y} = \left(\frac{Q_{2,i}R_i^2 - R_i\dot{R}_iT_i(y_0 + T_iv_{y,0} - y_i)}{R_i^3} + \frac{Q_{2,j}R_j^2 - R_j\dot{R}_jT_j(y_0 + T_jv_{y,0} - y_j)}{R_j^3}\right), \tag{A.17}$$

and

$$\frac{\partial(g_i - g_j)}{\partial v_z} = \left(\frac{Q_{3,i}R_i^2 - R_i\dot{R}_iT_i(z_0 + T_iv_{z,0} - z_i)}{R_i^3} + \frac{Q_{3,j}R_j^2 - R_j\dot{R}_jT_j(z_0 + T_jv_{z,0} - z_j)}{R_j^3}\right), \tag{A.18}$$

where

$$T_i = i\Delta T,$$

$$Q_{1,i} = x_0 - x_i + 2T_i v_{x,0} - T_i v_{x_i},$$

$$Q_{2,i} = y_0 - y_i + 2T_i v_{y,0} - T_i v_{y_i},$$

and

$$Q_{3,i} = z_0 - z_i + 2T_i v_{z,0} - T_i v_{z_i}.$$

The weight matrix $\boldsymbol{W}$ is ideally a function of the covariance matrix of the measurements, given by

$$\boldsymbol{W} = \begin{bmatrix} c^2 \boldsymbol{W}_{\text{TDOA}}^{-1} & \mathbf{0}_{N-1\times N-1} \\ \\ \mathbf{0}_{N-1\times N-1} & \frac{c^2}{f_c^2} \boldsymbol{W}_{\text{FDOA}}^{-1} \end{bmatrix}^{-1}, \tag{A.19}$$

where $\boldsymbol{W}_{\text{TDOA}}^{-1}$ and $\boldsymbol{W}_{\text{FDOA}}^{-1}$ are given by

$$\boldsymbol{W}_{\text{TDOA}}^{-1} = \begin{bmatrix} \sigma_{\text{T},2}^2 + \sigma_{\text{T},1}^2 & \sigma_{\text{T},2}^2 & 0 & \dots & 0 \\ \sigma_{\text{T},2}^2 & \sigma_{\text{T},3}^2 + \sigma_{\text{T},2}^2 & \sigma_{\text{T},3}^2 & \dots & 0 \\ 0 & \sigma_{\text{T},3}^2 & \sigma_{\text{T},4}^2 + \sigma_{\text{T},3}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{\text{T},N}^2 + \sigma_{\text{T},N-1}^2 \end{bmatrix} \tag{A.20}$$

and

$$\boldsymbol{W}_{\text{FDOA}}^{-1} = \begin{bmatrix} \sigma_{\text{F},2}^2 + \sigma_{\text{F},1}^2 & \sigma_{\text{F},2}^2 & 0 & \dots & 0 \\ \sigma_{\text{F},2}^2 & \sigma_{\text{F},3}^2 + \sigma_{\text{F},2}^2 & \sigma_{\text{F},3}^2 & \dots & 0 \\ 0 & \sigma_{\text{F},3}^2 & \sigma_{\text{F},4}^2 + \sigma_{\text{F},3}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_{\text{F},N}^2 + \sigma_{\text{F},N-1}^2 \end{bmatrix}, \tag{A.21}$$

respectively, where $\sigma_{\text{T},i}^2$ and $\sigma_{\text{F},i}^2$ are the variance of estimation errors in ith TOA and FOA measurements as indicated in (4.30) and (4.31), respectively. However, in practice, the covariance matrix is unknown. For SPIN, since we have rough position and velocity estimates obtained from 2-WLS, we use them to compute initial value of $\boldsymbol{W}$. We update $\boldsymbol{W}$ after each TWLS iteration based on the updated position and velocity estimates.