THESIS UNDER JOINT SUPERVISION
UNIVERSITA' MEDITERRANEA DI REGGIO CALABRIA
UNIVERSITAT JAUME I

DOCTORAL SCHOOL OF UNIVERSITAT JAUME I
EUROPEAN JOINT DOCTORATE MARIE SKLODOWSKA-CURIE
IN A NETWORK FOR DYNAMIC WEARABLE APPLICATIONS
WITH PRIVACY CONSTRAINTS

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE,
DELLE INFRASTRUTTURE E DELL'ENERGIA SOSTENIBILE (DIIES)
PHD IN
INFORMATION ENGINEERING
ciclo XXXV

DIRECT COMMUNICATION RADIO INTERFACE FOR NEW RADIO MULTICASTING AND COOPERATIVE POSITIONING

CANDIDATE

NADEZHDA CHUKHNO

SUPERVISORS

Dr. Sergio Trilles Oliver (University Jaume I)

Assoc. Prof. Giuseppe Araniti (University Mediterranea of Reggio Calabria)

Dr. Joaquin Torres-Sospedra (University of Minho)

Prof. Antonio Iera (University of Calabria)

REGGIO CALABRIA, JANUARY 2023

Direct Communication Radio Interface for New Radio Multicasting and Cooperative Positioning

Doctoral thesis submitted by **NADEZDA CHUKHNO** in order to be eligible for a double doctoral degree awarded by the University Mediterranea of Reggio Calabria and Universitat Jaume I

Dipartimento di Ingegneria dell'Informazione, delle Infrastrutture e dell'Energia Sostenibile (DIIES), PhD Course in Information Engineering, S.S.D. ING-INF/02, XXXV CICLO

European Joint Doctorate Marie Sklodowska-Curie in A Network for Dynamic Wearable Applications with Privacy Constraints (A-WEAR), Universitat Jaume I Doctoral School

AUTHOR:

Nadezda Chukhno

SUPERVISORS:

Assoc. Prof. Giuseppe Araniti

Prof. Antonio Iera

Dr. Sergio Trilles Oliver

Dr. Joaquin Torres-Sospedra

Reggio Calabria (Italy), January 2023

This dissertation is funded by the European Union's Horizon 2020 Research and Innovation programme under the Marie Skłodowska-Curie grant agreements No. 813278 (A-WEAR: A network for dynamic wearable applications with privacy constraints, http://www.a-wear.eu/)

NADEZDA CHUKHNO

DIRECT COMMUNICATION RADIO INTERFACE FOR NEW RADIO MULTICASTING AND COOPERATIVE POSITIONING

The Teaching Staff of the PhD course in
INFORMATION ENGINEERING
consists of:

Antonio Iera (coordinator)

Pier Luigi Antonucci,
Giuseppe Araniti,
Francesco Buccafurri,
Claudia Campolo,
Giuseppe Coppola,
Mariantonia Cotronei,
Lorenzo Crocco,
Dominique Dallet,
Claudio De Capua,
Francesco Della Corte,
Giuliana Faggio,
Gioia Failla,
Fabio Filianoti,
Patrizia Frontera,
Sofia Giuffrè,
Giorgio Graditi,
Voicu Groza,
Tommaso Isernia,
Gianluca Lax,
Aime Lay Ekuakille,
Gaetano Licitra,
Antonella Molinaro,
Andrea Morabito,
Carlo Francesco Morabito,
Giacomo Morabito,
Rosario Morello,
Fortunato Pezzimenti,
Filippo Pratico',
Domenico Rosaci,
Giuseppe Ruggeri,
Mariateresa Russo,
Antonino Vitetta

"Family is the most important thing in the world."

Princess Diana

Acknowledgments

This Ph.D. thesis and my numerous publications would not have been possible without the support of many people. First and foremost, I would like to express my sincere gratitude to my supervisors - *Giuseppe Araniti, Antonio Iera, Antonella Molinaro, Sergio Trilles Oliver, and Joaquin Torres-Sospedra* for their assistance at every stage of the 3 years of my Ph.D. I also would like to thank *Simona Lohan and Alex Ometov* for their guidance and help throughout the A-WEAR project. My gratitude extends to *Claudia Campolo* for her valuable feedback on the research questions, and to *Sergey Andreev* and *Olga Galinina* for their patience and time spent teaching me.

I would like to offer my special thanks to *Sara Pizzi* (and to her family), to my friend and colleague in parallel, for her support and love in the real world, and for her insightful comments and suggestions in the work world.

I could not have undertaken this journey without my sister *Olga Chukhno*. I am thankful for her unwavering support, positive thinking, encouragement, and love. Special thanks here.

Many thanks to my Italian family, which I obtained, and I am sure, I will never lose *Alessia Ferraro, Gianluca Brancati, Chiara Suraci, Federica Rinaldi, Domenico Zappalà* (big thanks for Tigro), *Angelo Tropeano, Vincenzo Violi, Pasquale Scopelliti, Gianmarco Lia, Sabrina Zumbo, Giada Battaglia, and others.*

I am also grateful to my wise friend - *Olga Vikhrova* for her constant assistance and help that she was offering starting from the begging of my life in Reggio Calabria, and of course, for wise pieces of advice in any matter. Big thanks go to *Lucie* and *Roman Klus*, who are always ready to help me.

I cannot forget to thank *all my family members* and *my boyfriend* for their unconditional support, anytime help, for motivating me throughout my journey of Ph.D. and this thesis, and for always being there for me. I do not have the words to tell you how truly fortunate I feel to have you in my life. It is a huge source of comfort to know how much I am genuinely loved. Thank you so much!

Abstract

Recently, the popularity of Millimeter Wave (mmWave) wireless networks has increased due to their capability to cope with the escalation of mobile data demands caused by the unprecedented proliferation of smart devices in the fifth-generation (5G). Extremely high frequency or mmWave band is a fundamental pillar in the provision of the expected gigabit data rates. Hence, according to both academic and industrial communities, mmWave technology, e.g., 5G New Radio (NR) and WiGig (60 GHz), is considered as one of the main components of 5G and beyond networks. Particularly, the 3rd Generation Partnership Project (3GPP) provides for the use of licensed mmWave sub-bands for the 5G mmWave cellular networks, whereas IEEE actively explores the unlicensed band at 60 GHz for the next-generation wireless local area networks. In this regard, mmWave has been envisaged as a new technology layout for real-time heavy-traffic and wearable applications.

This very work is devoted to solving the problem of mmWave band communication system while enhancing its vantages through utilizing the direct communication radio interface for NR multicasting, cooperative positioning, and mission-critical applications. The main contributions presented in this work include: *(i)* a set of mathematical frameworks and simulation tools to characterize multicast traffic delivery in mmWave directional systems; *(ii)* sidelink relaying concept exploitation to deal with the channel condition deterioration of dynamic multicast systems and to ensure mission-critical and ultra-reliable low-latency communications; *(iii)* cooperative positioning techniques analysis for enhancing cellular positioning accuracy for 5G+ emerging applications that require not only improved communication characteristics but also precise localization.

Our study indicates the need for additional mechanisms/research that can be utilized: *(i)* to further improve multicasting performance in 5G/6G systems; *(ii)* to investigate sidelink aspects, including, but not limited to, standardization perspective and the next relay selection strategies; and *(iii)* to design cooperative positioning systems based on Device-to-Device (D2D) technology.

Index terms: 5G, New Radio, millimeter Wave, sidelink, device-to-device (D2D), multihop, public safety, factory automation, collaborative localization, cooperative localization, wearables.

Sommario

Di recente, la popolarità delle reti wireless Millimeter Wave (mmWave) è aumentata grazie alla loro capacità di far fronte all'aumento della richiesta di dati mobili causata dalla proliferazione senza precedenti di dispositivi intelligenti nelle reti di quinta generazione (5G). La banda ad altissima frequenza, o mmWave, è un pilastro fondamentale per consentire di fornire la velocità di trasmissione dati dell'ordine dei gigabit prevista. Pertanto, secondo le comunità accademiche e industriali, la tecnologia mmWave, come 5G New Radio (NR) e WiGig (60 GHz), è considerata uno dei componenti principali delle reti 5G e oltre. In particolare, all'interno del 3rd Generation Partnership Project (3GPP) si prevede l'uso di sottobande mmWave con licenza per le reti cellulari 5G mmWave, mentre la IEEE valuta la possibilità di sfruttare la banda senza licenza a 60 GHz per le reti locali wireless di prossima generazione. A questo proposito, mmWave è stato concepito come un nuovo layout tecnologico per applicazioni wearable e real-time ad alto carico.

Questo lavoro di tesi si focalizza sui sistemi di comunicazione in banda mmWave, allo scopo di migliorare al contempo i vantaggi derivanti dall'utilizzo dell'interfaccia radio di comunicazione diretta per il multicasting in sistemi NR, dal posizionamento cooperativo e dalle applicazioni mission-critical. I principali contributi presentati in questo lavoro includono: *(i)* un insieme di modelli matematici e strumenti di simulazione per caratterizzare la fornitura di traffico multicast in sistemi direzionali mmWave; *(ii)* l'utilizzo della possibilità di inoltrare dati mediante comunicazioni sidelink per far fronte al deterioramento delle condizioni del canale dei sistemi multicast dinamici e per garantire comunicazioni mission-critical e ultra-affidabili a bassa latenza; *(iii)* l'analisi di tecniche di posizionamento cooperativo per migliorare l'accuratezza del posizionamento cellulare per le emergenti applicazioni 5G+ che richiedono non solo migliori caratteristiche in termini di comunicazione ma anche una localizzazione precisa.

Lo studio condotto fa emergere la necessità di ulteriori meccanismi/ricerche che possano essere utilizzati per: *(i)* migliorare ulteriormente le prestazioni del multicasting nei sistemi 5G/6G; *(ii)* studiare aspetti secondari come, ad esempio, prospettive di standardizzazione e strategie per la selezione dei nodi che operano come relay, e *(iii)* progettare sistemi di posizionamento cooperativo basati sulla tecnologia Device-to-Device (D2D).

Parole chiave: 5G, New Radio, onde millimetriche, sidelink, device-to-device (D2D), multi-hop, sicurezza pubblica, automazione industriale, localizzazione collaborativa, localizzazione cooperativa, wearable.

Resumen

Recientemente, se ha incrementado la popularidad de las redes inalámbricas mmWave debido a su capacidad para hacer frente a la creciente demanda de datos móviles provocada por la proliferación sin precedentes de dispositivos inteligentes de 5G. La banda de frecuencias extremadamente altas o mmWave representa uno de los pilares fundamentales en la provisión de las esperadas velocidades de datos gigabit. Por tanto, según las comunidades académicas e industriales, la tecnología mmWave, por ejemplo, 5G NR y WiGig (60 GHz), se considera uno de los principales componentes de las redes 5G y posteriores. En particular, el 3GPP contempla el uso de sub-bandas mmWave con licencia para las redes de telefonía móvil mmWave de 5G, mientras que el IEEE explora activamente la banda sin licencia de los 60 GHz para las redes inalámbricas de área local de próxima generación. En este sentido, se ha planteado la mmWave como una nueva disposición tecnológica para aplicaciones de tráfico pesado y wearables en tiempo real.

Este trabajo tiene como objetivo resolver el problema del sistema de comunicación en la banda mmWave, mejorando sus ventajas mediante la utilización de la interfaz de radio de comunicación directa para la multidifusión NR, el posicionamiento cooperativo y las aplicaciones de misión crítica. Las principales contribuciones presentadas en este trabajo incluyen: *(i)* un conjunto de marcos matemáticos y herramientas de simulación para caracterizar la entrega de tráfico de multidifusión en sistemas direccionales mmWave; *(ii)* la explotación del concepto de retransmisión de enlace lateral para hacer frente al deterioro de las condiciones del canal de los sistemas de multidifusión dinámicos y para garantizar comunicaciones de misión crítica y ultra fiables de baja latencia; *(iii)* el análisis de las técnicas de posicionamiento cooperativo para mejorar la precisión del posicionamiento celular en aplicaciones emergentes 5G+ que requieren no solo mejores características de comunicación, sino también una localización precisa.

Nuestro estudio pone de manifiesto la necesidad de mecanismos o investigaciones adicionales que puedan utilizarse: *(i)* para mejorar aún más el rendimiento de la multidifusión en los sistemas 5G/6G; *(ii)* para investigar los aspectos de los enlaces laterales, incluyendo, entre otros, la perspectiva de la estandarización y las próximas estrategias de selección de relés; *(iii)* para ser considerados mientras se diseñan los sistemas de posición cooperativa basados en la tecnología D2D.

Palabras clave: 5G, New Radio, onda milimétrica, enlace lateral, dispositivo a dispositivo (D2D), saltos múltiples, seguridad pública, automatización de fábricas, localización colaborativa, localización cooperativa, dispositivos portátiles.

Contents

1 Introduction . . . 1

1.1 Motivation . . . 1

1.2 Objectives and Methodology . . . 3

1.3 Contributions . . . 4

1.3.1 Research outputs . . . 5

1.3.2 Publications . . . 8

1.4 Thesis outline . . . 10

2 Delivering Multicast Traffic in mmWave Systems . . . 11

2.1 Optimal Multicasting in 5G/6G mmWave Systems . . . 11

2.1.1 Motivation . . . 11

2.1.2 System Model . . . 12

Antenna Model . . . 12

Propagation Model . . . 13

Blockage Model . . . 15

2.1.3 Analysis . . . 16

Single-RAT Operation . . . 16

Multi-RAT Operation . . . 19

Solutions to Optimal Multicasting . . . 22

2.1.4 Performance Assessment . . . 33

Single-RAT . . . 33

Multi-RAT . . . 41

2.2 Conclusions . . . 43

3 5G NR Sidelink Multi-Hop Transmission . . . 47

3.1 Optimal Scheduling for Highly Directional Sidelink-assisted Multicasting . . . 47

3.1.1 Motivation . . . 47

3.1.2 System Model . . . 49

Deployment and Traffic Model . . . 49

Antenna Model . . . 49

Propagation and Blockage Model . . . 51
Mobility Model . . . 51
3.1.3 Analysis . . . 51
Framework Description at a Glance . . . 51
Step I – Multicast Group Formation . . . 53
Step II – Optimization . . . 55
Proposed Heuristic . . . 57
3.1.4 Performance Assessment . . . 59
Effect of Linkage Function . . . 60
Effect of Mobility . . . 61
Complexity vs. Energy Performance Trade-off . . . 61
Effect of Users' Distribution . . . 63
Heuristic Evaluation . . . 63
Effect of Transmit Power . . . 65
3.2 5G NR Sidelink Multi-Hoping in Public Safety and Factory Automation Scenarios . . . 67
3.2.1 Motivation . . . 67
3.2.2 NR Sidelink in a Nutshell: Why NR Sidelink? . . . 69
3.2.3 NR Sidelink as a Tool to Support Public Safety and Factory Automation Use Cases . . . 70
Public Safety . . . 71
Factory Automation . . . 72
Performance Indicators . . . 72
3.2.4 Performance Assessment . . . 73
Factory Automation Scenario . . . 74
Public Safety: Fire Brigade Scenario . . . 77
3.3 Conclusions . . . 78

4 Cooperative Positioning . . . 81
4.1 D2D-based Cooperative Positioning Paradigm for Future Wireless Systems . 81
4.1.1 Motivation . . . 81
4.1.2 5G Positioning Applications and Enabling Technologies . . . 82
Non-3GPP Technologies . . . 83
3GPP Technologies . . . 85
4.1.3 Positioning methods . . . 85
Indoor Localization vs. Outdoor Localization . . . 86
Indirect Localization vs. Direct Localization . . . 86
Active Localization vs. Passive Localization . . . 87
Centralized Localization vs. Distributed Localization vs. Decentralized 87
Absolute Localization vs. Relative Localization . . . 88

Non-cooperative Localization vs. Cooperative Localization 88
4.1.4 D2D-based Cooperative Positioning 88
D2D Technologies Comparison 89
Review of D2D-based Cooperative Positioning 90
4.1.5 Lessons Learned 96
4.2 A Theoretical Model for Cooperative RSSI-based Localization with D2D versus RIS 97
4.2.1 Motivation 97
4.2.2 Cellular BS-MT and Relay Model 99
D2D-aided Localization Using Power Measurements 99
RIS-aided Localization Using Power Measurements 103
4.2.3 Passive Positioning with RISs vs Relays 104
4.2.4 Performance Assessment 106
Simulation Setup 106
Equal Placement of RISs and Relays 109
Random Placement of Relays 109
Random Placement of Relays Within 20 m Radius 110
NLoS: Random Placement of Relays 112
Effects of the Number of RIS Elements 112
Summary 113
4.3 Conclusions 113

5 Conclusions 115
5.1 Summary 115
5.2 Future Research 116
5.2.1 mmWave Multicasting 116
5.2.2 Sidelink Relaying 117
5.2.3 Cooperative Positioning 118

References 123

List of Figures

2.1 Ratio of occupied to available resources as function of cell radius, $K = 10$, $C = 25$ Mbps, $W = 50$ MHz [1]. 34
2.2 Optimal number of beams in multi-beam system as function of cell radius, $K = 10$ UEs, $C = 25$ Mbps, $W = 50$ MHz [1]. 36
2.3 Average number of UEs per beam as function of cell radius, $K = 10$ UEs, $C = 25$ Mbps, $W = 50$ MHz [1]. 37
2.4 Ratio of occupied to available resources as function of number of UEs, $R = 250$ m, $C = 25$ Mbps, $W = 50, 100, 200$ MHz, $L = 1$ beam [1]. 37
2.5 Subgroup assignment accuracy, σ, for $H_2 = 5000$, $R = 250\,\mathrm{m}$, $K = 10$ [2]. 38
2.6 Subgroup assignment accuracy, σ, for $H_1 = H_2 = 5000$, $K = 13$ [2]. 39
2.7 Variables' importance estimates [2]. 40
2.8 Performance metrics when mmWave resources are utilized whenever possible (mmWave RAT priority): mmWave – $\mu_m = 3$, μWave – $\mu_\mu = 0$. 41
2.9 Performance metrics for weighted optimization function: mmWave – $\mu_m = 3$, μWave – $\mu_\mu = 0$. 41

3.1 Sidelink-assisted multicast system illustration. 50
3.2 Flow diagram of proposal. 52
3.3 Impact of linkage functions on energy consumption for hierarchical clustering: (a) over 200 time slots, (b) mean value. Uniform user distribution. 60
3.4 Energy consumption over time for pedestrian mobility when considering: (a) uniform and (b) 2 clusters. Black lines are drawn in case MGF is rerun at every time slot. 62
3.5 Energy consumption for multicast users moving with different speeds in case of uniform distribution (each second bar is when we rerun MGF at every time slot). 62
3.6 Complexity gap vs. energy performance gain compared to no MGF rerunning for uniform distribution of users: (a) MGF, (b) SA3M without restrictions (c) SA3M with restrictions. 63

3.7 Effect of different distributions of multicast users on (a) energy consumption, (b) network throughput, and (c) energy efficiency. MGPF is executed at every time slot. 64
3.8 Energy consumption for heuristic solution varying $S_{\text{thr,h}}$ and D2D_{thr} thresholds. 64
3.9 Heuristic compared to MGF and SA3M algorithms for (a) CQI 6, (b) CQI 7, and (c) CQI 8. D2D_{thr}=50 m. Effect of different distributions of multicast users on (a) energy consumption, (b) network throughput, and (c) energy efficiency. MGPF is executed at every time slot. 64
3.10 Effect of transmit power on (a) energy consumption, (b) latency, and (c) network throughput. Heuristic with D2D_{thr}=50 m, MCS 8, and $W = 1$ GHz. . 66
3.11 Effect of transmit power on (a) energy consumption, (b) latency, and (c) network throughput. Heuristic with D2D_{thr}=50 m, MCS 8, and $W = 400$ MHz. 66
3.12 Public safety use case illustration. "R" stand for a relay. "R" can transmit data via unicast and groupcast. In this work, we use only unicast. Note that broadcast is only used for LTE. 71
3.13 Factory automation use case illustration. 72
3.14 Multi-hop establishment illustration: (a) chain, (b) concurrent transmissions. 75
3.15 Factory automation: latency as a function of packet size (blue); network throughput as a function of packet size (red). 76
3.16 Public safety (packet size ranges from 10 to 300 byte): energy efficiency as a function of packet size (blue); power consumption as a function of packet size (red). 78

4.1 D2D-aided positioning illustration. 89
4.2 LoS probability as a function of 2D distance between the BS and the MT according to 3GPP UMi Street Canyon model. 102
4.3 SNR comparison of RIS and relays. 106
4.4 Visualisation of devices deployments: considered use cases. 107
4.5 Visualisation of median and mean on skewed and normal distributions. 108
4.6 Visualisation of triangulation. Pink and blue circles are real and estimated positions. 109
4.7 CDF of individual error. 110
4.8 CDF of individual error, SNR for RIS with different number of antenna elements, LoS, PLE=2.1, SF=4. 110

List of Tables

2.1 Notation and parameters used in this chapter. 14
2.2 Parameters induced by 5G NR BS antenna arrays. 15
2.3 Interpretability characteristics of main classifier types. 29
2.4 Algorithms' execution time. 34
2.5 Subgroup and resource matching accuracy, $H_1 = 5000$, $H_2 = 5000$, $K = 13$. . 39

3.1 System modeling notation 50
3.2 5G NR numerology and subcarrier spacing [3]. 55
3.3 Default parameters for numerical evaluation. 60
3.4 Algorithms' Complexity, seconds. 62
3.5 CQI, MCS, spectral efficiency, and SNR mapping for 3GPP NR. 65
3.6 Simulation parameters factory automation scenario 74

4.1 Vertical Use Cases for Localization and Requirements [4, 5] 84
4.2 D2D Technologies Comparison 91
4.3 The State-of-the-art on D2D-based Indoor Positioning 92
4.4 The State-of-the-art on D2D-based Outdoor Positioning 93
4.5 Main Notation for Cooperative RSSI-based Localization 100
4.6 Propagation Models [6–8] 101
4.7 RIS vs Relay 105
4.8 Default Parameters 108
4.9 Individual errors, [m] 111

Abbreviations

2D	two-dimensional
3D	three-dimensional
2G	second-generation mobile telecommunications
3G	third-generation mobile telecommunications
3GPP	3rd Generation Partnership Project
4G	fourth-generation mobile telecommunications
5G	fifth-generation
6G	sixth-generation
AI	Artificial Intelligence
AOA	Angle Of Arrival
AOD	Angle Of Departure
B5G	beyond 5G
BLE	Bluetooth Low Energy
BS	Base Station
BPP	Bin Packing Problem
CAM	Cooperative Awareness Message
CQI	Channel Quality Indicator
D2D	Device-to-Device
DENM	Decentralized Environmental Notification Message
DL	Downlink
eMBB	Enhanced Mobile Broadband
FD	Full-Duplex
FOA	frequency of arrival
FDOA	frequency difference of arrival
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
GNR	Gain-to-Noise Ratio
GPS	Global Positioning System
HARQ	Hybrid Automatic Repeat Request
HD	Half-Duplex

HPBW Half Power Beamwidth
HSFM Headed Social Force Model
IAB Integrated Access and Backhaul
IEEE Institute of Electrical and Electronics Engineers
IIPS Infrastructure-based Indoor Positioning System
IIoT Industrial Internet of Things
IoT Internet of Things
ITS Intelligent Transportation System
KPI Key Performance Indicator
kNN K-Nearest Neighbors
LBS Location-based Services
LB Local Branching
LoS Line-of-Sight
LTE Long Term Evolution
LTE-A Long-Term Evolution - Advanced
MARL Multi-Agent Reinforcement Learning
MBMS Multimedia Broadcast Multicast Services
MCS Modulation and Coding Scheme
MCPTT Mission-Critical Push-to-Talk
MCData Mission-Critical Data
MCVideo Mission-Critical Video
MEC Mobile Edge Computing
ML Machine Learning
MIMO Multiple-Input–Multiple-Output
MIP Mixed-Integer Programming
MKF Multi-Kalman Filter
mmWave Millimeter Wave
MSE Mean Squared Error
MGF multicast group formation
MT Mobile Terminal
NLoS Non-Line-of-Sight
NR New Radio
NT Network Throughput
NFV Network Function Virtualization
NN Neural Network
OTDOA/UTDOA Observed/Uplink Time Difference of Arrival
O2I Outdoor-to-Indoor
PLE Path Loss Exponent
PPP Precise Point Positioning
PPP Poisson Point Process

PTM Point-to-Multipoint
PSSCH Physical Sidelink Shared Channel
PSCCH Physical Sidelink Control Channel
PRB Primary Resource Block
QoS Quality of Service
RAT Radio Access Technology
RL Reinforcement Learning
RF Radio Frequency
RFID Radio Frequency Identification
RSS Received Signal Strength
RIS Reconfigurable Intelligent Surface
RINS Relaxation-Induced Neighborhood Search Heuristic
RUS Reflection Unit Set
RMSE Root Mean Squared Error
RTK Real-Time Kinematic
SA3M Sidelink-Assisted Multiple Modes mmWave
SDN Software-Defined Networks
SNR Signal-to-Noise Ratio
SVM Support Vector Machine
SISO Single-Input Single-Output
TBS Terrestrial Beacon Systems
TDOA Time Difference Of Arrival
THz Terahertz
TOA Time Of Arrival
TTFF Time to First Fix
UAV Unmanned Aerial Vehicle
UE User Equipment
UL Uplink
UMi Urban Microcell
URLLC ultra-Reliable Low-Latency Communications
UWB Ultra-Wide Band
UTDOA Uplink TDOA
V2V Vehicle-to-Vehicle
V2X Vehicle-to-Everything
VR Virtual Reality
Wi-Fi Wireless Fidelity
WLAN Wireless Local Area Network
WPAN Wireless Personal Area Network
WSN Wireless Sensor Network
XR Extended Reality

1

Introduction

In this chapter, we explain the motivation of this research and its aim. We define central research objectives and summarize our contributions. The chapter also covers our research methodology and the structure of the thesis.

1.1 Motivation

mmWave band transmissions allow wireless technologies to meet the high data rate requirements of bandwidth-hungry applications, such as Extended Reality (XR) and multimedia services. This is one of the main advantages of 5G NR mmWave small cells, which are considered as one of the main components of future 5G+ networks [9]. Using multicast via Point-to-Multipoint (PTM) communications in these small cells may help to improve further the spectrum efficiency. Multicasting, which is under consideration by the 3GPP for Release 17 [10] of 5G systems, can provide substantial improvements in terms of system efficiency, user experience, and total network throughput [11], which is a critical feature for ultra-high-speed data transmissions.

Although multicasting has been widely investigated in traditional omnidirectional communications (i.e., at sub-6 GHz bands), the *design of efficient directional mmWave multicasting* techniques has to account for the limited coverage of directional mmWave communications [12]. Since mmWave is prone to blockages and suffers from high propagation loss, it can severely affect the performance of the multicast link. In case one user in a multicast group suffers from blockage, two options are possible: *(i)* all users experience this poor channel condition (human blockage takes 15-25 dB [13–15] from the signal-to-noise ratio), or *(ii)* the blocked user is served by the base station through unicast communication. In the thesis, **we take heed of the mentioned gap in supporting multicast traffic delivery in directional systems by offering novel optimal and heuristic multicast delivery and group formation strategies.**

Moreover, in practical scenarios, the problem of multicasting with directional beams calls for new strategies that are simpler to compute, compared to the non-polynomial optimal ones, while still guaranteeing near-to-optimal performance [16]. For this purpose, we also use machine learning approaches to provide close-to-optimal solutions very fast.

Furthermore, multi-hop relaying schemes are considered as one of the key interest areas in future 5G+ systems. With D2D communications enabled, users close to the base station

can serve as relays towards cell-edge users in their proximity, interested in the same multicast content, by using more robust D2D links [17, 18]. Hence, several non-adjacent links may be active at the same time, thus enabling concurrent transmissions to achieve better system performance. It has been demonstrated that mmWave and D2D symbiosis can bring throughput performance improvement up to 2.3 times [19]. Furthermore, concurrent transmissions and D2D-enabled communications in directional multicast systems help to reduce energy consumption, as required by battery-constrained wearable devices [20, 21].

To this end, D2D-aided multicasting in mmWave directional systems has been investigated in several recent studies. In [22], an efficient heuristic is designed for multicast data delivery, where D2D multi-hop and concurrent transmissions are jointly exploited to achieve lower energy consumption compared to a series of unicast transmissions. More recently, in [23] and [20], an optimal multicast scheduling problem is formulated, with D2D links and concurrent transmissions, through a mixed-integer non-linear program, which is known to be NP-hard. Heuristic solutions with cubic complexity are also designed. A similar approach is proposed in [24], where multicast scheduling jointly exploits relaying and spatial sharing properties of mmWave networks to minimize the overall data delivery time. In [21], an optimal D2D-enabled multicast scheduling policy is proposed by constructing an integer linear program problem with the goal of minimizing energy consumption in mmWave cellular networks.

Here we claim that in a single-beam system, D2D links can improve performance compared to unicast communications in terms of transmission delay, energy consumption, and overall network throughput. Hence, due to the nature of multicast transmissions where the users with the worst channel conditions define the data rate of the whole group (that becomes more critical in dynamic scenarios), one may need to exploit relaying functionality to maintain the link quality of multicast transmissions or/and improve it employing D2D. **This thesis solves this problem through the mix of optimal and machine learning methods.** In addition, existing studies only investigate the special case of one relay node (two-hop) sidelink operation [25–28], we investigate the case of an arbitrary number of hops considering mission-critical scenarios.

While all mentioned before research gaps were connected to the communication perspective, **the thesis also addresses the research gap in improving 5G localization accuracy demanded by emerging applications**. The Non-Line-of-Sight (NLoS) problem that may appear in high-frequency bands, such as mmWave, is one of the most challenging problems for sixth-generation (6G), which can drastically reduce localization accuracy [29–31]. To this end, the cooperative positioning technique can be used. In this thesis, we exploit the two technologies to implement cooperative localization and provide a performance comparison from communication and localization views.

1.2 Objectives and Methodology

Based on the challenges and research gaps identified in Section 1.1, three **research objectives** (ROx) have been formulated and subsequently addressed to answer the **central research question** of this thesis *how to schedule the resources for highly directional multicast and sidelink-assisted multicast communications efficiently, and whether D2D and RIS technology can help positioning by increasing its accuracy.*

- **RO1. Support multicast traffic in mmWave networks with single- and multi-beam directional antennas.**
  Transmissions at mmWave use highly directional antennas to guarantee the gigabit capabilities and overcome the short propagation range, thereby suffering from the limited coverage caused by the oxygen absorption and severe path loss and making the multicast fashion more complex. The former drawback makes it unfeasible to serve users spread over large regions at a time with one beam due to the decrease in antenna gain. The latter is an effect of the directionality of mmWave systems, which complicates multicast deployment by posing additional challenges (e.g., beam steering and proper selection of beamwidth). Hence, the proper beamwidth and data rate setting is one of the most challenging issues in multicast with directional antennas. This calls for new optimal solutions, heuristic strategies, and statistical learning methods to properly schedule the transmissions of multicast traffic.
- **RO2. Improve the performance of mmWave multicast directional systems for dynamic scenarios and for mission-critical services.**
  An additional challenge emerges in directional multicast systems in the case of a *non-static scenario*. In mmWave multicast systems, beams are steered in between users to cover multiple receivers at a time, leading to signal deterioration or even connection disruption between the Base Station (BS) and mobile receiver if the latter resides close to the beam edge. Therefore, guaranteeing coverage in the presence of moving users is becoming increasingly challenging. Even though research efforts have also been put on issues related to unicast mobility and beam tracking, group mobility has received low interest from the community. To this end, this objective studies the use of sidelink to assist mmWave multicasting, thereby improving performance. Moreover, this objective reviews the NR sidelink applicability for public safety and factory automation applications, thereby covering mission-critical and ultra-Reliable Low-Latency Communications (URLLC) scenarios.
- **RO3. Improve the performance of cellular positioning and analyze cooperative positioning methods and their impact on accuracy and coverage.**
  Acquiring accurate location information from Mobile Terminals (MTs) is becoming increasingly crucial for achieving 5G and 6G applications requirements. And this is not only true for location-based services but also for improving wireless communication performance in various ways, including channel estimation, beam alignment, medium access

control, routing, and network optimization. In this vein, inter-agent measurements and indirect Reconfigurable Intelligent Surface (RIS) links appear to provide additional position information and virtual Line-of-Sight (LoS), hence boosting localization accuracy. Although much effort has been devoted to evaluating the performance of D2D and RIS assisted localization systems separately, a comparison of these technologies from a localization perspective and the synergies resulting from their *rendezvous* have not been sufficiently investigated despite the fact that they may produce a new perspective for research and industrial communities. This objective aims to analyze cooperative positioning methods utilizing D2D and RIS technologies and their impact on accuracy and coverage.

1.3 Contributions

The main contributions of the thesis can be summarised as follows:

1. Chapter 2:
   - **optimal solution** for multi-beam mmWave BS operation minimizing the amount of resources required to serve User Equipments (UEs) based on multi-period variable cost and size bin packing problem;
   - **heuristic algorithms** characterized by polynomial complexity and allowing to achieve close approximations of the optimal solution;
   - problem formalization and computation of the exact (**globally optimal**) and **approximate simulated annealing** solution for multicast optimization in dual-mode mmWave/μWave BS deployments;
   - implementation and comparison of the set of **machine learning algorithms**;
2. Chapter 3:
   - **design of a multicast services delivery framework**, composed of two steps, that considers a multi-mode approach (sidelink, mmWave unicasting, and multicasting);
   - **formulation of the transmission scheduling problem** solved by an optimization problem that aims at maximizing the network throughput;
   - **exploitation of a machine learning technique for clustering users**, working as a pre-optimization step able to reduce the complexity of the optimal solution.
   - **provision of a low-complexity heuristic** solution, which offers results comparable to the optimal scheduling.
   - **analysis of complexity vs. performance trade-off** while accounting for the user's mobility and provision of practical hints to reach the trade-off.
   - presentation of numerical results demonstrating that **transmit power together with transmission bandwidth can be adjusted to reduce the total power consumption**.
   - an **overview of the main functionalities** and features of NR sidelink;

- **possibility of using NR sidelink communication for public safety and factory automation** scenarios, demonstrating noticeable end-to-end latency and energy efficiency performance improvement;
- investigation of **multi-hop sidelink** operation;

3. Chapter 4:
   - detailed **state-of-the-art of D2D-aided cooperative positioning**, which can be determined as the information exchange among MTs intending to increase their localization accuracy;
   - **system-level modeling** for D2D- and RIS-based positioning systems for Received Signal Strength (RSS) based localization. RSS readings observed at the MT can be employed to estimate the corresponding distances from surrounding BSs through mathematical models (known as path loss models) that describe signal attenuation as a function of distance;
   - **novel approach to integrate** D2D and RIS communication model into positioning algorithms by using the statistical mean and median for sensor fusion at the location-level;
   - **identification of relevant use cases** for RIS-aided localization and RIS communications in general.

1.3.1 Research outputs

The detailed contributions and relevant publications are organized in three blocks, each corresponding to a research objective from the list presented in Section 1.2. Contributions C1 are included in chapter 2, chapter 3 contains our contributions C2, and chapter 4 describes contributions C3.

C1. Contributions in supporting multicast traffic in mmWave networks with single- and multi-beam directional antennas by proposing optimal, heuristic, and statistical methods solutions.

Multicasting is becoming more and more important in the Internet of Things (IoT) and wearable applications (e.g., high-definition video streaming, virtual reality gaming, and public safety, among others) that require high bandwidth efficiency and low energy consumption. In this regard, mmWave communications can play a crucial role in efficiently disseminating large volumes of data as well as enhancing the throughput gain in 5G and beyond networks. There are, however, challenges to face in view of providing multicast services with high data rates under the conditions of short propagation range caused by high path loss at mmWave frequencies. Indeed, the strong directionality required at extremely high frequency bands excludes the possibility of serving all multicast users via a single transmission. Therefore, multicasting in directional systems consists of a sequence of beamformed transmissions to serve all multicast group members, subgroup by subgroup. We address the challenge of optimal multicasting in 5G mmWave systems by presenting a globally optimal solution for

multi-beam antenna operation. The optimization problem is formulated as a special case of multi-period variable cost and size bin packing problem that allows not imposing any constraints on the number of the beams and their configurations. We also propose heuristic solutions having polynomial time complexity.

On top of this, new deployment options, such as dual mmWave and microwave (μWave) deployments and new antenna design solutions, add to the complexity of the problem. To this end, the resource allocation task for multicast services in dual mmWave/μWave deployments with multi-beam directional antennas is addressed as a multi-period variable cost and size bin packing problem is studied. This latter is solved, and the globally optimal solution is characterized. To decrease complexity, we also propose and test the simulated annealing approximation as well as two relaxation techniques.

However, the bin packing problem is known to be NP-hard, and the solution time is practically unacceptable for large multicast group sizes. To this aim, we further develop and test several machine learning alternatives to address this issue. The numerical analysis shows that there is a trade-off between accuracy and computational complexity for multicast grouping when using decision tree-based algorithms. A higher number of splits offers better performance (i.e., prediction accuracy) at the cost of an increased computational time.

These contributions have been included in the following publications:

- Chukhno, N, Chukhno, O., Moltchanov, D., Molinaro, A., Gaidamaka, Yu., Samouylov, K., Koucheryavy, Ye., and Araniti, G. "Optimal Multicasting in Millimeter Wave 5G NR with Multi-beam Directional Antennas." IEEE Transactions on Mobile Computing, 2021.
- Chukhno, N., Chukhno, O., Moltchanov, D., Gaydamaka, A., Samuylov, A., Koucheryavy, Y., Molinaro, A., Iera, A. and Araniti, G. "Optimal Multicasting in Dual mmWave/μWave 5G NR Deployments with Multi-Beam Directional Antennas." IEEE Transactions on Vehicular Technology, 2022, (submitted).
- Chukhno, N., Chukhno, O., Moltchanov, D., Gaydamaka, A., Samuylov, A., Koucheryavy, Y., Molinaro, A., Iera, A. and Araniti, G. "The Use of Machine Learning Techniques for Optimal Multicasting in 5G NR Systems." IEEE Transactions on Broadcasting, 2022, (Early Access).

C2. Contributions in improving performance of mmWave multicast directional systems in the presence of dynamic users/blockers by designing a multicast services delivery framework enhanced by D2D links and further elaboration on the concept of multi-hop D2D ProSe communications while referring to public safety and factory automation sample use cases.

Recent years' technological developments in the field of telecommunications have brought revolutionary modifications and enhancements also in wireless communications systems. Among these, the exploitation of the mmWave spectrum stands out for its capability to provide ultra-high transmission data rates. However, its full adoption beyond the fifth generation (5G+/6G) multicast-capable systems remains hindered, mainly due to mobility ro-

bustness issues. We address this issue by proposing an optimal "sidelink-assisted" multicast scheduling in a mobile multi-mode system, including sidelink/D2D, unicast, and multicast transmissions. As a pre-optimization step of the optimal solution for the sidelink-assisted multiple modes mmWave scheduling, we introduce a clustering technique based on machine learning and perform a detailed analysis of the effects of various system parameters on performances. Particularly, we numerically show the trade-off between complexity and performance depending on the mobility pattern and users' distribution. Our results also demonstrate that energy consumption/latency trade-off can be reached via proper transmit power and bandwidth adjustments.

Further, we study the deployment of D2D communications (also known as ProSe or sidelink transmissions) in cellular networks, taking advantage of proximity, multi-hop, and spatial reuse gains. We describe the main advancements of NR sidelink compared to Long-Term Evolution - Advanced (LTE-A) sidelink. Then, we run a simulation campaign to test D2D-based ProSe for public safety and factory automation scenarios with their mission-critical requirements and ultra-reliable low-latency communications, respectively. A preliminary study on NR sidelink usage for both considered use cases is performed, aiming to identify the main advantages and disadvantages thereof.

Our research on sidelink relaying functionalities has been published in the following papers:

- Chukhno, N., Chukhno, O., Pizzi, S., Molinaro, A., Iera, A., Araniti, G. "Optimal Scheduling for Highly Directional Sidelink-assisted Multicast Communications". IEEE Transactions on Vehicular Technology Journal, 2022, (submitted).
- Chukhno, N., Orsino, A., Torsner, J., Iera, A., Araniti, G. "5G NR Sidelink Multi-Hop Transmission in Public Safety and Factory Automation Scenarios". IEEE Network, 2023.

C3. Contributions in analyzing cooperative positioning methods by means of D2D and RIS technologies and their impact on accuracy and coverage, as well as in proposing a loose-coupling sensor fusion based on a statistical metric.

The next generation of high-accuracy positioning services are required to satisfy the sub-meter accuracy level for more than 95 % of the network area, including indoor, outdoor, and urban deployments. For this purpose, 5G NR technology is designed to facilitate high-accuracy continuous localization. In 5G systems, the existence of high-density small cells and the possibility of the D2D communication between mobile terminals paves the way for cooperative positioning applications. From the standardization perspective, 5G technology is already under consideration (5G NR Release 16) for ultra-dense networks enabling cooperative positioning and is expected to achieve the ubiquitous positioning of below one-meter accuracy, thereby fulfilling the 5G requirements. We analyze the strengths and weaknesses of D2D as an enabling technology for cooperative cellular positioning (including two D2D approaches to perform cooperative positioning).

Further, we research cooperative positioning techniques by means of D2D and RIS technologies (to ensure virtual LoS transmissions) leveraging RSS-based ranging. In particular, the theoretical models for D2D- and RIS-aided positioning are provided, and the pros and cons of each of the two methods are discussed from both a communication and a localization perspective. Numerical results reveal the use cases advantageous for RIS and D2D usage. Then, based on the results, useful guidelines are derived on the optimal sensor fusion metric – median – that minimizes the mean absolute error of the cooperative localization.

Reproducible Research: The open-source code is available from `https://github.com/NadezhdaChukhno/Cooperative-RSSI-based-Localization-using-cellular-data-D2D-and-RIS`.

These contributions have been published in the following papers.

- Chukhno, N., Trilles, S., Torres-Sospedra, J., Iera, A., and Araniti, G. "D2D-based cooperative positioning paradigm for future wireless systems: A survey." IEEE Sensors Journal 22, no. 6 (2021): 5101-5112.
- Chukhno, N., Trilles, S., Torres-Sospedra, J., Iera, A., and Araniti, G. "D2D-aided versus RIS-aided Cooperative Positioning: Theoretical Model for RSSI-based Ranging and Performance Comparison." IEEE Transactions on Vehicular Technology Journal, 2022, (submitted).

1.3.2 Publications

The complete list of the author's publications produced during the Ph.D. period includes 7 papers related to the subject of the thesis and mentioned in Section 1.3.1, and 10 papers not included in the thesis.

1. Chukhno, N, Chukhno, O., Moltchanov, D., Molinaro, A., Gaidamaka, Yu., Samouylov, K., Koucheryavy, Ye., and Araniti, G. Optimal Multicasting in Millimeter Wave 5G NR with Multi-beam Directional Antennas. IEEE Transactions on Mobile Computing (2021).
2. Chukhno, N., Chukhno, O., Moltchanov, D., Gaydamaka, A., Samuylov, A., Koucheryavy, Y., Molinaro, A., Iera, A. and Araniti, G. "Optimal Multicasting in Dual mmWave/μWave 5G NR Deployments with Multi-Beam Directional Antennas." IEEE Transactions on Vehicular Technology, 2022, (submitted).
3. Chukhno, N., Chukhno, O., Moltchanov, D., Gaydamaka, A., Samuylov, A., Koucheryavy, Y., Molinaro, A., Iera, A. and Araniti, G. "The Use of Machine Learning Techniques for Optimal Multicasting in 5G NR Systems". IEEE Transactions on Broadcasting, 2022, (Early Access).
4. Chukhno, N., Chukhno, O., Pizzi, S., Molinaro, A., Iera, A., Araniti, G. "Optimal Scheduling for Highly Directional Sidelink-assisted Multicast Communications". IEEE Transactions on Vehicular Technology Journal, 2022, (submitted).
5. Chukhno, N., Orsino, A., Torsner, J., Iera, A., Araniti, G. "5G NR Sidelink Multi-Hop Transmission in Public Safety and Factory Automation Scenarios". IEEE Network, 2023.

6. Chukhno, N., Trilles, S., Torres-Sospedra, J., Iera, A., and Araniti, G. "D2D-based cooperative positioning paradigm for future wireless systems: A survey". IEEE Sensors Journal 22, no. 6 (2021): 5101-5112.
7. Chukhno, N., Trilles, S., Torres-Sospedra, J., Iera, A., and Araniti, G. "D2D-aided versus RIS-aided Cooperative Positioning: Theoretical Model for RSSI-based Ranging and Performance Comparison". IEEE Transactions on Vehicular Technology Journal, 2022, (submitted).
8. Chukhno, N., Chukhno, O., Pizzi, S., Molinaro, A., Iera, A., and Araniti, G. "Efficient Management of Multicast Traffic in Directional mmWave Networks". IEEE Transactions on Broadcasting 67, no. 3 (2021): 593-605.
9. Chukhno, O., Chukhno, N., Araniti, G., Campolo, C., Iera, A. and Molinaro, A., 2020. "Optimal placement of social digital twins in edge IoT networks". Sensors, 20(21), p.6181.
10. Chukhno, N., Chukhno, O., Araniti, G., Iera, A., Molinaro, A. and Pizzi, S. "Challenges and Performance Evaluation of Multicast Transmission in 60 GHz mmWave". In International Conference on Distributed Computer and Communication Networks (pp. 3-17). Springer, Cham, 2020, September.
11. Ometov, A., Chukhno, O., Chukhno, N., Nurmi, J. and Lohan, E.S. "When wearable technology meets computing in future networks: a road ahead". In Proceedings of the 18th ACM International Conference on Computing Frontiers (pp. 185-190), 2021, May.
12. Ometov, A., Shubina, V., Klus, L., Skibińska, J., Saafi, S., Pascacio, P., Flueratoru, L., Gaibor, D.Q., Chukhno, N., Chukhno, O. and Ali, A. "A survey on wearable technology: History, state-of-the-art and current challenges". Computer Networks, 193, p.108074, 2021.
13. Chukhno, N., Chukhno, O., Pizzi, S., Molinaro, A., Iera, A. and Araniti, G. "Unsupervised Learning for D2D-Assisted Multicast Scheduling in mmWave Networks". In 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB) (pp. 1-6). IEEE, 2021, August.
14. Chukhno, O., Chukhno, N., Galinina, O., Andreev, S., Gaidamaka, Y., Samouylov, K. and Araniti, G. "A Holistic Assessment of Directional Deafness in mmWave-based Distributed 3D Networks". IEEE Transactions on Wireless Communications, 21(9), pp.7491-7505, 2022.
15. Chukhno, O., Chukhno, N., Araniti, G., Campolo, C., Iera, A. and Molinaro, A. "Placement of Social Digital Twins at the Edge for Beyond 5G IoT Networks". IEEE Internet of Things Journal (Early Access), 2022.
16. Brancati, G., Chukhno, O., Chukhno, N., Araniti, G. "Reconfigurable Intelligent Surface Placement in 5G NR/6G: Optimization and Performance Analysis". In 2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). IEEE, 2022.
17. Chukhno, O., Chukhno, N., Pizzi, S., Molinaro, A., Iera, A., Araniti, G. "Modeling Reconfigurable Intelligent Surfaces-aided Directional Communications for Multicast Services".

In GLOBECOM 2022-2022 IEEE Global Communications Conference (pp. 5850-5855). IEEE (2022, December).

1.4 Thesis outline

The thesis is organized into 4 chapters, their content is briefly described below.

- **Chapter 1. Introduction** contains the motivation, structure, and contributions of this work.
- **Chapter 2. Delivering Multicast Traffic in mmWave Systems** presents methods for multicast traffic delivery in directional mmWave systems.
- **Chapter 3. 5G NR Sidelink Multi-Hop Transmission** is dedicated to the analysis of the potential benefits coming from D2D technology in the case of mobile multicast systems and public safety and factory automation applications.
- **Chapter 4. Cooperative Positioning** covers the state-of-the-art collaborative localization methods using D2D technology and integrating the transmission and localization components of wireless systems and introduces a theoretical model based on RSS ranging for D2D- and RIS-aided cooperative positioning.
- **Chapter 5. Conclusions** includes the summary of research outcomes and a discussion of future research avenues.

The final part of the document includes the bibliography.

2

Delivering Multicast Traffic in mmWave Systems

This chapter overviews multicasting in mmWave directional systems. We first provide an optimal multicast scheduling in 5G mmWave systems by implementing a globally optimal solution for a single- and multi-beam antenna operation. Then, we offer optimal multicasting in dual-mode mmWave sub-6 GHz hybrid deployments when both types of Radio Access Technology (RAT) can be utilized to serve multicast UE. Finally, we show how to utilize low-complexity heuristics and Machine Learning (ML) techniques to reduce the complexity of an optimal solution.

2.1 Optimal Multicasting in 5G/6G mmWave Systems

2.1.1 Motivation

The growth in demand for mobile multimedia services poses considerable challenges in providing reliable service quality, with the support of a large number of users competing for limited radio resources in cellular networks [32]. The NR technology is expected to be the primary enabler of the 5G cellular system's air interface. While the basic functionality of NR has already been specified in 3GPP Rel-15 [33] and Rel-16 [34], several advanced functionalities are still not defined. One of these critical functionalities is multicasting, which has been planned for 3GPP Rel-17 onwards [10, 35].

Multicasting is a prominent technique applied to improve bandwidth efficiency compared to unicast transmission [36]. In the multicast regime, BS can transmit the packet to many users simultaneously using the same band and Modulation and Coding Scheme (MCS). In the microwave spectrum with typically omnidirectional transmissions, multicast is a natural scheme to implement. However, in highly directional systems, i.e., mmWave band communications considered for NR, the use of extremely directional radiation patterns at the BS's antennas poses some challenges to the multicast operation design, which still remain unsolved or even unaddressed [12, 37].

In exchange for the promised extraordinary rates at the air interface, mmWave NR systems bring the following hurdles [38]. First of all, the use of highly directional antenna radiation patterns does not allow to serve simultaneously, in a single transmission, all the UE, which belong to the same multicast session and are located in very large regions [39]. Indeed, the Signal-to-Noise Ratio (SNR) decreases with larger beams. Secondly, NR is expected to work with considerably larger antenna arrays, hence increasing the design complexity with

respect to relatively simple microwave antenna configurations [40]. These issues are further exacerbated by the adverse properties of mmWave propagation, including severe free-space attenuation [41] and vulnerability to blockage [42]. Finally, the capability of modern antenna arrays to utilize multiple beams at the same time with potentially varying Half Power Beamwidth (HPBW) adds further degrees of freedom to multicast group formation and scheduling, significantly complicating their design. However, when multiple beams are available, the width of numerous beams to be swept simultaneously has to be properly selected, under the total transmission power constraint per antenna. This means that compared to single-beam systems, power has to be split among beams in a sophisticated manner.

The question of efficient multicasting in wireless systems has been addressed recently. Particularly, optimal solutions for single-beam antenna design have been proposed so far in [43]. Furthermore, there are a number of heuristic solutions for single-beam antennas [44, 45]. While several heuristics for multi-beam NR antenna designs have also been proposed [46, 47], no globally optimal solution is available. Without a globally optimal solution, it is impossible to fully benchmark existing solutions and develop enhancements.

This section fills the above-mentioned gaps by presenting a globally optimal solution for multi-beam antenna operation by explicitly considering mmWave specifics, including directional multi-beam antennas, signal propagation, and blockage. The optimization problem is first reduced to the special case of multi-period variable cost and size Bin Packing Problem (BPP) having well-known numerical solution algorithms, such that one may not place any constraints on the number of the beams and their HPBW. To account for multi-beam specifics, we select the optimization criterion to be the ratio of the amount of occupied resources to the overall resources in the system. We then proceed to formulate heuristic algorithms capable of approaching the globally optimal solution and extend the problem to multi-RAT case. We then use ML algorithms, including decision trees, random forests, and several types of neural networks for multicast grouping. Here, the exact solution, which is only feasible for a limited number of UEs in the multicast group, is utilized to obtain a training dataset for ML algorithms. The performance of the algorithms is finally compared based on the minimum amount of used resources as a metric of interest.

2.1.2 System Model

In this subsection, antenna, propagation, and blockage models. The main notations used throughout this chapter are gathered in Table 4.5.

Antenna Model

The main feature of antenna arrays in the context of multicasting is their directivity and gain in transmit/receive directions. The former parameter is often captured by utilizing HPBW, i.e., the angle, where the emitted power decreases by a factor of two. Detailed antenna radiation pattern model capturing not only the main lobe but side and back lobes is defined

in [48] as the superposition of individual elements. However, as the model is algorithmic in nature, it can only be utilized in simulation-based system-level performance evaluation.

A simplified model utilized for mmWave performance optimization and evaluation purposes is a cone-type model, see, e.g., [49], where the radiation pattern is represented as a conical zone with an angle of α coinciding with the HPBW of the antenna array. By following [50], the HPBW is proportional to the number of elements in the appropriate plane. Specifically,

$$\alpha = 2|\theta_m - \theta_{3db}|, \tag{2.1}$$

where θ_{3db} is the angle at which the value of the radiated power is 3 dB below the maximum, and θ_m is the location of the array maximum. The latter is given by $\theta_m = \arccos(-\beta/\pi)$, where β is the phase excitation difference affecting the physical orientation of the array. We assume $\theta_m = \pi/2$ for $\beta = 0$. The 3 dB point is provided by

$$\theta_{3db}^{\pm} = \arccos[-\beta \pm 2.782/(N\pi)], \tag{2.2}$$

and N is the number of antenna elements.

The gain over the HPBW can be found as [50]

$$G_A(\theta_{3db}^{\pm}) = \frac{1}{\theta_{3db}^{+} - \theta_{3db}^{-}} \int_{\theta_{3db}^{-}}^{\theta_{3db}^{+}} \frac{\sin(N\pi\cos(\theta)/2)}{\sin(\pi\cos(\theta)/2)} d\theta. \tag{2.3}$$

Note that a reliable approximation for HPBW of the main lobe can be obtained by utilizing $102^\circ/N$ [50]. Antenna gain over the main lobe in the appropriate plane can be approximated by the number of antenna elements, see Table 2.2 providing the comparison between gain calculated by utilizing (2.3) and approximation.

The described baseline model can be further extended or simplified according to the modeling needs.[1] Specifically, when there is no notable difference between BS and UE heights, a triangle model from [51] can be utilized. Alternatively, one may add a spherical component around the transmitter and appropriately divide the power between the main lobe and side and back lobes to represent parasite power. Throughout this chapter, we use the model as per Equation (2.3). Note that for multi-beam operation, the power budget has to be properly split among the beams.

Propagation Model

The SNR at the receiver located at the distance of y from the NR BS along the propagation path is

$$S(y) = \frac{P_A G_A G_U}{(N_0 W + M_I) L(y)}, \tag{2.4}$$

[1] The source code that generates planar antenna arrays, HPBWs, and gains is available at `https://github.com/NadezhdaChukhno/planar-antenna-array`

Table 2.1. Notation and parameters used in this chapter.

Fixed parameters with default values		
Parameter	**Definition**	**Value**
f_c	Carrier frequency, GHz	28 GHz
W	Available bandwidth, MHz	50 MHz
h_A	Height of NR BS, m	10 m
h_U	Height of UEs, m	1.5 m
h_B	Height of blockers, m	1.7 m
μ	5G NR numerology	3
M	Number of time slots in 1 ms subframe	8
L	Number of beams in the system	1,3,5
$P_{\max}$	Total available power, W	33 dBm
G_A, G_U	Antenna array gains at NR BS and UE ends, dBi	var/5.57 dBi
N_0	Power spectral density of noise, dBm/Hz	-174 dBm/Hz
N_A, N_U	Number of antenna elements at NR BS and UE	var/4 el
M_I	Interference margin	3 dB [52]
K	Number of multicast users	2-30
C	Bitrate of multicast session, Mbps	25 Mbps
w_{PRB}	Size of PRB, MHz	1.44 MHz
Δ	Subcarrier spacing, MHz	0.12 MHz
S_{th}	SINR threshold, dB	-9.47 dB
R_b	Number of available PRBs	32
R	Service (cell) area radius, m	250 m
Intermediate parameters		
$L(y)$	Path loss in linear scale	
$L_{dB}(y)$	Path loss in decibel scale	
X_A, Y_A	Coordinates of NR BS	
X_U, Y_U	Coordinates of UEs	
D	BSs intersite distance, m	
y	Three-dimensional distance between UE and NR BS, m	
y_{2D}	Two-dimensional distance between UE and NR BS, m	
$\theta^{\pm}_{3db}$	Upper and lower 3-dB points of antenna array, $^\circ$	
θ_m	Location of array maximum, $^\circ$	
β	Antenna array orientation, $^\circ$	
A_i, ζ	Propagation coefficients	
α	HPBW of a linear antenna array, rad	
$p_B(y)$	Distance-dependent blockage probability	
$S(y)$	Signal-to-interference-plus-noise ratio, SINR, dB	
P_A	Transmit power, W	
s_j	Spectral efficiency of the worst user in group G_j, bit/s/Hz	
Q	Number of carriers in a time slot	
c_j	Channel gain-to-noise ratio for beam j	
h_j	Channel gain for beam j	
σ_j	Standard deviation of the noise for beam j	

where P_A is the NR BS transmit power, G_A and G_U are the antenna array gains at the NR BS and the UE ends, respectively, N_0 is the power spectral density of noise, W is the operating bandwidth, $L(y)$ is the linear path loss. We capture the interference from the adjacent NR

Table 2.2. Parameters induced by 5G NR BS antenna arrays.

Array	HPBW,° [53]	Gain, dBi	Gain, linear
64x4	1.59	17.59	57.51
32x4	3.18	14.58	28.76
16x4	6.37	11.57	14.38
8x4	12.75	8.57	7.20
4x4	25.5	5.57	3.61
2x4	51	2.643	1.84
1x4	102	2.58	1

BSs via interference margin M_I in (2.4). For a given NR BS deployment density, one may estimate it by employing stochastic geometry-based models [49].

Following [54], the path loss measured in dB is

$$L_{dB}(y) = 32.4 + 21 \log_{10} y + 20 \log_{10} f_c, \tag{2.5}$$

where f_c is the carrier frequency in GHz and y is the three-dimensional (3D) distance between the NR BS and the UE. By concentrating on the averaged traffic load and channel conditions, we omit the consideration of small-scale fading. Nevertheless, the framework provided in what follows allows utilizing more complex models to capture propagation conditions. For example, the small-scale fading can be added to the model by assuming certain fading phenomena, such as Rayleigh [55], Rician [56], Nakagami-m [57], Noyt [58], or Weibull phenomena [59,60]. Those fading channels include multipath scattering effects, time dispersion, and Doppler shifts that arise from relative motion between the transmitter and receiver. Note that the introduction of an additional random variable to the considered propagation model, i.e., $P_R = FAy^{\gamma}$, where F follows the desired distribution, will affect the results quantitatively while preserving the same qualitative trend.

Blockage Model

We assume that blockers might temporarily block the LoS path between the UE and the NR BS. Depending on the current link state (LoS non-blocked or blocked) and the distance between the NR BS and the UE, the session employs an appropriate MCS to maintain reliable data transmission. The attenuation due to the human body blockage is assumed to be 15 dB [14].

The path loss in the form of (2.5) can be represented in the linear scale by utilizing the model in the form of Ay^{ζ}, where A and ζ are the propagation coefficients. Introducing the coefficients (A_1, ζ) and (A_2, ζ) that correspond to LoS non-blocked and blocked conditions, we have

$$A_1 = 10^{2 \log_{10} f_c + 3.24}, A_2 = 10^{2 \log_{10} f_c + 4.74}, \zeta = 2.1. \tag{2.6}$$

We note that the considered model can be extended to a model with building blockages and corresponding LoS/NLoS states. To this purpose, one may introduce the coefficients

(A_1, ζ_1), (A_1, ζ_2), (A_2, ζ_1), and (A_2, ζ_2) that correspond to LoS non-blocked, NLoS non-blocked, LoS blocked, and NLoS blocked conditions, respectively with $\zeta_1 = 2.1$, $\zeta_2 = 3.19$.

The value of SNR at the UE can then be written as

$$S(y) = \frac{P_A G_A G_U}{N_0 W + M_I} \left[\frac{y^{-\zeta}}{A_1} [1 - p_B(y)] + \frac{y^{-\zeta}}{A_2} p_B(y) \right], \tag{2.7}$$

where $p_B(y)$ is the blockage probability at the 3D distance y [42], which is calculated as

$$y = \sqrt{(X_A - X_U)^2 + (Y_A - Y_U)^2 + (h_A - h_U)^2}, \tag{2.8}$$

where (X_A, Y_A, h_A) and (X_U, Y_U, h_U) are the coordinates of the NR BS and the multicast user, respectively.

We note that blockage impairments in sub-6 GHz bands are in the range of $2-4$ dB and thus often neglected in performance models [61].

2.1.3 Analysis

Due to the highly directional nature of 5G mmWave systems, the usage of directional radiation patterns at the BS's antennas poses new challenges to the multicast operation design, which remain unsolved [12, 37]. For instance, in a single-beam system, all the UE devices, which belong to the same multicast session and are located in different cell regions, cannot be served within a single transmission. Differently, operation over larger beams limits the communications distances and also leads to inefficient use of radio resources due to lower MCSs. When multi-beam antennas are utilized, the total transmission power constraint per antenna introduces a similar effect and should be considered when selecting the width of beams to be swept simultaneously. We solve the optimization problem for a certain time interval, during which we assume that the radio channel conditions and the multicast group composition remain unchanged.

In this subsection, we describe the general framework for optimal multicast scheduling in 5G mmWave systems by implementing a globally optimal solution for single- and multi-beam antenna designs. This framework can further be extended to capture various operational specifics of 5G deployments. As an example, we demonstrate how to adapt it to the case of optimal multicasting in dual-mode mmWave sub-6 GHz hybrid deployments when both types of RATs can be utilized to serve multicast UEs. Finally, we show how to utilize low-complexity heuristics and ML techniques to reduce the complexity of an optimal solution.

Single-RAT Operation

The multicast multi-beam operation optimization problem can be formalized as a subclass of BPP [62], one of the most studied combinatorial problems. In BPP, a collection of items of various sizes must be either packed into a minimum number of identical bins, packed so that the items are evenly distributed, or filled in the most time-efficient manner. The variable-sized BPP [63] represents a new variant of BPP that attempts to reduce the cost of assigning

items to particular bins. The authors offer BPP settings, in which the item allocation cost to a bin is explicitly accounted for and may or may not be primarily determined by the item's volume. In our formulation, UEs represent items, whereas a beam is a bin for a subgroup of UEs. The goal is to minimize the cost in terms of the ratio of occupied resources to the total available resources for assigning multicast group's UEs to subgroups, with each subgroup served by a directional beam [1].

Consider the tri-sector cellular architecture, wherein each BS covers a 120° sector and operates with a directional antenna array that contains $L \geq 1$ beams. The set of K UEs that make up a multicast group is denoted as $\mathcal{K} = \{1, \ldots, K\}$. We assume an OFDMA-based system, where M designates the time horizon's length, i.e., the number of time slots in the time horizon (one subframe of 1 ms), with the index $t \in \mathcal{T}$, $\mathcal{T} = \{1, \ldots, M\}$, of each time slot. The number of time slots M depends on the chosen NR numerology μ. The maximum number of Primary Resource Blocks (PRBs) available in the system is restricted by MLR_b, where R_b is the available number of resource blocks in the system for a beam at time slot t for given numerology μ and operating frequency f_c. The potential maximum number of subgroups served within the time horizon is restricted by ML.

The number of possible subgroups when considering all combinations of K UEs of the multicast group scales as $2^K - 1$ [43]. Hence, we introduce $\mathcal{K}_j$, which denotes the set of UEs forming subgroup j, $j \in \mathcal{J}, \mathcal{J} = \{1, \ldots, 2^K - 1\}$, and $|\mathcal{K}_j|$ is the number of UEs in subgroup j. For instance, for $K = 3$ UEs, the number of subgroups' options is 7 and these feasible options are

$$\begin{aligned} &\mathcal{K}_1{=}\{1\},\ \mathcal{K}_2{=}\{2\},\ \mathcal{K}_3{=}\{3\},\ \mathcal{K}_4{=}\{1,2\}, \\ &\mathcal{K}_5{=}\{1,3\},\ \mathcal{K}_6{=}\{2,3\},\ \mathcal{K}_7{=}\{1,2,3\}. \end{aligned} \tag{2.9}$$

To proceed, we need to define a *suit* $\mathcal{G}_k$ as the collection of subgroup's indices, $\mathcal{G}_k \subset \mathcal{J}$, corresponding to the combination of subgroups $\mathcal{K}_j$, $j \in \mathcal{J}$, that covers all the UEs of the multicast group without their repetition, $k = 1, 2, ..., |\Omega|$, where Ω is the set of all such combinations.[2]

In the case of $K = 3$, the set Ω contains $\mathcal{G}_1 \sim \mathcal{K}_1 \cup \mathcal{K}_2 \cup \mathcal{K}_3$, $\mathcal{G}_2 \sim \mathcal{K}_3 \cup \mathcal{K}_4$, $\mathcal{G}_3 \sim \mathcal{K}_2 \cup \mathcal{K}_5$, $\mathcal{G}_4 \sim \mathcal{K}_1 \cup \mathcal{K}_6$, $\mathcal{G}_5 \sim \mathcal{K}_7$ with $|\Omega|$ being 5. Therefore, the following condition should be held:

$$\bigcup_{j \in \mathcal{G}_k} \mathcal{K}_j = \mathcal{K}, \qquad k = 1, 2, ..., |\Omega|, \tag{2.10}$$

as well as

$$\mathcal{K}_{j_1} \bigcap \mathcal{K}_{j_2} = \emptyset, \qquad j_1 \neq j_2,\ \forall j_1, j_2 \in \mathcal{G}_k, \tag{2.11}$$

meaning that each multicast UE has to be included in one subgroup only. Note that the set $\mathcal{K}_j$ of UEs forming subgroup j determines the directionality of the beam $\theta_{m,j}$, HPBW α_j required to cover all UEs in subgroup j and distance L_j from the BS to the farthest UE.

[2] Note that a suit is actually set of subsets. We utilize this term suit for clarity of further exposure.

For a single-beam system, $L = 1$, all the subgroups with indices included in suit $\mathcal{G}_k$ are served sequentially by one beam. For a multi-beam system, $L > 1$, we define the subset of subgroups' indices $\mathcal{G}_k^l \subseteq \mathcal{G}_k$, $l = 1, 2, ..., L$, which are scheduled for beam l. Hence, suits $\mathcal{G}_k^l$ should satisfy the following conditions:

$$\mathcal{G}_k = \bigcup_{l=1}^{L} \mathcal{G}_k^l,$$
$$\mathcal{G}_k^{l_1} \bigcap \mathcal{G}_k^{l_2} = \emptyset,\ l_1 \neq l_2,\ \forall l_1, l_2 \in \{1, 2, \ldots, L\}. \tag{2.12}$$

We introduce a binary indicator $g_j^t \in \{0, 1\}$ to designate the subgroup assignment decision variable at time slot t. More precisely, $g_j^t = 1$ if subgroup j is served at time slot t, and $g_j^t = 0$ otherwise. Then, a vector-indicator $\mathbf{g}^t = (g_1^t, \ldots, g_{|\mathcal{J}|}^t)$ represents subgroups that are served at time slot t.

We assume that at time slot t at most L beams can be simultaneously swept, or, equally, L subgroups can be served, that is,

$$\sum_{j \in \mathcal{G}_k} g_j^t \leq L,\ \forall t \in \mathcal{T}. \tag{2.13}$$

The model does not limit the scheduler's beam assignment; however, a suit service time should not exceed the scheduling time horizon that may depend on implementation, i.e.,

$$\sum_{j \in \mathcal{G}_k^l} \sum_{t \in \mathcal{T}} g_j^t \leq M,\ \forall l = 1, ..., L, \forall k = 1, ..., |\Omega|. \tag{2.14}$$

Furthermore, we have to take into account the total transmit power budget per antenna that serves subgroup j when dealing with a multi-beam system

$$\sum_{j \in \mathcal{G}_k} g_j^t P_j \leq P_{\max},\ \forall t \in \mathcal{T}, \tag{2.15}$$

where P_j can be calculated as

$$P_j = \frac{A_1 A_2 S_{th}(N_0 W + M_I)}{G_A G_U L_j^{\zeta_1} \left[A_2(1 - p_B(L_j)) + A_1 p_B(L_j)\right]}, \tag{2.16}$$

where $P_{\max}$ corresponds to the overall emitted antenna power to be split between beams, S_{th} is the SNR threshold corresponding to a chosen NR MCS [3], N_0 is the thermal noise at 1 Hz, W is the operational bandwidth, M_I is the interference margin, whereas the transmit antenna gain G_A depends on HPBW α_j, $\alpha_j = 102/N_j$ [53], G_A, G_U are in linear scale, L_j is the distance between the BS and the farthest UE in the subgroup j.

The SNR of subgroup j is defined according to (2.7) by substituting L_j for y. Consider that a multicast session requires a constant bit rate of C bps. Then, to calculate the number of resources required from BS to provide a multicast service with bit rate C, one needs to know the Channel Quality Indicator (CQI) and MCS values, and SNR to spectral efficiency mapping. In our analysis, we use MCS mappings from [64], but these parameters are typically vendor-specific.

The cost of the multicast service delivered to subgroup j is the function $a_j = f(P_j, N_j, C)$, where P_j is the transmit power of the corresponding beam, N_j is the number of antenna

elements used to form the radiation pattern of the beam, and C is the required session bit rate, i.e.,

$$a_j = C/s_j w_{\text{PRB}}, \tag{2.17}$$

where s_j is a spectral efficiency in bps/Hz of the farthest UE in subgroup j and w_{PRB} is a PRB size.

We emphasize that the scheduler's time slot assignment is written in vector $\mathbf{g}_j = (g_j^1, \ldots, g_j^M)$ with

$$\sum_{t \in \mathcal{T}} g_j^t = \left\lceil \frac{a_j}{R_b} \right\rceil, \; j \in \mathcal{J}. \tag{2.18}$$

The following condition on the number of resources allocated to subgroup j served by a beam should also be held

$$a_j \leq MR_b, \; j \in \mathcal{J}. \tag{2.19}$$

Finally, in (2.14) and (2.19), the constraint on the maximum number of available resources in the system should hold, i.e.,

$$\sum_{j \in \mathcal{G}_k} a_j \leq MLR_b, \; j \in \mathcal{J}, k = 1, ..., |\Omega|. \tag{2.20}$$

The proposed BPP formalism can be utilized to formulate single- and multi-beam multicast optimization problems. It also allows for extensions to the case of multiple RATs as discussed below.

Multi-Beam Antennas Optimization

In the case $L \geq 1$, the goal is to find an optimal grouping of multicast UEs that minimizes the total multicast service cost in terms of ρ - the ratio of occupied PRBs to the total available number of PRBs for the entire time horizon. Thus, the optimization problem can be written in the following form:

$$\min_{k \in 1, \ldots, |\Omega|} \sum_{j \in \mathcal{G}_k} \frac{a_j}{MLR_b}, \tag{2.21}$$

$$\text{s.t. } (2.10), (2.11), (2.12), (2.13), (2.14), (2.15), (2.19), (2.20),$$

with ρ as the objective function.

Single-Beam Antennas Optimization

In the case $L = 1$, the entire transmit power budget at BS is allocated to a single beam, $P_j = P_1 = P_A$. Hence, we can use either the optimization problem defined above in (2.21) or the conventional resource minimization task [45]. That is,

$$\min_{k \in 1, \ldots, |\Omega|} \sum_{j \in \mathcal{G}_k} a_j, \tag{2.22}$$

$$\text{s.t. } (2.10), (2.11), (2.12), (2.13), (2.14), (2.15), (2.19), (2.20).$$

Multi-RAT Operation

We now proceed with extending an optimal multicast scheduling formalism to multi-RAT 5G mmWave systems. The goal of the model remains the same. The scheduler aims to minimize

total service cost in terms of the ratio of occupied PRBs to the total number of available PRBs, ρ, for the entire time horizon, thereby finding the optimal grouping of multicast UEs while considering the possibility of transmission over two technologies, e.g., mmWave/μWave.

No Service Priorities

Considering the case of no external priorities, the problem formulation can be written in a similar way as described above for single-RAT with multi-beam antennas by introducing the variables with indices $_m$ and $_\mu$ for mmWave and μWave technologies, respectively,

$$\min_{k\in 1,\dots,|\Omega|} \sum_{j\in\mathcal{G}_k} \left[\frac{a_{j,m}}{M_m L_m R_{b,m}} + \frac{a_{j,\mu}}{M_\mu L_\mu R_{b,\mu}}\right], \tag{2.23}$$

$$\text{s.t. } (2.11), (2.12)$$

where

$$\mathcal{G}_k = \left(\bigcup_{l_m=1}^{L_m} \mathcal{G}_k^{l_m}\right) \bigcup \left(\bigcup_{l_\mu=1}^{L_\mu} \mathcal{G}_k^{l_\mu}\right), \tag{2.24}$$

$$\mathcal{G}_k^{l_{m_1}} \bigcap \mathcal{G}_k^{l_{m_2}} = \emptyset, l_{m_1} \neq l_{m_2},\ \forall l_{m_1}, l_{m_2} \in \{1, ..., L_m\},$$

$$\mathcal{G}_k^{l_{\mu_1}} \bigcap \mathcal{G}_k^{l_{\mu_2}} = \emptyset, l_{\mu_1} \neq l_{\mu_2},\ \forall l_{\mu_1}, l_{\mu_2} \in \{1, ..., L_\mu\},$$

$$\mathcal{G}_k^{l_m} \bigcap \mathcal{G}_k^{l_\mu} = \emptyset,\ \forall l_m \in \{1, ..., L_m\}, l_\mu \in \{1, ..., L_\mu\}, \tag{2.25}$$

$$\sum_{j\in\mathcal{G}_k^{l_m}} g_{j,m}^{t_m} \leq L_m, \forall t_m \in \mathcal{T}_m, \sum_{j\in\mathcal{G}_k^{l_\mu}} g_{j,\mu}^{t_\mu} \leq L_\mu, \forall t_\mu \in \mathcal{T}_\mu, \tag{2.26}$$

$$\sum_{j\in\mathcal{G}_k^{l_m}} \sum_{t_m\in\mathcal{T}_m} g_{j,m}^{t_m} \leq M_m, \sum_{j\in\mathcal{G}_k^{l_\mu}} \sum_{t_\mu\in\mathcal{T}_\mu} g_{j,\mu}^{t_\mu} \leq M_\mu, \tag{2.27}$$

$$\sum_{j\in\mathcal{G}_k^{l_m}} g_{j,m}^{t_m} P_{j,m} \leq P_{\max,m}, \forall t_m \in \mathcal{T}_m, \sum_{j\in\mathcal{G}_k^{l_\mu}} g_{j,\mu}^{t_\mu} P_{j,\mu} \leq P_{\max,\mu}, \forall t_\mu \in \mathcal{T}_\mu, \tag{2.28}$$

$$a_{j,m} \leq M_m R_{b,m}, a_{j,\mu} \leq M_\mu R_{b,\mu}, \tag{2.29}$$

$$\sum_{j\in\mathcal{G}_k^{l_m}} a_{j,m} \leq L_m M_m R_{b,m}, \sum_{j\in\mathcal{G}_k^{l_\mu}} a_{j,\mu} \leq L_\mu M_\mu R_{b,\mu}. \tag{2.30}$$

The constraints (2.11) and (2.12) are responsible for suits' formation. Constraints (2.24) and (2.25) ensure that the scheduling satisfies the constraints of serving multicast UEs by utilizing a single technology only. With constraint (2.26), we require that the system should comply with the constraint on the maximum number of subgroups to be served at each time slot t_m and t_μ. This implies that at a time slot at most L_m and L_μ, beams can be simultaneously swept. The constraint (2.27) guarantees that the suit service time does not exceed the subframe duration for all considered numerologies for any $l_m = 1, \dots, L_m$ and $l_\mu = 1, \dots, L_\mu$. The constraint (2.28) ensures that the transmit power budget per antenna that serves subgroup $j \in \mathcal{G}_k^{l_m}$, $j \in \mathcal{G}_k^{l_\mu}$, $k = 1, \dots, |\Omega|$, over mmWave/μWave bands is satisfied at any $t_m \in \mathcal{T}_m$ and $t_\mu \in \mathcal{T}_\mu$. Finally, the constraint (2.29) and (2.30) impose resource constraints assigning a beam to the subgroup for all the service time for any $j \in \mathcal{J}$.

We emphasize that (2.23) reflects the implicit mmWave or μWave priorities. In the first case, the system selects mmWave band to serve a set of UEs $\mathcal{K}_j$, $j \in \mathcal{J}$, if $P_{j,m} \leq P_{\max,m}$.

This means that the mmWave BS is utilized until it fails to perform successful data delivery, and μWave technology is only used when some multicast UEs reside outside of the coverage of mmWave BS. By analogy, μWave priory ensures that the set $\mathcal{K}_j$ is served by μWave BS, if $P_{j,\mu} \leq P_{\max,\mu}$.

Weighted Priority Service

An operator's technology selection may be influenced by the available spectrum, deployment area, traffic conditions, and other factors. We offer the following weighted optimization function to fulfill these specific requirements:

$$\min_{k\in 1,\ldots,|\Omega|} \sum_{j\in\mathcal{G}_k} \left[w\frac{a_{j,m}}{M_m L_m R_{b,m}} + (1-w)\frac{a_{j,\mu}}{M_\mu L_\mu R_{b,\mu}} \right], \tag{2.31}$$

$$\text{s.t. } (2.11), (2.24), (2.25), (2.26), (2.27), (2.28), (2.29), (2.30),$$

where w is the weight factor.

The weight parameter w in (2.31) can be introduced to provide weighted priority in technology selection. When considering the coexistence of unicast and multicast traffic, one may set $w = \min(1, R^2/R_m^2)$ with R and R_m being the service area and mmWave cell radii, respectively, making w proportional to the coverage distance. The motivation is that the objective function in (2.31) maximizes the resources available for a new session when the geometric locations of unicast sessions are uniformly distributed throughout the dual-mode BS coverage region. Alternatively, the weight w can be set proportionally to the operator's utility, depending on these factors.

More Than Two RATs

Different from the single-RAT networks, multicast user grouping that minimizes total service cost and mapping these subgroups onto multiple RATs for parallel transmission in the multi-RAT networks can be determined. In our framework, we consider minimization of the ratio of utilized to available resources, ρ, while satisfying the service requirements. Thus, similarly to the two-RAT scheme, the scheduler aims to minimize total delivery cost in terms of ρ during the entire time horizon considering the possibility of transmission over all available technologies. For more than two RATs, one may use the formulation described in subsection 2.1.3 by adding more components associated with all available technologies. Alternatively, the optimization criteria can be latency minimization, data rate maximization, etc. In general, for more than two RATs considered, the optimization function takes the following form

$$\min_{k\in 1,\ldots,|\Omega|} \sum_{j\in\mathcal{G}_k} \sum_{\eta\in H} w_\eta \frac{a_{j,\eta}}{M_\eta L_\eta R_{b,\eta}}, \tag{2.32}$$

where η represents the RAT index, $\eta \in H$, H is a set of RATs.

By combining multiple technologies, the effective service area of a multi-RAT solution will be extended to the coverage of all technologies onboard. Moreover, the reliability can be significantly improved compared to any single-RAT connectivity. Note that, in general, the choice of technology depends on the used application.

Algorithm 1: Single-RAT Optimal BPP formalism, $L \geq 1$

Input: $(X_U(i), Y_U(i), h_U), i \in \mathcal{K}$
Output: Optimally global solution for multicast grouping
Create $2^K - 1$ multicast subgroups of UEs
for *each subgroup* $\mathcal{K}_j$ **do**
    find the farthest UE i and the distance from BS to this UE: $y \leftarrow \max_{i \in \mathcal{K}_j} y_i$ as (2.8);
    find HPBW needed to cover subgroup $\mathcal{K}_j$
    $\alpha_j = \arccos\left(\frac{(X_U(i)X_U(i')+Y_U(i)Y_U(i')+h_U^2}{y(i)y(i')}\right)$ {α_j *is chosen according to the angle between two edge UEs* i *and* i'}
    find P_j from (2.16) using $L_j = y$;{$P_j = P_A$ *is fixed for* $L = 1$}
    find the cost a_j from (2.17);
end
Solve the problem by using (2.21) with exhaustive search for $L \geq 1$ or (2.22) for $L = 1$.

Solutions to Optimal Multicasting

General BPPs, wherein a given set of items of various sizes has to be packed into the fewest number of unit capacity bins, belong to the NP-complete decision problem [63]. In this subsection, we provide solutions to previously introduced multicast problems, including those based on exact branch-and-cut and branch-and-bound methods, various relaxation approaches, meta-heuristics, and ML methods.

Single-RAT Optimal Solution

The pseudo-code in Algorithm 8 describes the exact globally optimal solution for single-RAT multicasting formulated as (2.21) for $L \geq 1$ and (2.22) for $L = 1$. The algorithm employs our analytical framework described above to obtain optimal subgroup formation for UEs making a multicast group and corresponding resource allocation. Note that for the case $L = 1$, the power of a beam is $P_j = P_1 = P_A$. To reduce the complexity of the solution (which might be important for some applications working in real-time), we propose practical heuristic algorithms to solve the problem at hand.

Single-RAT Heuristic Solution

The proposed heuristic algorithm is suitable for the case $L \geq 1$ and consists of two stages: stage 1 - subgroups formation and stage 2 - beam assignment and power allocation. The second stage is also logically divided as follows: *(i)* selection of the subgroups, which have to be served at a time slot simultaneously, *(ii)* water-filling for detecting the maximum power allocation that can be assigned to all of the beams simultaneously, and *(iii)* the subsequent refinement of the allocations for selected beams. We underline that we consider multi-beam transmissions starting from the second stage, which implies that the power-budged constraints (2.15) per antenna have to be guaranteed. In other words, for the single-

beam antennas, $L = 1$, only the first stage is required (see Algorithm 2), whereas, in the case of $L \geq 1$, additional steps have to be performed (we refer to Algorithm 4).

Subgroups Formation. At this stage, we create subgroups to serve all UEs making a multicast group within a time horizon. This procedure can be completed in two ways, as detailed below.

Subgroup Formation Option 1.1. To implement beam assignment, we adapt the incremental multicast grouping algorithm for directional mmWave networks initially proposed in [45] to the multi-beam case, i.e., $L > 1$. It is worth mentioning that for $L = 1$, we utilize the approach described in [45] with the only modification of the objective function. Specifically, we determine the number and width of the beams required to optimize the multicast transmission performance by employing the resource utilization minimization criteria. The pseudo-code is shown in Algorithm 2, wherein the output of the algorithm contains the number of subgroups, n, required to serve set $\mathcal{K}$ of multicast UEs, $1 \leq n \leq |\mathcal{J}|$; the set of subgroups, $\mathcal{S}_1^M, \ldots, \mathcal{S}_n^M$, that covers all UEs $\mathcal{K}$ from a multicast group without their repetition; and required beam transmit power for each subgroup, $P_1^M, \ldots, P_n^M$.

The list of UEs to be allocated into subgroups is referred to as $\mathcal{A}$. Initially, we include all UEs of the multicast group, i.e., set $\mathcal{K}$, to list $\mathcal{A}$ (line 3). We also introduce the 3D distance-vector $\mathbf{y} = (y_1, y_2, ..., y_i, ..., y_K)$, each element of thereof represents the distance between the BS antenna and UE i as per (2.8), where i is the index of the UE. Vector $\mathbf{\Phi} = (\phi_1, ..., \phi_K)$ takes into account UEs' reference angles in the azimuth plane (lines 4-5). The amount of utilized resources is initially set to 0 in line 7. The algorithm iteratively partitions the UEs from list $\mathcal{A}$ into multiple subgroups, as seen in line 9. Particularly, the minimization function is set to infinity on line 10. Here, the minimization function reflects the occupied per UE resources for each multicast subgroup. The algorithm begins by selecting the furthest UE from list $\mathcal{A}$ with a distance y and its reference angle in the azimuth plane of ϕ_y (lines 12-13).

Then, adaptive beamforming is used based on the UE's location, wherein one beam pattern can be selected to transmit with a chosen MCS. Line 15 collects all UEs covered by a beam with width α directed toward the UE with reference angle ϕ_y and with distance y in the multicast subgroup $\mathcal{S}_\alpha$. We underline that the transmit power for each beam with width α is computed according to (2.16) substituting L_j with y for $L \geq 1$. The maximum available power P_A is used for the transmission when $L = 1$. Recall that for a single-beam operation, i.e., $L > 1$, unlike the approach described in [45], we consider the minimization of the ratio of occupied to available resources as the objective function (line 18). Here, s_α is a spectral efficiency for a beam with width α and corresponds to s_j in (2.17). As a result, the algorithm selects the best α for the chosen in line 12 UE and removes all the UEs served by the beam with width α UEs from the list $\mathcal{A}$ (line 29). Algorithm 2 comes to a stop either when all UEs have been serviced (i.e., list $\mathcal{A}$ is empty) or when there are no resources available in the system.

Subgroup Formation Option 1.2. Another approach for subgroup formation is presented in Algorithm 3. First, this algorithm chooses the farthest UE i from the BS and

Algorithm 2: Single-RAT Heuristic Stage 1 Option 1.1, $L \geq 1$

1 **Input:** $(X_U(i), Y_U(i), h_U), i \in \mathcal{K}$
2 **Output:** n; $\mathcal{S}_1^M, ..., \mathcal{S}_n^M$; $P_1^M, ..., P_n^M$;
3 $\mathcal{A} \leftarrow \mathcal{K}$, $\mathcal{K} = \{1, \ldots, K\}$;
4 $\mathbf{y} = (y_1, ..., y_K)$ as (2.8);
5 $\mathbf{\Phi} = (\phi_1, ..., \phi_K)$; {*reference angles*}
6 $n \leftarrow 0$; {*subgroups counter*}
7 $a_{\text{sum}} \leftarrow 0$;{*occupied resources collector*}
8 $\mathcal{S}_n^M \leftarrow \emptyset$;
9 **while** $\mathcal{A} \neq \emptyset$ *or* $a_{sum} < MLR_b$ *or* $n < ML$ **do**
10 $\quad \text{MIN}_Q \leftarrow \infty$;
11 $\quad n \leftarrow n + 1$;
12 $\quad y \leftarrow \max_{i \in \mathcal{A}} y_i$;
13 $\quad \phi_y \leftarrow \phi(y)$;
14 $\quad$ **for** $\alpha \in \Omega_\alpha = \{\alpha_{\min}, ..., \alpha_{\max}\}$ **do**
15 $\quad\quad \mathcal{S}_\alpha = \{i \in \mathcal{A} : \phi_y - \alpha/2 \leq \phi_i \leq \phi_y + \alpha/2\}$;
16 $\quad\quad$ calculate P_α from (2.16);
17 $\quad\quad$ **if** $P_\alpha \leq P_{\max}$ **then**
18 $\quad\quad\quad Q_\alpha = \frac{C}{s_\alpha w_{\text{PRB}} |\mathcal{S}_\alpha|}$;
19 $\quad\quad\quad$ **if** $MIN_Q > Q_\alpha$ **then**
20 $\quad\quad\quad\quad \text{MIN}_Q \leftarrow Q_\alpha$;
21 $\quad\quad\quad\quad \mathcal{S}_n^M \leftarrow \mathcal{S}_\alpha$;
22 $\quad\quad\quad\quad P_n^M \leftarrow P_\alpha$;
23 $\quad\quad\quad\quad a_n \leftarrow \frac{C}{s_\alpha w_{\text{PRB}}}$;
24 $\quad\quad\quad$ **end**
25 $\quad\quad$ **else**
26 $\quad\quad\quad$ go to line 29;
27 $\quad\quad$ **end**
28 $\quad$ **end**
29 $\quad \mathcal{A} \leftarrow \mathcal{A} \setminus \mathcal{S}_n^M$;
30 $\quad a_{\text{sum}} \leftarrow a_{\text{sum}} + a_n$;
31 **end**
32 **return** n, $\mathcal{S}_1^M, ..., \mathcal{S}_n^M$, $P_1^M, ..., P_n^M$.

identifies the subgroup $\mathcal{K}_j$, such that $i \in \mathcal{K}_j$, to serve at the smallest value $a_j/|K_j|$, $j \in \mathcal{J} = \{1, ..., 2^K - 1\}$ (lines 8-10). The motivation behind this approach is that the algorithm can cover more UEs when sweeping the beam by choosing the farthest UE from the multicast group. Further, to provide a less complex solution while preserving the intention to

Algorithm 3: Single-RAT Heuristic Stage 1 Option 1.2, $L \geq 1$

1 **Input:** $(X_U(i), Y_U(i), h_U), i \in \mathcal{K}$
2 **Output:** n, $\mathcal{S}_1^M, ..., \mathcal{S}_n^M$, $P_1^M, ..., P_n^M$;
3 Create $2^K - 1$ multicast subgroups of UEs, $\mathcal{J} = \{1, ..., 2^K - 1\}$;
4 $\mathcal{A} \leftarrow \mathcal{K}$, $\mathcal{K} = \{1, \ldots, K\}$;
5 $n \leftarrow 0$; {*subgroups counter*}
6 **while** $\mathcal{A} \neq \emptyset$ **do**
7 $n \leftarrow n + 1$;
8 find the farthest UE $i \in \mathcal{A}$ and the distance from BS to this UE: $y \leftarrow \max_{i \in \mathcal{A}} y_i$ as (2.8);
9 find all subgroups $\mathcal{K}_j$, $j \in \mathcal{J}$, such as $i \in \mathcal{K}_j$;
10 find subgroup $\mathcal{S}_n^M$ such as $i \in \mathcal{K}_j$, with the smallest utilized resources per UE: $\mathcal{S}_n^M \leftarrow \min_{j \in \mathcal{J}, i \in \mathcal{K}_j} a_j/|K_j|$;
11 $\mathcal{A} \leftarrow \mathcal{A} \setminus \mathcal{S}_n^M$;
12 remove from $\mathcal{J}$ all subgroups that contain UEs from $\mathcal{S}_n^M$;
13 **end**
14 **return:** n, $\mathcal{S}_1^M, ..., \mathcal{S}_n^M$, $P_1^M, ..., P_n^M$.

minimize the ratio ρ of occupied to available resources, the algorithm selects the beam with the smallest value of utilized resources per UE, $a_j/|K_j|$. Then, the algorithm erases selected UEs from the list $\mathcal{A}$ (line 11) and repeats the process for the remaining UEs (lines 6-12). We emphasize that all subgroups $\mathcal{K}_j$ from $\mathcal{J} = \{1, ..., 2^K - 1\}$ that contain the served UEs are also excluded (line 12). By doing this, Algorithm 3 significantly reduces the complexity while keeping comparable performance with the optimal solution obtained by Algorithm 8.

Beam Assignment and Power Allocation. The pseudo-code of stage 2, where beam assignment and power allocation is performed, is presented in Algorithm 4. Let $\mathcal{S}^M$ denote the set of subgroups being selected at stage 1 of the proposed heuristics. For the time horizon, the algorithm's objective is to find the subgroups that will be served simultaneously in each time slot and the transmit power for corresponding beams to minimize the total ratio ρ of occupied to available resources. Accordingly, the algorithm works until all subgroups are deleted from $\mathcal{S}^M$ (lines 5-22) and outputs the number of time slots m to serve all subgroups. We also introduce the $\mathcal{D}^{(m)}$ to denote a set of subgroups to be served at the current time slot m. The algorithm selects the worst in terms of the needed power as per (2.16) subgroup from $\mathcal{S}^M$ and adds this subgroup to the set $\mathcal{D}^{(m)}$ (lines 7-9). If the power budget constraint $P_{\max}$ allows adding more subgroups to the set $\mathcal{D}^{(m)}$, the algorithm selects the best subgroup in terms of the required transmission power and adds it to the set $\mathcal{D}^{(m)}$ (lines 10-19). The number of subgroups in $\mathcal{D}^{(m)}$ is restricted by L. When the set $\mathcal{D}^{(m)}$ is determined, the power

Algorithm 4: Single-RAT Heuristic Stage 2, $L > 1$

1 **Input:** $\mathcal{S}_1^M, ..., \mathcal{S}_n^M$; $P_1^M, ..., P_n^M$;
2 **Output:** m, $\mathcal{D}^{(k)}$, $P_j^{*(k)}$, $j = 1, ...n$, $k = 1, ...m$;
3 $\mathcal{S}^M \leftarrow \{\mathcal{S}_1^M, ..., \mathcal{S}_n^M\}$;
4 $m \leftarrow 0$; {*time slot counter*}
5 **while** $\mathcal{S}^M \neq \emptyset$ **do**
6 $m \leftarrow m + 1$;
7 $k_{\max} \leftarrow \arg\max_{j \in \mathcal{S}^M} P_j$;
8 $P_{\text{sum}} \leftarrow P_{k_{\max}}^M$;
9 $\mathcal{D}^{(m)} \leftarrow \mathcal{S}_{k_{\max}}^M$;
10 **if** $\mathcal{S}^M \setminus \mathcal{D}^{(m)} \neq 0$ **then**
11 **for** $j = 2 : L$ **do**
12 $k_{\min} \leftarrow \arg\min_{j \in \mathcal{S}^M \setminus \mathcal{D}^{(m)}} P_j$;
13 **if** $P_{sum} + P_{k_{\min}}^M \leq P_{\max}$ **then**
14 $\mathcal{D}^{(m)} \leftarrow \mathcal{D}^{(m)} \cup \mathcal{S}_{k_{\min}}^M$;
15 **else**
16 go to line 20;
17 **end**
18 **end**
19 **end**
20 Perform water-filling for $\mathcal{D}^{(m)}$ and obtain $P_j^{*(m)}$ from (2.33)-(2.35);
21 $\mathcal{S}^M \leftarrow \mathcal{S}^M \setminus \mathcal{D}^{(m)}$;
22 **end**
23 **return:** m, $\mathcal{D}^{(k)}$, $P_j^{*(k)}$, $j = 1, ...n$, $k = 1, ...m$.

water-filling algorithm (the two options described below) chooses the power such that the utilized resources are minimized (line 20).

Option 2.1. Traditional power water-filling. We now introduce $c_j = |h_j|/\sigma_j^2$ as a channel Gain-to-Noise Ratio (GNR), where h_j is a channel gain, and σ_j is a standard deviation of the noise for the beam corresponding to the subgroup j. In a traditional water-filling algorithm, the channel with high c_j receives more power, which leads to a higher system capacity. Note that GNR is related to the SNR as $S_j = P_j^M c_j$. Then at time slot k the SNR $S_j^{(k)} = \min_{i \in \mathcal{D}^{(k)}} S(y_i)$, $j \in \mathcal{D}^{(k)}$, $k = 1, ..., m$.

The beam's power allocation of the water-filling approach is the solution of the following optimization task for the optimal power $P_j^{*(k)}$ for subgroup j at time slot k:

$$\left(P_1^{*(k)},...,P_{|\mathcal{D}^{(k)}|}^{*(k)}\right) \leftarrow \max_{\left(P_1,...,P_{|\mathcal{D}^{(k)}|}\right)} \sum_{j=1}^{|\mathcal{D}^{(k)}|} \log(1+P_j c_j),$$

$$\text{s.t. } P_j \geq 0, \quad \forall j \in \mathcal{D}^{(k)}, \quad \sum_{j=1}^{|\mathcal{D}^{(k)}|} P_j = P_{\max}, \forall k = 1,...,m, \tag{2.33}$$

where $|\mathcal{D}^{(k)}|$ is the number of subgroups that have to be served simultaneously at time slot k, $|\mathcal{D}^{(k)}| \leq L$. Note that the first constraint implies that the power allocation is non-negative, while the second constraint limits the power budget of the system. The sought optimal transmit power $P_j^{*(k)}$ is

$$P_j^{*(k)} = (1/\xi^* - 1/c_j)^+, \tag{2.34}$$

where $1/\xi^*$ is the maximum power that can be allocated for each subgroup, $x^+ = \max(x, 0)$.

The problem in (2.33) is convex in nature. Since the maximization of concave function (2.33) is equivalent to the minimization of a convex function, we have

$$\xi^* \leftarrow \min_{\xi} \sum_{j=1}^{|\mathcal{D}^{(k)}|} \log\left(1+P_j^{*(k)}c_j\right) - \xi\left(\sum_{j=1}^{|\mathcal{D}^{(k)}|} P_j^{*(k)} - P_{\max}\right), \forall k = 1,...,m,$$

$$\text{s.t. (2.34)}. \tag{2.35}$$

Option 2.2. Resource-Based Power Water-Filling. Alternatively, one may employ resource information to execute water-filing. This option allocates additional power to those subgroups, resulting in the most significant reduction in the amount of utilized resources.

Machine Learning Solution Algorithms

Another approach to solving the multicast grouping problem is to utilize ML techniques. Below, we consider three classes of algorithms depending on their complexity, including: *(i)* regression models, *(ii)* decision trees and forests, and *(iii)* neural networks.

One may utilize the obtained data from the direct solution for the limited number of UEs in a multicast group to design an algorithm capable of solving the multicast grouping problem for more UEs in a multicast group. Thus, the suitable class of ML algorithms is supervised algorithms. In a supervised learning model, the algorithm learns on a labeled dataset (e.g., data from the direct solution) and provides the results that the algorithm can evaluate in terms of accuracy based on the training set. The algorithm can be implemented as an offline or online learning tool within the recently standardized ML framework for 5G systems, see [65–67].

The exploited ML algorithm has to be as simple as possible to run in real-time on the BS side when a new UE joins the multicast group or some UE leaves it. Therefore, the execution time and the training phase (preferably) have to be short. Since we aim at practical implementation, low-complexity ML tools receive priority in what follows. For this reason, one may include into consideration simple algorithms such as supervised classification and decision trees. We further consider random forests and neural networks to evaluate whether advanced techniques may provide more accurate results. Note that the decision tree is computationally faster than the random forest because of the ease of generating rules.

Several factors must be considered in a random forest classifier to interpret the patterns among the data points.

One may utilize two types of similarity metrics to evaluate the accuracy of ML algorithms. The first metric of interest, σ, is based on the exact matching of the number of subgroups and UEs assigned to these subgroups. More precisely, the following criterion can be utilized

$$\sigma = \frac{\text{number of correctly classified data}}{\text{number of test data}} \times 100\%. \tag{2.36}$$

Observe that if the match between ML and optimization results is perfect, the ratio of occupied to available resources, ρ, for both metrics coincide, i.e., $\rho_{opt} = \rho_{ML}$. However, due to the discrete nature of resource allocation and mapping between MCSs and spectral efficiency, the considered metrics ρ_{opt} and ρ_{ML} might be close even when the different number of subgroups and UEs assignment to these subgroups is observed. Since resource utilization is the main metric of interest, in addition to perfect matching between UEs assignments, one may also consider the second metric, i.e.

$$\gamma = \frac{\rho_{ML}}{\rho_{opt}} \times 100\%, \tag{2.37}$$

that measures the closeness of resource allocation produced by the considered ML algorithm and optimization framework.

Due to the availability of the training datasets, supervised learning algorithms, whose most common learning task is *classification*, can be utilized. For the problem at hand, one may consider the following supervised learning algorithms:

- *Decision Trees* are supervised algorithms used both for classification and regression. Their main advantage is the building of an interpretable model. Thus, they are also known as white-box algorithms [68].
- *Logistic Regression* is used for classification problems to assign observations to a discrete set of classes. This technique transforms the output by using the logistic sigmoid function to return a probability value class mapping.
- *Naive Bayes* is a simple but powerful classification algorithm, "probabilistic classifier", based on Bayes' theorem with the assumption of conditional independence among considered features of objects.
- *Support Vector Machine (SVM)* is an algorithm that can distinguish between two or more classes by defining a hyperplane that separates those classes. The support vectors are the closest points to the hyperplane. A change in the support vector results in a modification of the hyperplane [69]. SVM can be used for solving classification and regression problems.
- *K-Nearest Neighbors (kNN)* is a famous algorithm used both for solving classification and regression problems. The output of the algorithm is obtained by comparing the input with known data [69].
- *Neural Network (NN)* classifiers are used for multiclass classifications. These models typically outpace other algorithms in prediction accuracy. The flexibility of NN models increases with the number and size of connected layers.

Table 2.3. Interpretability characteristics of main classifier types.

Algorithm (Classifier)	**Interpretability**
Decision Trees	Easy
Ensemble Classifiers	Hard
Logistic Regression	Easy
Naive Bayes Classifiers	Easy
Support Vector Machine	Easy for linear kernel, hard for others
Nearest Neighbor Classifiers	Hard
Neural Network Classifiers	Hard

Table 2.3 provides the characteristics of considered algorithms in terms of interpretability, which allows for a better understanding of obtained solutions. This feature makes it easy to avoid solution mistakes or errors and compensate for them. With higher interpretability, one can understand how ML model makes its decision.

All the abovementioned algorithms can solve the required classification problem, namely, assigning an observation (UE) to one of the classes (multicast subgroups). In general, the following model's features can be considered as parameters that form the dataset for the training of supervised algorithms: *(i)* UE's coordinates X_U, Y_U, *(ii)* number of UEs K, *(iii)* service area radius R, *(iv)* bandwidth W, *(v)* session data rate C, and *(vi)* number of subgroups (clusters). Model's features form predictor's set $\mathcal{P}$. We choose these parameters since they all may affect the results of the classification. Later, we explore which of those parameters have a higher importance level. The algorithms learn from the training dataset of size H_1 (provided by the optimization presented in subection 2.1.3) by predicting the data and adjusting it for the correct answer of multicast subgroups formation 5G NR systems.

To prepare the data for training ML algorithms, one may first use the exact solution to obtain optimal multicast grouping and beam selection and then train the ML models using these data. To evaluate the accuracy of models, the following metrics can be utilized: *(i)* σ based on the correct classification of UEs and *(ii)* γ based on the resulting resource utilization matching. Finally, since the optimal solution is feasible only for a limited number of UEs, one needs to test the extrapolation capabilities of ML models by training them on the reduced number of UEs and then assessing the accuracy for a larger number of UEs.

Multi-RAT Optimal Solutions

Similarly to a single-RAT optimal solution, the formalized optimization problem for the case of multi-RAT operation can be classified as a special class of BPP, where items of various sizes are packed into a finite number of bins, each of a fixed capacity with the goal of minimizing the cost of assigning the items to the particular bins. The pseudo-code in Algorithm 5 defines the globally optimal solution according to (2.23), i,e., suit $\mathcal{G}_k^*$ of subgroup's indices in the form of (2.24). Algorithm 5 belongs to the class of NP-hard problems with exponential complexity.

For a limited number of UEs in the coverage area of BS, the direct solution of the problems in (2.23) and (2.31) can be adopted by using, e.g., branch-and-cut or branch-and-

Algorithm 5: Multi-RAT Optimal BPP formalism, $L \geq 1$

Input: $(X_U(i), Y_U(i), h_U), i \in \mathcal{K}$
Output: Optimal solution $\mathcal{G}_k^*$ for multicast grouping in form of (2.24)
Create $2^K - 1$ multicast subgroups of UEs
for *each subgroup* $\mathcal{K}_j$ **do**
    find the farthest UE i and the distance from BS to this UE: $y \leftarrow \max_{i \in \mathcal{K}_j} y_i$;
    find HPBW needed to cover the subgroup $\mathcal{K}_j$
    $\alpha_j = \arccos\left(\frac{(X_U(i)X_U(i')+Y_U(i)Y_U(i')+h_U^2}{y(i)y(i')}\right)$; {$\alpha_j$ *as the angle between two edge UEs* i *and* i'}
    calculate P_j as (2.16);{$P_j = P_{\max}$ *is fixed for* $L = 1$}
    find the cost $a_{j,m}, a_{j,\mu}$ from (2.17);
end
Solve by exerting (2.23) with the exhaustive search.

bound techniques. Some of these solutions allow controlling heuristic behavior focusing on the solution's integrity rather than its optimality. In general, the following solutions have been shown to significantly improve the heuristic behavior of Mixed-Integer Programming (MIP):

- **Metaheuristics.** These are general frameworks to build heuristics, often using combinatorial formulations. Metaheuristic rules and principles can be used to create heuristics for resolving mathematical programming problems. For example, **Local Branching (LB)** is based on the idea of altering neighborhoods throughout the search to obtain the best feasible solution [70]. LB is a technique created based on the exact method. The difference is that the LB is limited in time to solve a problem. If this period elapses without the optimal solution being determined, LB stops and returns the best-known solution.
- **Neighbourhood Search Methods.** Relaxation-Induced Neighborhood Search Heuristic (RINS) is a heuristic that explores the neighborhood of a valid solution to discover an improved one [71]. Continuous relaxation of the MIP model is used to build a promising neighborhood, which is formulated as another MIP (known as the sub-MIP). Limiting the number of nodes in the search tree truncates the sub-MIP optimization.
- **Randomization Methods.** Since the formulated problem is NP-hard, one may adopt a heuristic *simulated annealing* solution, which is a stochastic global search optimization algorithm and is known to be efficient for BPPs [72]. We specify the initialization and implementation parts as well as parameterize the technique below.

Multi-RAT Heuristic – Simulated Annealing

Simulated annealing is based on the principle of randomizing the local search strategy and accepting changes that, with some probability, make the result better. The approach imitates the annealing of metals in thermodynamics, which involves exposing the metal to a very high temperature and then allowing it to cool down slowly to create the required shape

with a defect-free structure. As a result, using an appropriate temperature cooling schedule is a critical idea in simulated annealing. Several variations of the simulated annealing method differ in the distribution and temperature reduction rule, resulting in specific disadvantages and benefits such as speed, the assurance of reaching the global minimum, and execution complexity.

A critical part of the simulated annealing algorithm is the temperature control rule. Each control rule reduces the temperature at a different rate, and each method is better at optimizing a certain type of model. The main types of temperature control rules are as follows:

- Linear rule: $T = T - \omega_c$;
- Geometric rule: $T = T\omega_c$ (frequently used);
- Slow-reduction rule: $T = T/(1 + \omega_c T)$, ω_c is a constant;
- Fast annealing: $T = T/k$ [73];
- Very fast annealing: $T = T\exp(-c_i k^{1/D}), i = 1, ..., D$, where D is dimension of the search space (number of variables in the cost function), i denotes i-th variable of the cost function and, as understood from this equation, various annealing processes can be considered for different variables, c_i is a constant that can have different values depending on the problem [73];
- Boltzmann annealing: $T = T/\ln k$ [74].

To solve the multi-RAT multicast problem, one may utilize the *standard simulated annealing* methodology [75], presented in Algorithm 6 to obtain heuristic solution denoted as $\tilde{\mathcal{G}}_k$. First, we define problem-specific choices, including the form of the objective function $c(S)$ and the strategy for obtaining solution S. Theoretically, the initial solution does not affect the final result. However, several experiments have shown that using a good heuristic to obtain an initial solution occasionally results in a faster convergence to the optimal solution [76, 77]. To achieve the global minimum, the number of steps, $MaxIt$, in the inner loop of Algorithm 6 must be larger than the number of points in the solution space, i.e., $MaxIt > |Q|$, leading to the futility of the approach [75].

The following heuristic described in detail in [12] is used to acquire a good initial solution and, hence, minimize the number of steps in the simulated annealing approach. First, the farthest UE from the BS is chosen. Then, by altering the set of predefined beamwidths, one that demands the lowest number of utilized resources per UE is selected to serve the corresponding UEs. Note that all UEs covered by the beam are included in the corresponding subgroup. All the selected UEs are then removed from the set of all multicast group's UEs, and the algorithm again selects the farthest UE from the remaining set of UEs. The algorithm operates until there are no UEs left.

We now describe the general logic that handles the operation of the Algorithm 6 itself. Here, the initial temperature is set in the temperature parameter, while the temperature reduction is a cooling function ω_c, $0 < \omega_c < 1$. At each iteration k, the temperature is cooled

Algorithm 6: Multi-RAT Simulated Annealing

1 **Input:** $(X_U(i), Y_U(i), h_U), i \in \mathcal{K}$
2 **Output:** Heuristic solution $\tilde{\mathcal{G}}_k$ for multicast grouping in form of (2.24)
3 Generate a feasible initial solution S;
4 Setup initial temperature $T > 0$;
5 Setup the cooling rate ω_c;
6 **while** $T = 1$ **do**
7 $k \leftarrow 0$;{ *number of iterations*}
8 **while** $k < MaxIt$ **do**
9 Select a neighbor S' of S;
10 $\Delta c = c(S') - c(S)$;
11 **if** $\Delta c \leq 0$ **then**
12 $S \leftarrow S'$;
13 **else**
14 $S \leftarrow S'$ if $random(0,1) < \exp(\frac{-\Delta c}{T})$;
15 **end**
16 $k \leftarrow k + 1$;
17 **end**
18 $T = T\omega_c$;
19 **end**

down by ω_c. We denote the number of neighbors to visit at each iteration as $MaxIt$. A stopping criterion can be either the condition $T = 1$ or the lack of significant improvement in two consecutive executions of the objective function of the outer loop. Also, to stop the algorithm, one may utilize the criterion of reaching a solution that does not exceed a predefined cost. We use condition $T = 1$ to stop the algorithm. The objective function $c(S)$ represents the ratio of occupied to available resources, ρ, required by solution S, where S is a set that includes all the UEs once.

After defining the initial solution S and setting up the general execution parameters, such as initial temperature and cooling rate, the algorithm performs the outer "while" loop with fixed temperature (lines 4-17 of Algorithm 6). In the inner "while" loop, which executes $MaxIt$ times, the algorithm selects a random neighbor S' and performs the Metropolis test (see below) to accept the move from S to S' or not (lines 6-15). In the algorithm, the procedure of the random neighbor selection is as follows: *(i)* randomly generate set S' such that it covers all the UEs, *(ii)* calculate the required transmit power for S' based on the most robust SNR, *(iii)* perform water-filling for those subgroups that can be served simultaneously in a slot considering the power budget per antenna, $P_{\max}$, and *(iv)* compute $c(S') = \rho$. Note that if the cost fiction $\Delta c = c(S') - c(S)$ is non-positive, the move is always accepted. Otherwise, the change of solution is accepted with probability $P = e^{-\Delta c/T}$. Once

$MaxIt$ steps are completed, the temperature decreases (line 18), and the inner loop starts again. The algorithm works until the stop criterion is met.

For fixed T, the acceptance probability P is an exponentially decreasing function of Δc. Hence, as Δc increases, the acceptance probability quickly becomes very small. The Metropolis test [78] allows for leaving the local minimum encountered while wandering around the solution space within the inner loop. After performing $MaxIt$ steps, the temperature declines according to the temperature reduction law, and the inner loop starts again. For fixed Δc, the acceptance probability decreases with T, so the accepted moves are rarer in the consecutive execution of the inner loop.

Algorithm 6 is relatively simple to implement, but its efficient implementation requires tinkering with parameters and figuring out ways to reduce the run-time associated with computing the solution for values in the search space. The initial temperature typically is a large number. Then the inner while-end loop is executed $MaxtIt$ times, which is another parameter of the algorithm. As simulated annealing is a heuristic solution, in the next subsection, we explore the optimality and complexity of the simulated annealing algorithm when the number of neighbors to be explored, $MaxIt$, is 15, and the initial temperature is $T = 10$.

2.1.4 Performance Assessment

In this subsection, we provide the performance comparison of discussed in this chapter optimal algorithms, heuristics, and statistical methods. To this end, by analogy with the previous subsection, we start with single-RAT solutions. We then proceed with a dual connectivity case.

Single-RAT

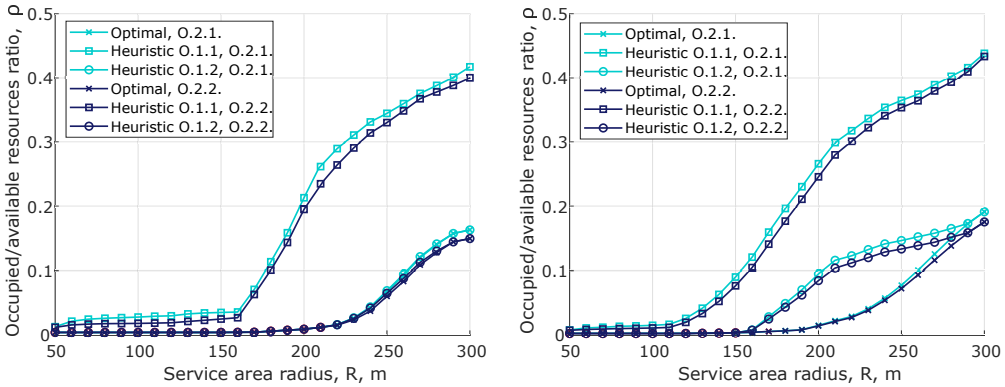
Execution Times

We start with the comparison of the computation complexity of the algorithms, as shown in Table 2.4. First, one can see that in the category of single-RAT multicast solutions, the heuristic solution offers a low-complexity scheduling scheme but at the expense of optimally (described below). Note that the run time of the heuristic solution for a single-RAT depends on the UE locations and the resulting configuration (see single-RAT heuristic stage 1 option 1.1). Moreover, Table 2.4 illustrates that the optimal solution is impractical when the number of UEs is high (more than 12 in single-RAT and more than 10 in dual-RAT).

Regarding the multi-RAT solutions, one may deduce that involving more than two technologies will result in an even higher execution time. One may use relaxation techniques (LB, RINS) to reduce the complexity of the optimal solution. However, the improvement is not significant. Instead, simulated annealing noticeably reduces the computational time. We emphasize that simulated annealing with the initial configuration obtained through heuristic (SA-H) and random initial configuration (SA) offer the same complexity since the heuristic solution is extremely low-complex.

Table 2.4. Algorithms' execution time.

Time/K	2	5	7	10	12	15	17	20	22	25	27	30
Single RAT (in minutes)												
Optimal	0.008	0.01	0.06	10.03	54.35	60 (limited)	-	-	-	-	-	-
Heuristic*	0.0021	0.0046	0.005	0.0043	0.0073	0.008	0.008	0.0085	0.0096	0.012	0.0115	0.017
*Note that the run time of the heuristic for single RAT depends on the UEs locations and the resulted configuration. In general, the complexity increases with the number of UEs.												
Dual Connectivity (in minutes)												
Optimal	0.15	0.89	14.37	29.50	60 (limited)	-	-	-	-	-	-	-
LB	0.13	0.88	14.2	28.70	60 (limited)	-	-	-	-	-	-	-
RINS	0.13	0.88	14.25	29.20	60 (limited)	-	-	-	-	-	-	-
SA-H	1	2.29	3.12	11.01	13.19	17.49	21.51	25.58	29.65	33.70	37.75	41.79
SA	1	2.29	3.12	11.01	13.19	17.49	21.51	25.58	29.65	33.70	37.75	41.79
Machine Learning, $R = 250\ m$, $H_2 = 5000$ (in seconds)												
Log. Regression	2.223	1.863	2.307	2.21	1.93	1.877	2.532	2.109	2.115	1.959	4.344	1.87
Kernel Naive Bayes	14.82	14.033	13.649	14.31	13.94	13.779	16.136	13.917	14.057	13.88	16.646	14.14
Random Forest	2.888	2.555	2.602	2.56	2.54	2.463	2.481	2.475	2.494	2.515	3.258	2.97
Narrow NN	0.184	0.137	0.1253	0.11	0.12	0.134	0.126	0.148	0.163	0.156	0.142	0.15
Weighted KNN	0.758	0.649	0.646	0.37	0.38	0.734	0.691	0.624	0.684	0.652	0.836	0.75
Cubic SVM	8.291	4.02	3.6771	5.6	7.36	5.093	7.097	5.547	10.934	10.698	4.956	8.31
Fine Tree	0.373	0.342	0.361	0.49	0.39	0.335	0.335	0.355	0.374	0.351	0.393	0.44
Coarse Tree	0.194	0.128	0.1297	0.15	0.19	0.1297	0.1345	0.129	0.133	0.133	0.131	0.16



(a) $L = 3$. (b) $L = 5$.

Fig. 2.1. Ratio of occupied to available resources as function of cell radius, $K = 10$, $C = 25$ Mbps, $W = 50$ MHz [1].

Finally, we comment on the complexity of ML solutions. Here, we note that the run time depends on the size of the testing dataset, H_2, and is not affected by the number of UEs. This proves that ML algorithms represent a good tool to work with a high number of UEs in the case of optimal multicast grouping.

Solutions' Performance and Water-Filling Comparison

To provide a comparison of the solutions designed for single-RAT systems, we start with Fig. 3.7 presenting the ratio of occupied to available resources, ρ, for the maximum number of beams $L = 3$ and $L = 5$ as a function of the cell area radius R. From these illustrations, we observe that the curves for $L = 3$, see Fig. 2.1(a), grow much slower with the increase in the cell radius than for $L = 5$, see Fig. 2.1(b). It is important to highlight that at smaller values of BS coverage radius R (e.g., approximately $50 - 100$ m), heuristic (O.1.2) and optimal solutions combine all UEs of the multicast group into a single subgroup. This explains the fact that the curves for $L = 5$ first show better performance and then demonstrate higher ρ values for all schemes. We also note that the reason behind the gap between the optimal solution and (O.1.2) for (O.2.1) and (O.2.2) heuristic options for $L = 5$ lies in the selected number of beams per time slot. More precisely, at R of approximately $150-250$ m the optimal solution utilizes one beam and several time slots, whereas heuristic solutions serve UEs with more than one beam within one time slot. Hence, we may deduce that at large distances, such as $150 - 240$ m, it is crucial to utilize one beam at a time to minimize ρ. Note that all the considered strategies utilize unicast mode to serve multicast UEs, i.e., use individual beam for each multicast UE, starting from around $R = 250$ m.

Analyzing the presented data further, one may also observe no significant difference between the types of power water-filling schemes, i.e., options (O.2.1) and (O.2.2), with the latter slightly outperforming the former. This modest superiority is intuitive and stems from the fact that water-filling (O.2.2) is based on the resource information feature. Similarly to $L = 1$, the heuristic option with exhaustive search (O.1.2) provides the best approximation of the optimal solution. However, as the maximum number of beams, L, increases, even this approximation starts to deviate from the optimal solution.

Optimal Number of Beams

The abovementioned conclusions on the utilized number of beams are further complemented by Fig. 2.2, which demonstrates the optimal number of beams, L_{opt}, as a function of the cell area radius. One may observe that the optimal solution selects only one beam per time slot until R reaches 230 m and 250 m for $L = 5$ and $L = 3$ beams. Further, as one may learn from the curves for $L = 3$, the optimal solution chooses one beam and several time slots when R is in the range of $240 - 250$ m, whereas the proposed heuristics (O.1.2) and (O.1.1) sweep two and three beams per time slot, respectively. Analyzing both Fig. 3.7 and Fig. 2.2, we can conclude that for the practical ranges of cell size and the considered number of UEs, the optimal solution always utilizes no more than $2 - 3$ beams.

Optimal Multicast Subgroup Size

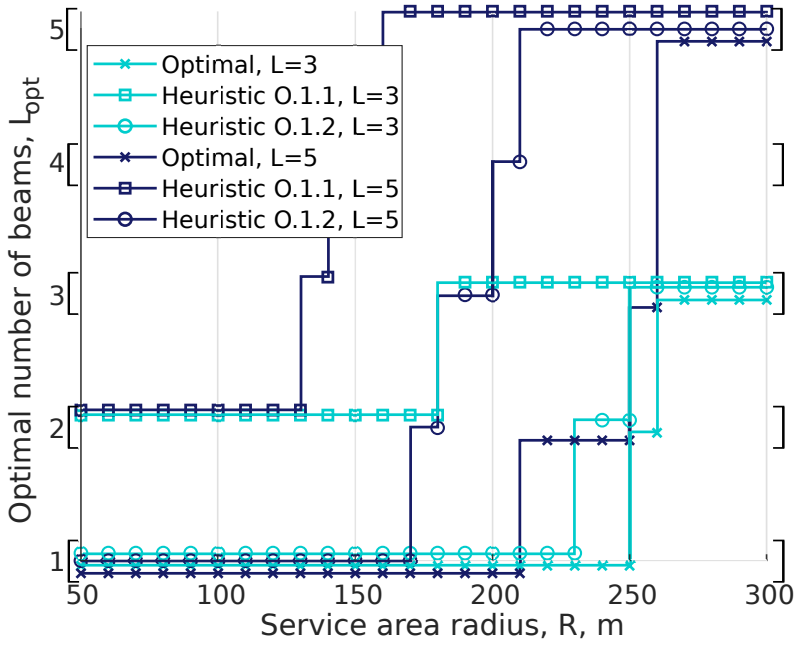


Fig. 2.2. Optimal number of beams in multi-beam system as function of cell radius, $K = 10$ UEs, $C = 25$ Mbps, $W = 50$ MHz [1].

We proceed with Fig. 2.3, which displays the average number of UEs served by a beam per time slot. The rationale for considering this metric is to assess the number of transmissions exploited to serve multiple UEs for various radii. The presented results confirm the statement derived from Fig. 3.7 that starting from 250 m, almost all the schemes use the unicast mode for $L = 5$ beams. Hence, Fig. 2.3 provides an insight into the efficiency of the multicast transmissions in mmWave networks. More precisely, it reflects situations where the system utilizes a lower resource ratio than that required by the unicast service, where UEs are serviced by individual beams (one UE per beam). One may observe that the system with $L = 3$ beams works better in terms of serving more UEs within a beam, which can be explained by the fact that, in general, the increase in the number of beams leads to a decrease in the number of UEs per beam. For small cell radii, a single beam (all UEs in one subgroup) is almost always utilized, while unicast service is only feasible for higher ones.

Extrapolation and Machine Learning

We now proceed with the assessment of the effect of the number of UEs, K, shown in Fig. 2.4 for cell radius of $R = 250$ m, requested rate of $C = 25$ Mbps, and three bandwidths, $W = 50$ MHz, $W = 100$ MHz, and $W = 200$ MHz. Note that for values of K higher than 20, we utilize quadratic extrapolation to construct the curves for the optimal solution. Analyzing the presented data, one may notice that the increase in the number of UEs leads to a rise in the ρ ratio for all the considered solutions. Indeed, higher values of K theoretically lead to either a higher number of subgroups or a higher number of UEs in a subgroup (which may negatively affect the multicast subgroup channel condition), thus increasing the ratio ρ of

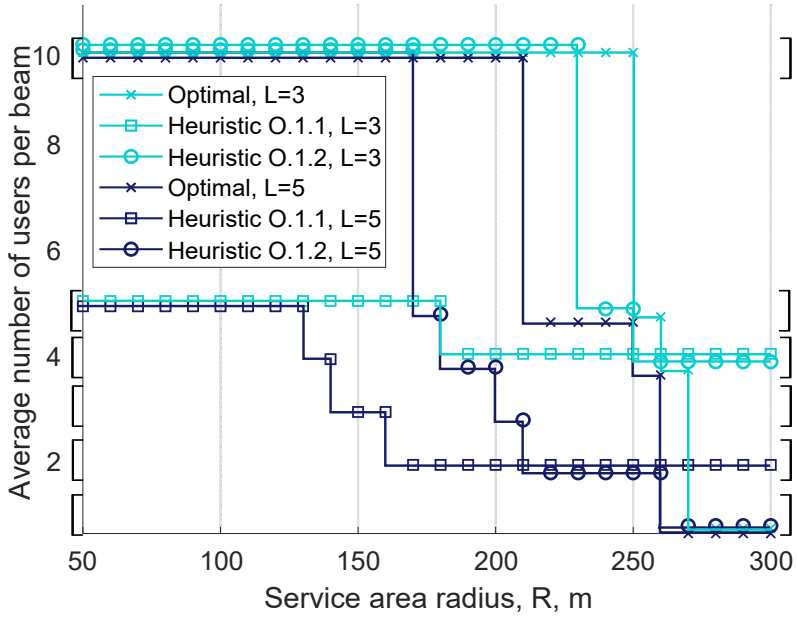


Fig. 2.3. Average number of UEs per beam as function of cell radius, $K = 10$ UEs, $C = 25$ Mbps, $W = 50$ MHz [1].

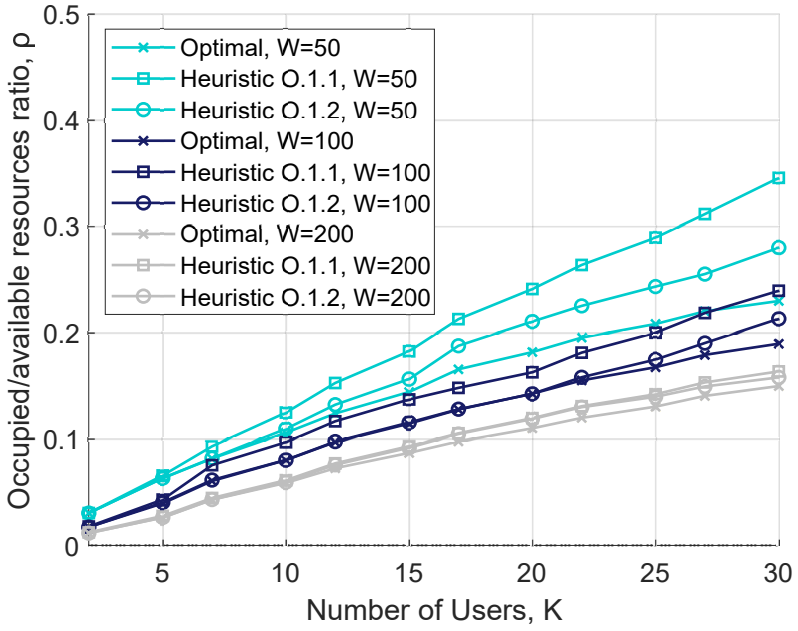


Fig. 2.4. Ratio of occupied to available resources as function of number of UEs, $R = 250$ m, $C = 25$ Mbps, $W = 50, 100, 200$ MHz, $L = 1$ beam [1].

utilized to available resources. Further, with the increase in K, the gap between the optimal and heuristic (O.1.2) solutions becomes larger. The impact of the increase in the available bandwidth W is also evident from Fig. 2.4. One may learn that for larger bandwidth of $W = 200$ MHz, the gap between the optimal and heuristic solutions is lower compared to $W = 100$ MHz and $W = 50$ MHz. The rationale is that the data transmission is much

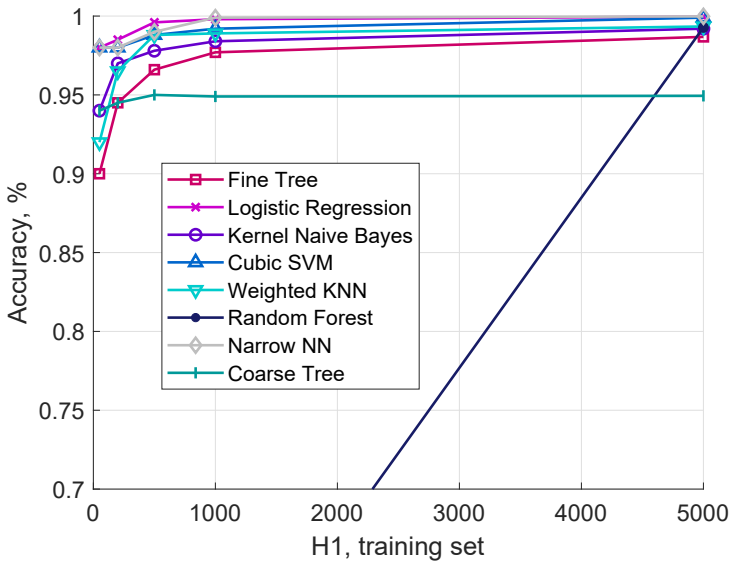


Fig. 2.5. Subgroup assignment accuracy, σ, for $H_2 = 5000$, $R = 250\,\text{m}$, $K = 10$ [2].

faster with a larger bandwidth. This, in turn, leads to a lower ratio of occupied to available resources. Therefore, this is an inherently quantitative effect as the difference between smaller values of ρ for a larger bandwidth is lower compared to the difference for a smaller bandwidth.

Regarding the ML algorithms implementation, the accuracy of UEs allocations to subgroups and resource matching is shown in Fig. 2.5. Here, we see that UE allocation to subgroups accuracy, σ, increases with the size of the training data set H_1 as expected. However, starting from approximately $H_1 = 1000$, the accuracy basically plateaus and does not increase any further. At the same time, note that perfect resource matching with optimal solution approach is observed for this considered distance even for very small values of H_1. Furthermore, the accuracy of all the considered algorithms (except for Random Forest) remains virtually unchanged when increasing the training sample size from $H_1 = 1000$ to higher values. This allows considering the latter as the lower bound on the training set size in practical implementations.

We now proceed with analyzing the extrapolation capabilities of the ML algorithms. To this aim, we train these algorithms by utilizing the training sample of length $H_1 = 1000$ for 10 UEs and then applying the trained algorithms to the system with 13 UEs. The accuracy metrics are calculated for the system with 13 UEs solved by applying the optimal solution. Fig. 2.6 shows the accuracy of the multicast subgroups formation for $H_1 = H_2 = 5000$ and $K = 13$ UEs. As one may observe, the match is perfect up until approximately $R = 250$ m and then drops abruptly for $R = 275$ m and beyond. The rationale is that the considered metric accounts for specific UEs classified into subgroups. Up until $R = 275\,\text{m}$, only one subgroup is utilized, explaining the perfect match between solutions. We also note that for the cell radius higher than $R = 300$ m, UEs are served individually using unicast transmissions.

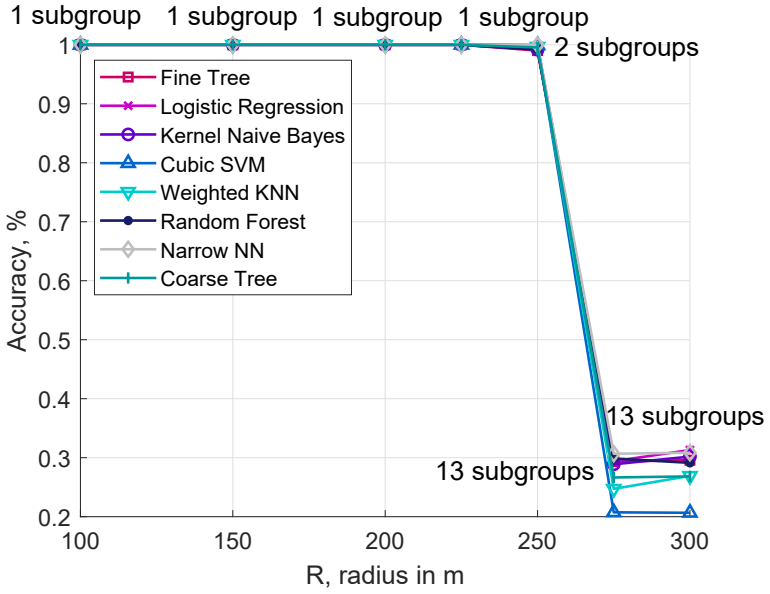


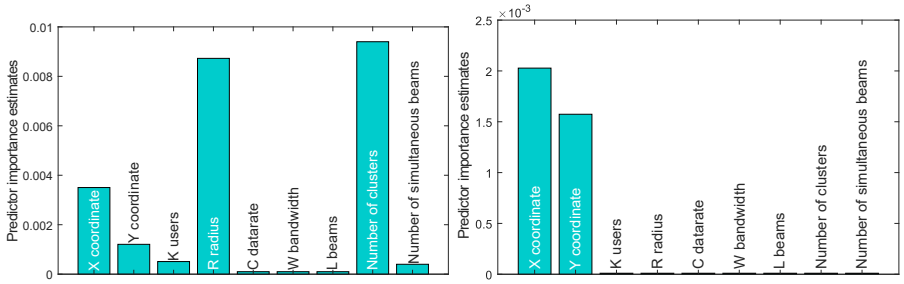
Fig. 2.6. Subgroup assignment accuracy, σ, for $H_1 = H_2 = 5000$, $K = 13$ [2].

Table 2.5. Subgroup and resource matching accuracy, $H_1 = 5000$, $H_2 = 5000$, $K = 13$.

Radius	**100m**	**150-225m**	**250m**	**275m**	**300m**
Fine Tree					
UE assignment, σ	100%	100%	99.02%	29.35%	29.58%
Resources, γ	100%	100%	100%	98.51	96.97%
Logistic Regression					
UE assignment, σ	100%	100%	99.96%	29.41%	31.30%
Resources, γ	100%	100%	100%	100%	98.53%
Kernel Naive Bayes					
UE assignment, σ	100%	100%	99.17%	28.88%	30.19%
Resources, γ	100%	100%	100%	98.44%	95.39%
Cubic SVM**					
UE assignment, σ	99.98%	NaN/100%	99.92%	20.74%	20.66%
Resources, γ	100%	NaN/100%	100%	85.00%	96.88%
Weighted KNN					
UE assignment, σ	100%	100%	99.67%	24.72%	26.91%
Resources, γ	100%	100%	100%	96.92%	98.53%
Random Forest					
UE assignment, σ	100%	100%	99.21%	29.86%	29.13%
Resources, γ	100%	100%	100%	96.92%	100%
Narrow NN					
UE assignment, σ	100%	100%	99.96%	30.67%	30.84%
Resources, γ	100%	100%	100%	98.53%	100%
Coarse Tree					
UE assignment, σ	100%	100%	99.55%	26.65%	26.83%
Resources, γ	100%	100%	100%	59.42%*	90.2%*

*the algorithm defines 5 clusters (on average) instead of 13

** no solution for 150, 200 m, accuracy is 100% is for 225 m



(a) Full dataset (b) Fixed radius $R = 300\,\mathrm{m}$

Fig. 2.7. Variables' importance estimates [2].

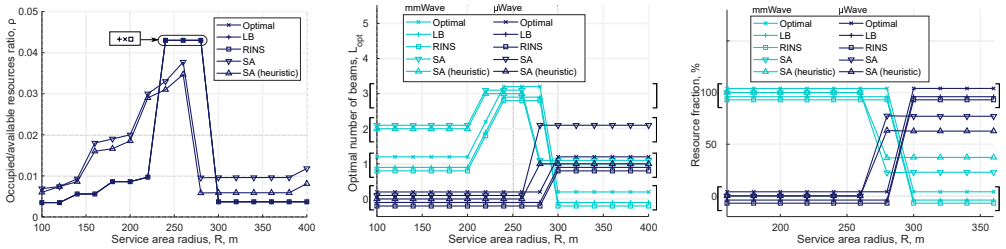
Being incapable of learning specific UE allocations to individual subgroups shown by σ does not indicate that the considered ML algorithms cannot learn other specifics of UE classification. To demonstrate it, we provide resource matching accuracy, γ, for various BS service area distances, R, in Table 2.5. As one may observe, several algorithms show excellent performance. Specifically, of interest are tree algorithms showing excellent extrapolation capabilities as well. As one may notice, Random Forest and Fine Trees provide almost 100% accuracy in terms of resource utilization, γ, over all the considered distances. Surprisingly, simple logistic regression with minimal computational complexity also shows excellent performance. By recalling that rather small computational efforts characterize trees, one may consider them as the best candidates for subgroups formation.

ML Predictors' Importance

Recall that to construct a dataset, we selected many variables of interest. However, these variables may or may not be utilized by the algorithms for classifications. Now, we proceed by exploring the question of what the dominant variables mostly affect the performance of the algorithms. To this aim, in Fig. 2.7, we provide predictor importance for the classification ensemble of decision trees. It computes the estimates of predictor importance for the dataset by summing these estimates over all weak learners in the ensemble. Note that a high value indicates that this predictor is important for the model.

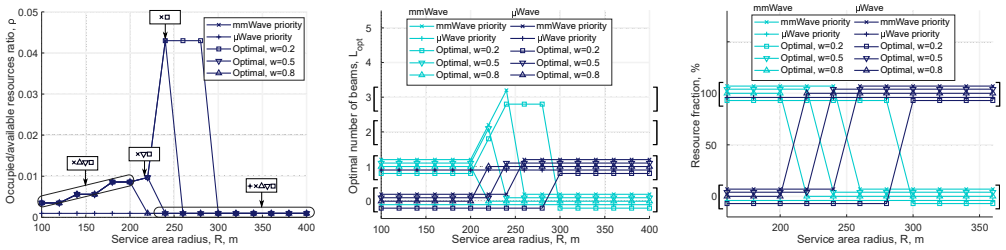
First, Fig. 2.7(a) illustrates the importance of the complete dataset, where we analyze the model's behavior as a function of the service area radius, R. Hence, it is predictable that R variable has high importance. However, we expected UEs' coordinates to be the most important model features. In contrast, the number of clusters obtained from the solution of the exact optimization problem and cell radius are two key predictors that affect the learning process, followed by UEs' coordinates.

Further, by studying Fig. 2.7(b), one can deduce that the importance of the predictors is dataset-specific. Here, fixing the radius R leads to the UEs' coordinates domination. This behavior can be explained by the fact that in directional multicast systems, the radius of the service area impacts the type of transmission utilized for service (i.e., multicast for multiple UEs or unicast for each multicast UE). Our numerical results confirm that the solution



(a) Ratio of occupied to available resources (b) Optimal number of beams (c) Utilized mmWave and μWave resources

Fig. 2.8. Performance metrics when mmWave resources are utilized whenever possible (mmWave RAT priority): mmWave – $\mu_m = 3$, μWave – $\mu_\mu = 0$.



(a) Ratio of occupied to available resources (b) Optimal number of beams (c) Utilized mmWave and μWave resources

Fig. 2.9. Performance metrics for weighted optimization function: mmWave – $\mu_m = 3$, μWave – $\mu_\mu = 0$.

mainly depends on the cell radius. For example, we can observe that a single subgroup is selected for the radius range $100 - 225$ m when we vary the number of UEs in the system. Then, for the range 275 m and further, unicast transmissions are exclusively utilized to serve multicast UEs, whereas the considered multicast group formation solutions can be utilized for the radii around 250 m.

Multi-RAT

mmWave Priority. Regime Switching

The results of the performance analysis when mmWave resources are utilized whenever possible are shown in Fig. 2.8 for mmWave numerology $\mu_m = 3$, μWave numerology $\mu_\mu = 0$, $K = 10$ UEs, $C = 5$ Mbps, $W_m = 100$ MHz, $W_\mu = 50$ MHz, $L_m = L_\mu = 5$ beams. Here, we start by analyzing the ratio of occupied to available resources, ρ, as a function of cell radius, R, illustrated in Fig. 2.8(a). As a general trend, one may notice that ρ grows with the increase in the cell radius until it reaches the distance at which no mmWave coverage is available due to propagation and blockage conditions. At this point, the system starts selecting μWave as a transmission technology. For example, in the case of the optimal solution, $R = 300$ m can be considered as a *threshold* that defines the change in the utilized transmission technology.

Once this threshold is exceeded, the optimal solution always chooses the subgroup containing all K UEs for μWave transmission.

We emphasize that the relaxation techniques (LB, RINS) show a perfect match with the globally optimal solution. On the other hand, the simulated annealing algorithms demonstrate slightly worse results but with better optimality vs. complexity trade-off than optimal solutions. By comparing the simulated annealing algorithms, we may learn that starting with a good solution (compared to the random one) at some points brings us a better value of ρ. This can be explained by the fact that heuristic-based simulated annealing can find a better solution by the time the stopping criterion is met. As our additional observation, we note that the fewer the number of subgroups chosen, the fewer resources they demand.

Optimal Number of Beams

We further comment on the optimal number of beams utilized in the multi-beam dual system as a function of the cell radius illustrated in Fig. 2.8(b). The optimal number of mmWave beams, L_m, starts with one beam (when all UEs form a single subgroup) and then increases up to 3 beams. On the contrary, up to one μWave beam can be swept at a time (and up to 2 μWave beams for random simulated annealing). As one may notice, μWave transmissions are utilized when mmWave fails to provide the service due to propagation conditions and blockage. We emphasize that μWave BS sweeps one beam as, first, it is possible to provide services to all UEs by using the wide beam (small propagation losses) and, second, it ensures the best ratio of occupied to available resources, ρ. We also note that the utilized HPBWs for μWave antennas are larger than those of mmWave technology as the former is employed for subgroups having UEs located farther away from each other. In contrast, mmWave technology typically serves individual UEs in the unicast way or very clustered subgroups of UEs.

RAT Priority Selection

Observe that μWave priority completely excludes mmWave resources, thereby fully loading μWave technology. A network operator may want to avoid it as μWave technology needs to be utilized in those areas not accessible for mmWave. On the other hand, the mmWave priority scheme exclusively utilizes mmWave resources up to a certain distance and then switches to μWave technology. In practice, an operator might have different preferences for balancing resource utilization between considered RATs. To this end, we continue by investigating the impact of the weighted optimization function on the system performance. The corresponding results are shown in Fig. 2.9 for mmWave numerology $\mu_m = 3$, μWave numerology $\mu_\mu = 2$, $K = 10$ UEs, $C = 5\,\text{Mbps}$, $W_m = 100\,\text{MHz}$, $W_\mu = 50\,\text{MHz}$, $L_m = L_\mu = 5$ beams.

By analyzing the data presented in Fig. 2.9, we emphasize that increasing w in (2.31) leads to the shift in the priority from mmWave to μWave. One may learn that at lower distances R, weights $w = 0.2, 0.5, 0.8$, do not affect the performance and provide results similar to the mmWave priority scheme. This can be explained by the fact that mmWave ensures more efficient resource utilization at smaller distances. Further, note that the choice of $w = 0.5$

produces a similar effect to mmWave priority; thereby utilizing μWave band resources only when mmWave service is infeasible due to propagation and blockage conditions. Alternatively, $w = 0.2$ increases the range of mmWave technology up to 280 m (compared to 240 m in the case of mmWave priority), whereas $w = 0.8$ shortens R to 200 m, thereby allowing μWave band usage. We can conclude that depending on the operator's preferences, weights can be properly adjusted to achieve a particular goal with respect to resource usage in dual-mode mmWave/μWave systems.

Our numerical results illustrate that properties of the optimal solution, such as resource utilization and the type of technology, heavily depend on the density of dual-mode BS deployments and RAT priority. Further, the utilized numerology may quantitatively affect the abovementioned conclusions, but the overall qualitative trends remain unchanged. The investigated RAT selection priorities reveal that when μWave RAT is prioritized for multicast service, mmWave resources are not utilized at all. However, by using weights for mmWave and μWave resources, the operator might achieve the desired balance by fitting its needs in a particular deployment. Finally, we note that the efficiency of resource utilization for multicast service may also be affected by the number of UEs and utilized numerologies.

2.2 Conclusions

This chapter introduced the concept of optimal multicasting in mmWave 5G systems. The capability of modern antenna arrays to utilize multiple beams simultaneously with potentially varying half-power beamwidth and asymmetric power allocation makes the problem of efficient multicast transmission in mmWave NR systems an extremely complex one. In this chapter, we solve this problem by developing an optimal multicast grouping and resource allocation solution. The approach is based on a variable-sized bin packing problem and is thus NP-hard. We have developed several heuristics with different complexities and approximation accuracies to provide practical algorithms with reduced computational requirements.

In our numerical results in the case of single-RAT, we utilize the developed optimal approach for benchmarking heuristic solutions. We show that a widely used group formation algorithm originally proposed in [44, 45] may drastically overestimate the amount of resources. The proposed exhaustive search group formation is nearly optimal but computationally intensive for large values of the number of users. The difference between the optimal and heuristic solutions increases with the number of users and the maximum number of supported beams by the antenna array and decreases with the amount of available bandwidth. The type of power allocation among the identified number of beams does not drastically affect the performance of the heuristic algorithms. Finally, for practical ranges of cell sizes and ranges of the number of users (10-50), the optimal amount of beams is always in the range of 2-3. For small cell radii, a single beam is almost always utilized, while unicast service is only feasible for higher ones. This makes the development of heuristic algorithms easier and levels down the requirements for practical antenna array implementations.

Further, by applying the discrimination procedure via comparing modeling and exact optimal solution to the considered set of ML approaches, we revealed that tree algorithms show the best performance for the multicast problem. The number of splits of the trees also matters as Fine and Bagged Trees outperform the Coarse Tree, which has a much smaller amount of splits. The factors mainly responsible for the accuracy of ML approximations are the cell service area and UE coordinates, in addition to "external" knowledge of the number of multicast subgroups provided during the training process. We also discovered a narrow range of the cell area radius R where one has to solve multicasting problems in 5G NR systems with directional systems. Specifically, multicasting with one wide beam for small cell radii leads to the optimal solution. For large cells, unicast transmissions represent the optimal solutions to the multicast problem. There is a narrow range between these two extremes, reported to be $225-275$ m for the considered system parameters, where the optimal solution is non-trivial.

Also, inspired by the prospective 5G NR integrated mmWave/μWave deployments and advanced antenna systems designs capable of simultaneously supporting multiple directional beams, we have provided a globally optimal solution for multicast grouping. Accounting for the NP-hard nature of the problem, we have then proposed and characterized the approximate simulated annealing approach as an efficient solution methodology.

Our numerical results illustrate that properties of the optimal solution, such as resource utilization and the type of utilized technology, heavily depend on the density of dual-mode BS deployments, RAT priority, and considered system parameters. There is a clear turning point for small dual-mode BS densities when the system switches from the regime when mmWave resources are utilized for service to the case when μWave technology is exclusively utilized. This point is dictated by the mmWave blockage and propagation conditions. The number of beams associated with optimal solution is upper limited by 3 for mmWave and by 2 for μWave technologies across all the considered densities of dual BS deployment. Moreover, in most cases, only one beam is utilized at μWave technology. Further, the utilized numerology may quantitatively affect the abovementioned conclusions, but the overall qualitative trends remain unchanged. The investigated RAT selection priorities reveal that when μWave RAT is prioritized for multicast service, mmWave resources are not utilized at all. However, by utilizing weights for mmWave and μWave resources, the operator might achieve the desired balance by fitting its needs in a particular deployment. Finally, we note that the efficiency of resource utilization for multicast service may also be affected by the number of UEs and utilized numerologies.

Concluding, we also note that the exact solution is feasible for up to 10-15 UEs in a multicast group, while relaxation techniques, such as LB and RINS heuristics, although producing a perfect match with the exact solution, do not reduce the solution time. The approximate simulated annealing techniques decrease the complexity leading to a linear increase in the solution time with the number of UEs. However, this happens at the expense of allocating 10-40% of more resources to serve the multicast group.

The materials of this chapter were published in [1] and [2].

3

5G NR Sidelink Multi-Hop Transmission

This chapter is dedicated to the deployment of D2D communications (also known as ProSe or sidelink transmissions) in cellular networks. Specifically, first, we exploit D2D communications to mitigate the limitations of mmWave multicast systems and provide an optimal solution to highly directional sidelink-assisted multicast communications scheduling considering dynamic systems (moving users). Further, we discuss the benefits of NR sidelink in the case of public safety and factory automation use cases and offer the simulation campaign. At the end of Section 3.2, important future directions for the NR sidelink development from a standardization perspective are highlighted.

3.1 Optimal Scheduling for Highly Directional Sidelink-assisted Multicasting

3.1.1 Motivation

Future wireless networks are anticipated to deliver a wide range of services requiring improved performance compared to the 5G in terms of delivered data rate, tolerated latency, mobility support, and massive access. Such services make use of different kinds of wearable devices, including head-mounted displays, motion-tracked controllers, haptic gloves, and body-tracking sensors [79].

Various future mobile applications, such as camera-assisted automotive driving, virtual reality with rich sensory information, and holographic communications, call for extra-high-demanding service delivery requirements that current communication technologies, operating in the low- and middle-frequency bands, are unable to meet. mmWave communication is considered as a viable way to break through this challenge, enabling multi-Gigabit/s data rates and ultra-low latencies for a high number of devices, due to its wide bandwidth and the compact antenna size allowed by short wavelength communications [39].

Nonetheless, in the case of group-oriented service delivery, mmWave technology alone is not sufficient to run future applications while assuring adequate spectral efficiency. Multicast can further enhance the superiority of mmWave communications by serving multiple users with a single transmission. This results in improved network utility achieved by saving spectrum resources. Coupling mmWave technology with multicast transmissions is becoming an important research trend toward increasing energy efficiency and network throughput [12].

However, in the case of sparse user deployments, serving the whole group with either one wide beam or a set of directional beams may significantly reduce the benefits of exploiting extremely high frequencies.

A technology that can be effectively exploited in such a scenario is D2D. Mainly, D2D involves two devices in close proximity communicating directly without a BS and is performed in 5G over the sidelink, which is defined as the interface between UEs for direct communications. Due to the capability of sidelink-aided communications to achieve ultra-low latency connectivity, high data rates, and ultra-high reliability [11], this technology is forecasted to play in 6G the same key role it actually has in 5G. The main reason why D2D can bring assistance in highly directional multicast communications is that beam narrowing may be achieved by excluding sparse users from the multicast transmission and replacing the transmission from the BS with the establishment of D2D links with nodes interacting in the local range. Moreover, sidelink communications, along with a wide range of technologies, such as sleeping and cell zooming, Massive Multiple-Input–Multiple-Output (MIMO), mmWave, and dense heterogeneous networks, can guarantee energy-efficient resource management [80].

Differently from existing works that focus solely on unicasting/multicasting in mmWave networks or provide heuristic solutions to the complex multicasting problem in directional systems (see, e.g., [43]), we focus on optimal multicast scheduling assisted by sidelink and unicast transmissions. In particular, we take a cue from our previous investigations on multicast data transmission optimization [12] and *consider scenarios that include multicast users moving at low speeds*, such as pedestrians equipped with wearable devices. Differently from [12], where we proposed a framework for mmWave beam coverage estimation, the idea is to split users into groups by leveraging fast algorithms (e.g., unsupervised hierarchical clustering [81], which has received much attention in the literature in this field), and then enable the system to adjust (optimize) the transmission mode to define the best option for users by accounting for distance and channel conditions as they move. Considered options are *(i)* unicasting, *(ii)* sidelink unicasting/multicasting, and *(iii)* mmWave multicasting, and, for each user's group and each occurring condition, the designed optimal policy makes the decision whether to widen the beam or use different beams.

The choice of the transmission mode selection for mobile users can be made with the help of either Reinforcement Learning (RL) algorithms, supervised learning, or optimization techniques. First, RL can outperform the optimization only if we have rapidly changing channels, coverage, and topology, i.e., a large state space [82, 83]. In this case, the problem can be written as Multi-Agent Reinforcement Learning (MARL) with centralized rewards where privacy is an issue [84, 85]. For this purpose, one needs to consider the protocols based on which users communicate when taking actions. Hence, it complicates the system with no particular effect. In [86], authors use multi-agent reinforcement learning for a static scenario. However, no multicast transmissions are considered wherein multiple users are served with the same beam, which makes the users' actions dependent on each other since the data rate of the group is limited by the user with the worst channel conditions and depends on the

users' mobility. Moreover, in our scenario, users do not communicate with each other while moving and are not aware of the other users' actions. Regarding supervised ML, we note that offline learning is not suitable since it is configuration-specific. Namely, the change in transmission power, number of users, area of interest, and other transmission parameters affect the final result. This means that for every configuration, we have to provide offline training. Therefore, our choice turns to an optimization tool.

Our objectives are two-fold: *(i)* throughput maximization required for bandwidth-hungry applications, such as Virtual Reality (VR), and *(ii)* mobility management in directional networks since we focus on a dynamic multicast scenario.

3.1.2 System Model

In this subsection, we introduce the reference scenario and describe traffic, antenna, propagation, blockage, and mobility models. The reference system is depicted in Fig. 3.1. The main notations used throughout this chapter are collected in Table 4.5.

Deployment and Traffic Model

We examine a 5G NR outdoor deployment, wherein all UE devices, such as XR glasses and wearable headsets, are provisioned with mmWave modules to be served by an NR BS that operates in the 28 GHz band. The height of the NR BS is set to h_A, and its coverage radius is R_d, within which all UEs can successfully receive data. The geometric locations of UEs are assumed to be scattered across a plane according to an independent homogeneous point process with a predefined density. In our system, all UEs, $\mathcal{N} = \{1, ..., N\}$, are assumed to be dynamic.

Assumption 1: Traffic Model. *We assume that all UEs from $\mathcal{N}$, located and moving within a certain area of interest, require the same multicast service.* In practical deployment, both multicast and unicast sessions may coexist. We do not consider unicast sessions and focus on a single multicast transmission, mainly to analyze the performance of the proposed framework in case of no "external disturbances" in the system. The problem of the joint management of unicast and multicast traffic is by itself a research problem that deserves particular attention [37, 87].

Note. We point out that the concepts of session type and delivery mode are different. In this study, one multicast session (i.e., data flow/content) only is considered, whereas both multicast and unicast transmission modes can be used to serve multicast UEs.

Antenna Model

We assume that devices transmit directionally with an antenna pattern that is akin to a conical shape, i.e., beamwidths are symmetric in both the vertical and horizontal planes. To

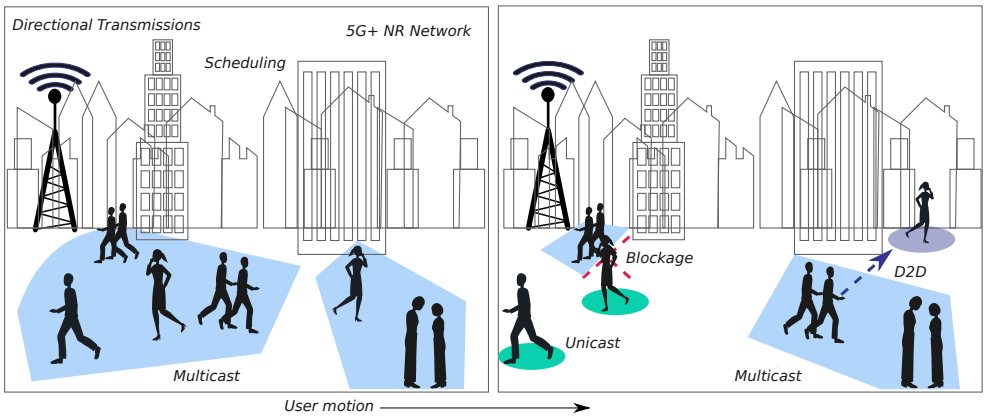


Fig. 3.1. Sidelink-assisted multicast system illustration.

this end, we approximate the beamforming pattern with the following transmit antenna gain as in [88, 89]:

$$G_{\text{tx}} = D_0 \rho(\alpha_i), \tag{3.1}$$

where D_0 is the maximum antenna directivity along the antenna boresight, α_i is the angular deviation of the transmit/receive direction from the boresight of a directional antenna for receiver i, $i \in \mathcal{N}$, and $\rho(\alpha_i) \in [0; 1]$ is a piecewise-defined linear function that scales the directivity D_0 with respect to the angular deviation [88, 89].

Table 3.1. System modeling notation

Parameter	**Definition**
f_c	Carrier frequency
W	Available bandwidth
R_d	Radius of area of interest
h_A	Height of NR BS
h_U	Height of UE
h_B	Height of blocker
r_B	Radius of blocker
y_i	Distance between UE i and NR BS
$p_B(y_i)$	Blockage probability of UE i
λ_B	Density of blockers
N	Number of multicast UEs
$\mathcal{N}$	Set of multicast UEs
D_0	Antenna directivity
α_i	Angular deviation from antenna boresight of UE i
$G_{\text{tx}}, G_{\text{rx},i}$	Antenna array gains at NR BS and UE i ends
P_T	Transmit power
$M_{S,nB}$, $M_{S,B}$	Fading margins
L_{dB}	Path loss in linear and decibel scales
A, ς	Propagation coefficients
S	Signal-to-noise ratio
N_0	Power spectral density of noise
v	UE's velocity
θ	Half-power beamwidth

Propagation and Blockage Model

Following 3GPP standard [6], we exploit the 3GPP Urban Microcell (UMi) street canyon path-loss model:

$$L_{\mathrm{dB}} = 32.4 + 21 \log_{10} y_i + 20 \log_{10} f_c, \tag{3.2}$$

where f_c is the carrier frequency in GHz and y_i is the 3D distance between the BS and the UE i.

5G NR systems that operate in a high-frequency band suffer from the presence of moving obstacles (called "blockers"), including humans and vehicles. Here, pedestrians are assumed to temporarily block the LoS path between the UE and the NR BS, i.e., causing blockage by the human body. This blockage attenuation B is considered to be 15 decibels. We also introduce shadow fading margins that are represented by $M_{S,B}$ and $M_{S,nB}$ for the blocked and non-blocked states, respectively. Then, the path loss in (3.2) may be written in a linear scale using $A y_i^{\varsigma}$, with A and ς being propagation coefficients:

$$\begin{aligned} A_{\mathrm{LoS,nB}} &= 10^{2 \log_{10} f + 3.24} M_{S,nB}, \varsigma_{\mathrm{LoS}} = 2.1, \\ A_{\mathrm{LoS,B}} &= 10^{2 \log_{10} f + 4.74} M_{S,B}, \varsigma_{\mathrm{LoS}} = 2.1. \end{aligned} \tag{3.3}$$

The blockers are modeled as cylinders with height h_B and radius r_B [42]. The number of blockers follows a Poisson distribution with density λ_B per square meter.

Then, the SNR in the propagation model can be represented as

$$S = \frac{P_T D_0 \rho(\alpha_i)}{N_0 W} \left(\frac{y_i^{-\zeta_{\mathrm{LoS}}}}{A_{\mathrm{LoS,nB}}} [1 - p_B(y_i)] + \frac{y_i^{-\zeta_{\mathrm{LoS}}}}{A_{\mathrm{LoS,B}}} p_B(y_i) \right), \tag{3.4}$$

where $p_B(y_i)$ is the blockage probability at the 3D distance y_i [42], N_0 is the noise power spectral density, and W is the operating bandwidth.

Mobility Model

We assume that UEs follow the social force-based mobility model that captures the realism of crowd behaviors [90]. More specifically, we apply the Headed Social Force Model (HSFM) proposed in [91], which can reproduce pedestrians moving together. The HSFM allows us to test the real-life scenario composed of several groups of moving UEs (with speed v) and is relevant to our system as we consider the multicast content delivery for a set of UEs.

3.1.3 Analysis

Framework Description at a Glance

The main goal of the proposed framework is to maximize the system throughput in the delivery of one multicast session to multiple UEs by dynamically selecting the transmission mode (i.e., unicast, multicast, or sidelink) towards each multicast group member according to UEs' mobility. In this subsection, we describe in detail our proposed solution to the

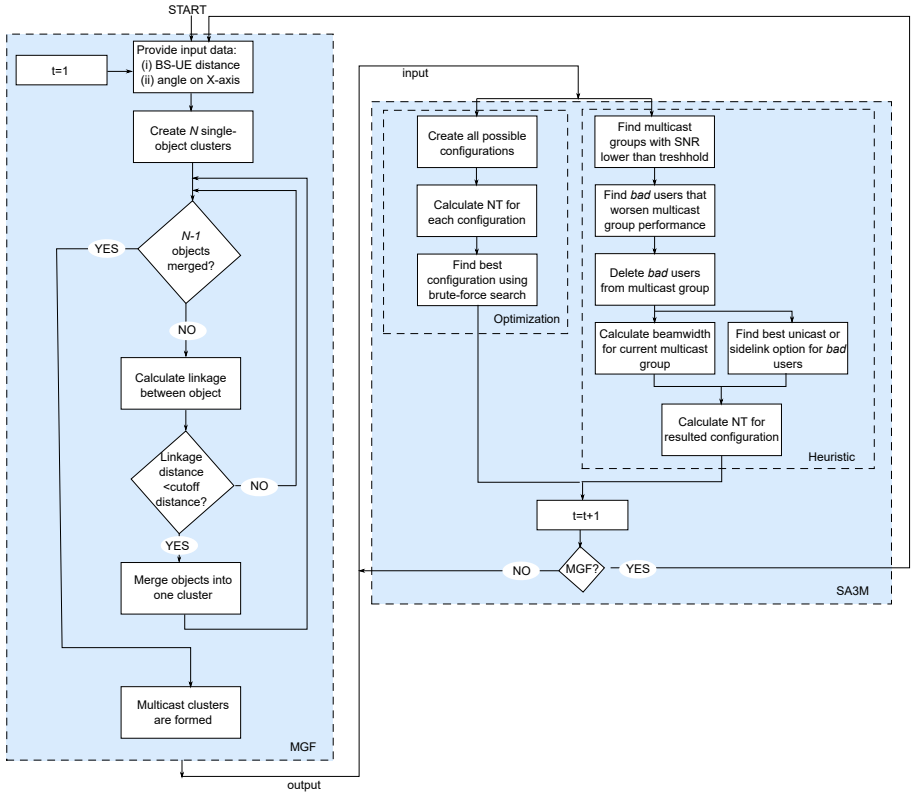


Fig. 3.2. Flow diagram of proposal.

dynamic sidelink-assisted mmWave scheduling problem. The framework consists of two steps, namely, the *multicast group formation (MGF)* and the *Sidelink-Assisted Multiple Modes mmWave (SA3M)* scheduling. The general flow diagram relevant to our proposal is presented in Fig. 3.2. After detailing how MGF works, we discuss the functioning of SA3M, where we also introduce the optimization objective and offer a low-complexity heuristic solution.

First, we perform the MGF step by applying a hierarchical clustering algorithm, which is an unsupervised ML technique that aims to find natural grouping based on the characteristics of the input data. We accomplish this task based on the information about the location of the UEs. We emphasize that an initial clustering of multicast UEs is performed to reduce the complexity of the optimal solution in the second step, as detailed below. Note that different clustering methods can be used at this stage. Among these, we chose hierarchical clustering because it is characterized by low complexity and considers the positions of UEs, which is essential for directional transmissions.

After clustering, starting from the obtained multicast groups configuration generated by MGF, as the UEs are moving with speed v, the system may proceed with SA3M and change the transmission mode either to *(i)* sidelink transmission, *(ii)* unicast mode, or *(iii)* recalculating transmission parameters for a multicast group (beamwidth adjustment). Specifically, starting from the configuration generated by MGF, the optimal SA3M algorithm exhaustively searches for all possible switching options, that is 1,2,..., or all N UEs can be served

via unicast while the rest UEs belong to multicast groups. Similarly, 1,2,..., or $N-1$ UEs can receive the data through sidelink ($N-1$ since at least one UE has to be a relay node towards the sidelink receivers). Note that, in case a UE needs to join another multicast cluster, we run MGF again, which serves as a means to build multicast clusters. Were MGF not implemented as an initial step, the algorithm should check all possible multicast clusters, that is $2^N - 1$ possible multicast groups (rather than the configuration that contains n groups selected by MGF). Then, the procedure described above should be run for sidelink and unicast options for each multicast configuration, which significantly complicates the model. Thus, the *main mission of MGF is to reduce the complexity of the SA3M step.* As an alternative to SA3M, we propose a heuristic solution requiring low run-time to be used for adjusting the transmission modes in order to improve the system performance.

Step I – Multicast Group Formation

Hierarchical clustering, applied for MGF, builds a binary merge tree. It starts from the data elements stored at the leaves (interpreted as singleton sets) and proceeds by merging two by two the *closest* subsets (stored at nodes) until the root of the tree contains all the elements of X. Specifically, in the beginning, each data point is assumed to be a separate cluster. Then, similar clusters are iteratively combined. We denote by $\Delta(X_k, X_j)$ the distance between any two subsets of X, called the linkage distance. This technique is also called *agglomerative hierarchical clustering* [92]. In our case, X represents an array with the observations, with at least one column and N strings (each string corresponds to a UE). The reference angle from the X-axis and the distance between the BS and every UE are used as the observations.

Let $D(x_k, x_j)$ denote the elementary distance between any two elements of X (e.g., Euclidean, Minkowski, Chebyshes, etc.). In order to select the closest pair of subsets at each stage of the hierarchical clustering, we define a subset distance $\Delta(X_k, X_j)$ between any two subsets of elements. When both subsets are singletons $X_k = x_k$ and $X_j = x_j$, then $\Delta(X_k, X_j) = D(x_k, x_j)$. There are four different methods implemented in scikit-learn to measure the similarity between clusters, i.e., four *common* linkage functions (also known as cluster-level scoring functions) that calculate the distance between clusters:

- Single linkage (SL) represents the shortest distance among all data points in two clusters:

$$\Delta(X_k, X_j) = \min_{x_k \in X_k, x_j \in X_j} D(x_k, x_j).$$

- Complete linkage (CL) represents the farthest distance among all data points in two clusters:

$$\Delta(X_k, X_j) = \max_{x_k \in X_k, x_j \in X_j} D(x_k, x_j).$$

- Average linkage uses the average distance between all pairs of objects in any two clusters:

$$\Delta(X_k, X_j) = \frac{1}{|X_k|} \frac{1}{|X_j|} \sum_{x_k \in X_k} \sum_{x_j \in X_j} D(x_k, x_j).$$

Algorithm 7: MGF

1 **Input:** X;
2 **Output:** Multicast clusters;
3 **Initialize** $\mathcal{G}_k = \{x_k\}$, $k = 1, ..., N$, $\mathcal{L} = \{\{x_1\}, \{x_2\}, ..., \{x_N\}\}$, *distance threshold*;
4 *counter* $\leftarrow N$;
5 **while** *counter* $\neq 2$ **do**
6 **Select** $\mathcal{G}_k$ and $\mathcal{G}_j$ from $\mathcal{L}$ such as $\Delta(X_k, X_j)$ is minimized along all pairs;
7 **if** $\Delta(X_k, X_j) <$ *distance threshold* **then**
8 **Merge** $\mathcal{G}_k \cup \mathcal{G}_j$;
9 $\mathcal{L} \leftarrow \mathcal{L} \setminus \mathcal{G}_k$;
10 $\mathcal{L} \leftarrow \mathcal{L} \setminus \mathcal{G}_j$;
11 $\mathcal{L} \leftarrow \mathcal{L} \cup (\mathcal{G}_k \cup \mathcal{G}_j)$;
12 *counter* $\leftarrow N - 1$;
13 **else**
14 go to line 5;
15 **end**
16 **end**
17 **return** $\mathcal{L}$;
18 **end**

- Ward linkage (appropriate for Euclidean distances only) uses inner squared distance, i.e., minimum variance algorithm:

$$\Delta(X_k, X_j) = \sqrt{\frac{2|X_k|X_j|}{|X_k| + |X_j|}} \|\overline{x}_k - \overline{x}_j\|_2,$$

  where $\|\|_2$ is the Euclidean distance, $\overline{x}_k$, $\overline{x}_j$ are the centroids of clusters X_k and X_j, respectively.

We note that, in the case of hierarchical clustering, the number of clusters may not be determined in advance as, for example, in the case of the k-means algorithm. Here, either a cutoff distance or a maximum number of clusters must be specified. In this study, we exploit a cutoff distance, which is the linkage distance threshold above which clusters will not be merged. The pseudo-code of the hierarchical clustering adapted for MGF is presented in Algorithm 7.

The algorithm assigns each observation in X to a single-object cluster (line 3). Then, the algorithm computes similarity information between every pair of objects $\mathcal{G}_k$ and $\mathcal{G}_j$ in the data set and uses a linkage function to group objects into a hierarchical cluster tree (line 6). Therefore, objects/clusters in close proximity are linked together if the result of the linkage function does not exceed the cutoff distance *distance threshold* (lines 7-14). This determines where to cut the hierarchical tree into clusters, thereby partitioning the data.

Table 3.2. 5G NR numerology and subcarrier spacing [3].

μ	**Df**$= 2^{\mu} \cdot 15$ **[kHz]**	**Bandwidth per RB [kHz]**	**TTI [ms]**	**Slots / ms**
0	15	180	1	1
1	30	360	0.5	2
2	60	720	0.25	4
3	120	1440	0.125	8
4	240	2880	0.0625	16

Since we start from $counter = |X| = N$ leaves to finish with a root containing the full set X, the algorithm performs exactly $N-1$ merge-operations. A straightforward implementation of this algorithm yields a cubic time complexity, in $O(N^3)$, since, in the k-th iteration of $N-1$ in total, all $\binom{N-1-k}{2}$ pairwise distances between the $N-k$ nodes in $\mathcal{L}$ are searched [93].

Step II – Optimization

Assumption 2: Transmission modes. *Unicast, sidelink unicast, and multicast are the transmission modes that can co-exist in a cell for the transmission of the same content.* The UE tunes into the corresponding channel for data reception according to the optimization problem described in the following.

In our study, we consider a dynamic scenario where time is divided into discrete time slots t of constant duration. 5G NR utilizes the scalable numerology that determines the subcarrier spacing, the number of slots in a subframe, and the slot duration (see Table 3.2). At each time slot t, UEs can be associated with different transmission modes depending on the channel conditions, i.e., as per (3.5), (3.8), (3.10), and the results of the optimization.

The SA3M optimization deals with choosing the best network configuration in terms of the considered metric of interest (see Algorithm 8). To create the network configurations, we assume the following rules: *(i)* UEs can not join a multicast group different from the one defined using Algorithm 7 (we rerun MGF to form distinct multicast groups at a given rerunning interval, see subsection 3.1.4 for consideration on how to set it); *(ii)* the predefined multicast transmission mode can be switched into unicast or sidelink for each UE following the SA3M algorithm to improve the network throughput.

As a preliminary step, the algorithm creates all possible network configurations that determine the transmission modes for all UEs (lines 4-5). Among them, the BS will choose (through exhaustive search) the one that optimizes the network performance. The number of possible network configurations is $2 \cdot 2^N - 1$ since there are 2^N possible combinations of 0 and 1, where 1 means that the UE remains in the multicast group determined by MGF and 0 represents a switch of the transmission mode to unicast. In the case of sidelink mode, we have $2^N - 1$ options as one of the devices should always be considered as a relay. Hence, all network configurations are in $2 \cdot 2^N - 1$.

Then, depending on the network configuration, UEs can be associated with different transmission modes, and SNR thereof is determined as follows.

Multicasting. Multicast services are multi-user specific, and the quality of the channel is determined by the UE experiencing the worst channel conditions, i.e.,

$$S_m(t) = \min_{i \in \mathcal{G}_j} \left(\frac{P_T D_0 \rho(\alpha_i(t))}{N_0 W} \left[\frac{y_i(t)^{-\zeta_{\text{LoS}}}}{A_{\text{LoS,nB}}(t)} [1 - p_B(y_i(t))] + \frac{y_i(t)^{-\zeta_{\text{LoS}}}}{A_{\text{LoS,B}}(t)} p_B(y_i(t)) \right] \right), \tag{3.5}$$

where $\mathcal{G}_j$ is the set of UEs in a multicast group covered by the same beam j, $|\mathcal{G}_j| \leq N, \mathcal{G}_j \subseteq \mathcal{N}$.

The time required for the transmission of a packet of size B to a multicast subgroup when experiencing the channel condition $S_m(t)$ can be calculated as

$$T_m(t) = \frac{B}{W_m \log_2(1 + S_m(t))}. \tag{3.6}$$

Hereinafter, we omit the slot notation (t) for the sake of space.

The HPBW θ required to serve subgroup $\mathcal{G}_j$ is given by:

$$\theta_{\mathcal{G}_j} = \arccos\left(\frac{X(i)X(i') + Y(i)Y(i') + Z(i)Z(i')}{y(i)y(i')} \right), \tag{3.7}$$

where multicast UEs i and i' are the two edge UEs in the group, i.e., the two farthest in term of angle between them.

Unicasting. mmWave unicast transmission facilitates expanding the coverage area by sweeping narrow beams (e.g., HPBW of 2°). Then, the UE that fails to be served as a part of multicast transmission can prefer unicasting with the following link quality and data transmission duration:

$$S_u = \frac{P_T D_0}{N_0 W} \left(\frac{y_i^{-\zeta_{\text{LoS}}}}{A_{\text{LoS,nB}}} [1 - p_B(y_i)] + \frac{y_i^{-\zeta_{\text{LoS}}}}{A_{\text{LoS,B}}} p_B(y_i) \right), \tag{3.8}$$

$$T_u = \frac{B}{W_u \log_2(1 + S_u)}. \tag{3.9}$$

D2D Unicasting. We assume in-band D2D, wherein UEs share the licensed uplink frequency resources with cellular communications. The channel link for UEs who choose to access the D2D multicast network can be determined as

$$S_d = \frac{P_{T,d} D_0}{N_0 W} \left(\frac{y_{i,d}^{-\zeta_{\text{LoS}}}}{A_{\text{LoS,nB}}} [1 - p_B(y_{i,d})] + \frac{y_{i,d}^{-\zeta_{\text{LoS}}}}{A_{\text{LoS,B}}} p_B(y_{i,d}) \right), \tag{3.10}$$

where $y_{i,d}$ is the distance between UE i and D2D transmitter, and the data transmission delay can be calculated as

$$T_d = \frac{B}{W_d \log_2(1 + S_d)}. \tag{3.11}$$

Assumption 3: D2D Relay Selection. *The BS selects the possible D2D transmitter (relay device) based on the following rules: (i) distance between a relay and a UE has to be within $D2D_{thr}$ (i.e, $y_{i,d} < D2D_{thr}$), (ii) we choose the closest relay among those that satisfy the previous condition, and (iii) a relay can transmit data to one or more UEs (more details are given in the following).* We assume that a relay can simultaneously receive and transmit data to the UE (i.e., Full-Duplex (FD) relaying). If no relay satisfies the described conditions,

Algorithm 8: Optimal SA3M

1 **Input:** Multicast clusters $\mathcal{L}$;
2 Coordinates of N multicast UEs $(X(i), Y(i), Z(i)), i \in \mathcal{N}$
3 **Output:** Optimal network configuration;
4 Create all 2^N possible network configurations considering unicasting;
5 Create all $2^N - 1$ possible network configurations considering D2D transmissions;
6 **for** *each network configuration* **do**
7 $\quad T_{\text{total}}^{\text{NC}} = \sum_{m \in \mathcal{G}} T_m + \sum_{u \in \mathcal{U}} T_u$.
8 **end**
9 Solve optimization as per (3.12).

a D2D transmission link cannot be established with a particular UE. In this case, unicast transmission shall be performed instead.

We consider two relay selection options for the purpose of accounting for different hardware on the devices. In the case referred to as "D2D communication without restrictions", a relay device can convey the traffic to more than one UE at a time. Differently, a simple device works in the category of "D2D with restrictions", wherein a relay can transmit data to only one UE at a time.

Assumption 4: Interference. *In our system, we assume that the power transmitted by the relay node is lower than the power emitted by the BS, that is $P_{T,d} < P_T$, which helps to avoid that D2D communication causes exceeding interference [94].* Recall that in-band D2D UEs reuse the same uplink resources of the mmWave cell that can cause interference.

Optimization Objective. A multicast UE can receive data at different rates depending on its current location and blockage conditions. The optimization objective consists in maximizing the Network Throughput (NT) (i.e., aggregated throughput optimization). NT is calculated as the sum of data rates that are delivered to all UEs in the network.

Here, the problem consists in solving the overall maximum throughput optimization problem that is formulated as follows:

$$\max \frac{BN}{\sum_{m \in \mathcal{G}} T_m + \sum_{u \in \mathcal{U}} T_u}, \tag{3.12}$$
$$\text{s.t. } S_m \geq S_{\text{thr}}, S_u \geq S_{\text{thr}}, S_d \geq S_{\text{thr}}, y_{i,d} < y_{\text{thr}}.$$

Proposed Heuristic

The proposed heuristic, detailed in Algorithm 9, works as follows. It starts to check if the SNR of each multicast group from $\mathcal{L}$ satisfies $S_{\text{thr,h}}$ (line 6). If the SNR of the group (i.e., the worst SNR value among the group members as per (3.5)) is lower than the $S_{\text{thr,h}}$ value, then the algorithm proceeds with checking every UE i in this group (lines 7-8). In this case, UE i is removed from the group and added to a separate one, $\mathcal{G}_g$, if its SNR value is below the

Algorithm 9: Heuristic Solution

1 **Input:** Multicast clusters $\mathcal{L}$;
2 Coordinates of N multicast UEs $(X(i), Y(i), Z(i)), i \in \mathcal{N}$
3 **Output:** Network configuration;
4 $g \leftarrow N$;
5 **for** *each* $\mathcal{G}_j \in \mathcal{L}$ **do**
6 **if** $S_{\mathcal{G}_j} < S_{thr,h}$ **then**
7 **for** *each UE* $i \in \mathcal{G}_j$ **do**
8 **if** $S_i < S_{thr,h}$ **then**
9 $g \leftarrow g + 1$;
10 $\mathcal{L} \leftarrow \mathcal{L} \setminus \mathcal{G}_j$;
11 $\mathcal{G}_j \leftarrow \mathcal{G}_j \setminus \{x_i\}$;
12 $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{G}_j$;
13 $\mathcal{G}_g \leftarrow \{x_i\}$;
14 $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{G}_g$;
15 **end**
16 **end**
17 **end**
18 **end**
19 **for** *each* $\mathcal{G}_j \in \mathcal{L}$ **do**
20 **if** $|\mathcal{G}_j| > 1$ **then**
21 calculate θ as per (3.7);
22 calculate S_m, T_m as per (3.5),(3.6);
23 **else**
24 find $\max_{m \in \mathcal{M}^*} \{S(y_{i,m}) | y_{i,m} < \mathrm{D2D_{thr}}\}$; $\{y_{i,m}$ is the distance between UEs i and m, $\mathcal{M}^*$ is a set of UEs served via multicast$\}$
25 calculate T_d as per (3.11);
26 calculate S_u, T_u as per (3.8),(3.9);
27 choose best option as $\min(T_d, T_u)$;
28 **end**
29 **end**
30 **return** $\mathcal{L}$;
31 **end**

threshold (lines 6-14). That is, the algorithm detects the UEs that deteriorate the multicast group performance. The second *for* cycle of the algorithm is responsible for the calculation of the beamwidth θ of the group (lines 16-19) and for the selection between sidelink and

unicast modes for single UEs (lines 19-23). All groups are already reformed at this stage, and the algorithm needs to adjust the beamwidth to be swept as per (3.7).

The computational complexity of the proposed algorithm is given by

$$O((|\mathcal{L}| \cdot N) + |\mathcal{L}|) = O(|\mathcal{L}| \cdot N) = O(N^2),$$

where each summons (on the left side of the expression) determines the complexity of each *for* cycle. Then, as each cycle is called in turn (sequential execution), the complexity of the algorithm is $O(|\mathcal{L}| \cdot N)$. We note that $|\mathcal{L}| = N$ in the worst case when we have only unicast UEs. Hence, the algorithm's complexity is polynomial and can be rewritten as $O(N^2)$. Note that our algorithm has embedded *for* cycle (lines 7-14). We highlight that this execution helps reduce the complexity when not all multicast groups contain a "bad" UE that deteriorates the group's performance. These two cycles could be substituted by one *for* cycle among all N UEs. That is, we could check all UEs without exclusion.

3.1.4 Performance Assessment

This subsection evaluates the performance of proposed MGF and SA3M algorithms, as well as the proposed heuristic solution. To this aim, we developed a simulation environment in MATLAB that accepts the default parameters summarized in Table 3.3. We first select the linkage function that works better for MGF. We then proceed with a numerical analysis of the introduced optimal SA3M algorithm and discuss the effects of mobility, complexity, and UEs' distribution on the system performance. Finally, we report on the performance of the proposed low-complexity heuristic and analyze the trade-off between transmit power and energy consumption.

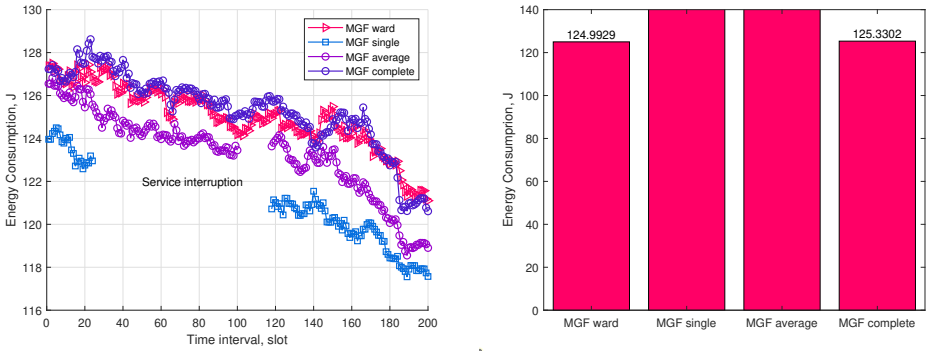
Without losing generality, we adopt the following parameters for the performance evaluation. We distribute $N = 10$ UEs within a sector of radius 100 m according to a Poisson Point Process (PPP) and Matérn cluster point process with 2 clusters. The transmission parameters are modeled as indicated in the previous subsection 3.1.3 with the operating frequency of 28 GHz and transmit power of 46 dBm. The bandwidth is 1 GHz [95] and the noise figure is 7.6 dB. The beam parameters are adjusted depending on the position of UEs in the multicast groups, whereas unicast and sidelink UEs utilize the antenna with beamwidth of 3.18°. The mobility pattern of UEs is simulated as HSFM.

The metrics of interest we analyze in the following are: *(i)* energy consumption, measured in joules (J), computed as the number of power units consumed over transmission time, *(ii)* network throughput representing the sum of data rates delivered to all UEs in the network, and *(iii)* energy efficiency calculated as network throughput divided by consumed energy. Note that all three metrics are linked, and maximization of throughput leads to optimizing all of them. Therefore, for the sake of space, we plot only one of the metrics depending on the investigated scenario. We emphasize that metrics are calculated for the resulted configuration after *(i)* all the two algorithms (MGF+SA3M) for optimization are performed, and the results are obtained, *(ii)* heuristic provided the solution.

Table 3.3. Default parameters for numerical evaluation.

Parameter	**Value**
Carrier frequency, f_c	28 GHz
Available bandwidth, W	1 GHz
Height of NR BS, h_A	10 m
Height of UE, h_U	1.5 m
Height of blocker, h_B	1.7 m
Radius of blocker, r_B	0.4 m
Density of blockers, λ_B	0.3 bl./m^2
SNR threshold, S_{thr}	-9.47 dB
SNR threshold for heuristic, $S_{\text{thr,h}}$	6.367 dB (CQI 8)
Transmit power, P_T	46 dBm
D2D transmit power, $P_{T,d}$	10 dBm
Power spectral density of noise, N_0	-174 dBm/Hz
UE planar antenna elements, N_U	4 el
Velocity, v (pedestian/segway)	0.69 m/sec, 11 m/sec
Service area radius, R	100 m [1, 43]
Number of multicast UEs, N	10 [1, 43]
Fading margins, $M_{S,nB}$, $M_{S,B}$	4/8.2 dB
Packet size, B	1 Gb
Linkage function	ward

Effect of Linkage Function



(a) (b)

Fig. 3.3. Impact of linkage functions on energy consumption for hierarchical clustering: (a) over 200 time slots, (b) mean value. Uniform user distribution.

In our model, we test four general linkage functions for clustering under MGF that may affect performance at the NR BS. We first analyze their impact on energy consumption, which is one of the metrics of interest, as illustrated in Fig. 3.3. By observing Fig. 3.3(a), one can notice the decreasing trend of the curves, i.e., an improvement in energy consumption due to the capability of better tracking UE mobility by rerunning the MGF algorithm at every time slot. It is worth noting that *single* and *average* linkages show better performance in terms

of the minimal value of energy consumption while at times failing to maintain minimum allowable service quality (i.e., CQI 1). Then, as indicated in Fig. 3.3(b), *ward* function provides slightly better performance in terms of mean energy consumption compared to *complete* and much better than the other two, meaning that ward and complete functions can capture directional multicast transmission features. Hence, in the remainder of the study, we consider ward linkage as a default parameter for the MGF Algorithm. However, we highlight that linkage functions can perform differently depending on input parameters for clustering.

Effect of Mobility

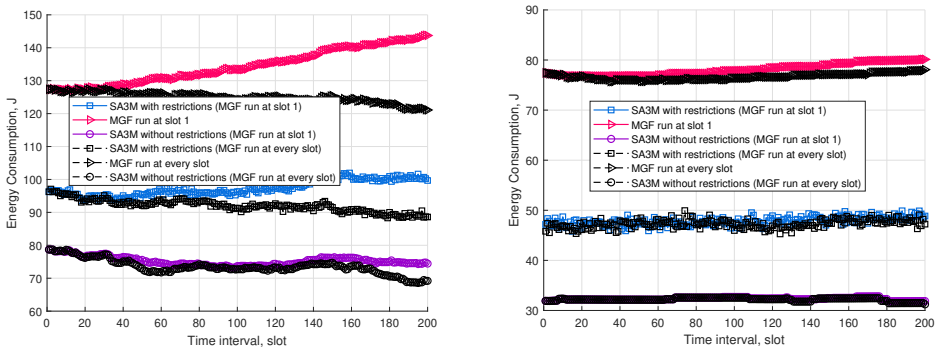
We start our primary evaluation campaign with the analysis of the MGF and SA3M algorithms' performance over time, as shown in Fig. 3.4. To this aim, we run MGF and SA3M with and without restrictions for sidelink relaying in two different modes: *(i)* MGF together with SA3M launched at every time slot and *(ii)* MGF launched at time slot 1 only (no rerunning) and SA3M run every time slot for the two considered user distributions with pedestrian mobility. We note that no rerunning of MGF (see colored curves compared to black ones) affects the performance over time even though the speed and users' mobility are the same. In particular, for uniform distribution, we can see the most noticeable difference between the two running modes of the algorithms. In contrast, the rerunning produces almost no improvement for the 2 cluster distribution as seen in Fig. 3.4(b). This effect can be explained by the fact that, in the case of PPP, UEs are spread around the area of interest. In general, the distance between every two uniformly distributed UEs in the network is higher than in the case of the cluster distribution, which impacts the performance since, in this case, D2D transmissions have to be performed over longer distances and wider beam should be swept to cover multicast groups.

Analyzing further the effect of mobility for segway with $v = 11\,\mathrm{m/sec}$ and pedestrians $v = 0.69\,\mathrm{m/sec}$ in Fig. 3.5, we learn that for faster speeds, MGF rerunning plays a crucial role in maintaining the performance level in dynamic scenarios. As expected, the gap between rerunning vs. no rerunning of MGF is higher for segway mobility. The average performance improves by 11% and 7% for segway and walking UEs, respectively. Hence, it is highly recommended to rerun MGF to maintain the required performance level. In the following subsection, we comment on the rerunning interval of the MGF algorithm and on the complexity of the algorithms.

Complexity vs. Energy Performance Trade-off

We run the simulations via MATLAB R2021a on an Intel(R) Core(TM) i5-7200U CPU @2.50 GHz at 2.71 GHz with 8.00 GB RAM. The observed complexity of the proposed algorithms is summarized in Table 3.4.

We now analyze the complexity/energy performance trade-off (please refer to Fig. 3.6). For this reason, we run additional simulations to compare the performance of SA3M with



(a) Uniform (b) 2 clusters

Fig. 3.4. Energy consumption over time for pedestrian mobility when considering: (a) uniform and (b) 2 clusters. Black lines are drawn in case MGF is rerun at every time slot.

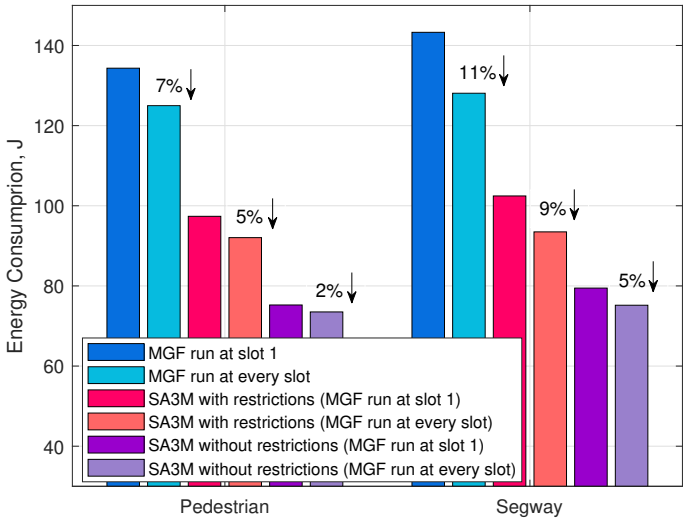
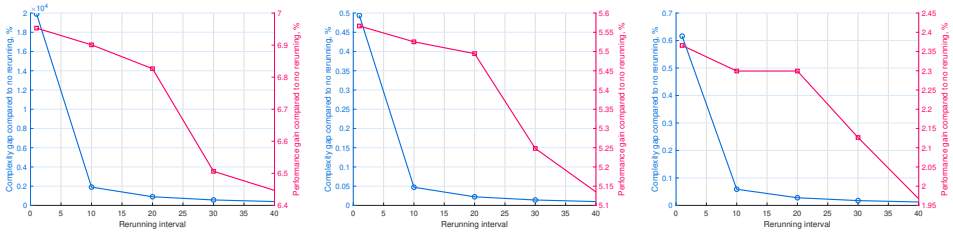


Fig. 3.5. Energy consumption for multicast users moving with different speeds in case of uniform distribution (each second bar is when we rerun MGF at every time slot).

Table 3.4. Algorithms' Complexity, seconds.

MGF	SA3M with restrictions	SA3M without restrictions	Heuristic
0.027602	4.4580, of which:	5.5605, of which:	0.031314
	- 3.870358 to create unicast groups	- 3.870358 to create unicast groups	
	- 0.002252 to create sidelink groups	- 0.002252 to create sidelink groups	
	- 0.585398 to run algorithm	- 1.687939 to run algorithm	

or without MGF rerunning in terms of *(i)* energy consumption gain on the right y-axis and *(ii)* complexity gap on the left y-axis as a function of the MGF rerunning interval. First, let us analyze how MGF behaves when considering rerunning interval values ranging from 1 (i.e., the algorithm runs every single slot) to 40 (i.e., the algorithm runs every 40^{th} slot). By



(a) (b) (c)

Fig. 3.6. Complexity gap vs. energy performance gain compared to no MGF rerunning for uniform distribution of users: (a) MGF, (b) SA3M without restrictions (c) SA3M with restrictions.

observing Fig. 3.6(a), it emerges a high increase in complexity (up to 20000%) for MGF when the rerunning interval is every slot compared to the "no rerunning" case. Moreover, one may observe a noticeable drop in complexity for rerunning interval values ranging from 1 to 10. By further increasing the rerunning interval, the performance gap between "no rerunning" and "rerunning" slowly decreases, as confirmed by Fig. 3.5.

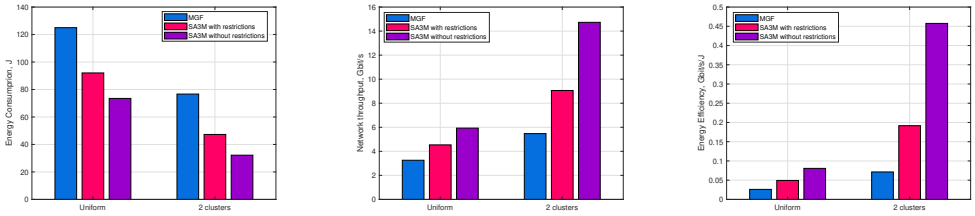
Similar trends are observed in Fig. 3.6(b) and Fig. 3.6(c), in which, however, the observed quantitative increase in complexity is not so significant. This can be explained by the fact that the total complexity of SA3M is vastly greater than that of MGF. Hence, we may conclude that MGF complexity does not contribute to the overall complexity of SA3M. On the other hand, shortening the rerunning interval might be crucial for fast UE speeds depending on the mobility pattern. Thus, taking into account both above-mentioned considerations, our recommendation is to keep the rerunning interval in the range of 10-20 slots, which represents a good trade-off between complexity level and achievable energy performance.

Effect of Users' Distribution

This subsection illustrates the results of our evaluation campaign in terms of three system metrics, namely energy consumption, network throughput, and energy efficiency, as a function of users' distribution. We consider PPP and Matérn cluster point process with 2 clusters. As shown in Fig. 3.7, for all considered metrics, the Matérn cluster distribution of UEs within the sector provides better results compared to the uniform one. The reason is that in the case of Matérn cluster distribution, its randomly located points tend to form random clusters, which is beneficial for multicasting, rather than UEs scattered around the area of interest.

Heuristic Evaluation

To evaluate the performance of the heuristic algorithm, let us examine the impact of two parameters: *(i)* SNR threshold, $S_{\mathrm{thr,h}}$, for removing a multicast UE from the group (as



(a) (b) (c)

Fig. 3.7. Effect of different distributions of multicast users on (a) energy consumption, (b) network throughput, and (c) energy efficiency. MGPF is executed at every time slot.

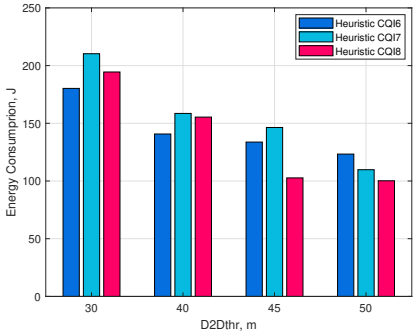
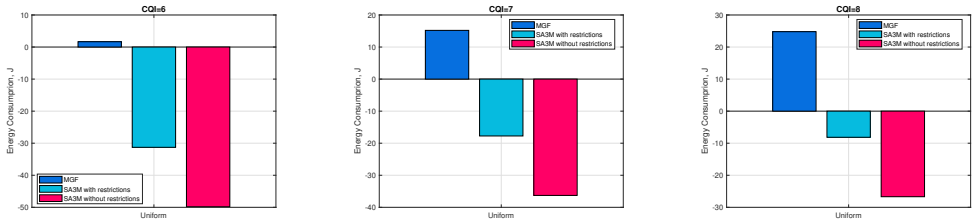


Fig. 3.8. Energy consumption for heuristic solution varying $S_{\text{thr,h}}$ and D2D_{thr} thresholds.

for lines 6-14 of the Heuristic Algorithm) and assigning unicast/sidelink transmissions, and *(ii)* distance threshold at which sidelink communication can be established, D2D_{thr}. For SNR thresholds, we use MCS mappings from [64] provided in Table 3.5.



(a) (b) (c)

Fig. 3.9. Heuristic compared to MGF and SA3M algorithms for (a) CQI 6, (b) CQI 7, and (c) CQI 8. D2D_{thr}=50 m. Effect of different distributions of multicast users on (a) energy consumption, (b) network throughput, and (c) energy efficiency. MGPF is executed at every time slot.

The impact of D2D_{thr} is evaluated in Fig. 3.8. Here, it is shown how a less stringent threshold (i.e., D2D_{thr}=50 m) for distances within which devices can establish sidelink communication outperforms all other thresholds. The reason is that our proposed low-complexity heuristic, rather than implementing an exhaustive search for all possible configurations, works

Table 3.5. CQI, MCS, spectral efficiency, and SNR mapping for 3GPP NR.

CQI	**MCS, code rate x 1024**	**Spectral efficiency** (bits/symbol)	**SNR threshold** (dB)
0	Out of range		
1	QPSK, 78	0.1523	-9.478
2	QPSK, 120	0.2344	-6.658
3	QPSK, 193	0.3770	-4.098
4	QPSK, 308	0.6010	-1.798
5	QPSK, 449	0.8770	0.399
6	QPSK, 602	1.1758	2.424
7	16QAM, 378	1.4766	4.489
8	16QAM, 490	1.9141	6.367
9	16QAM, 616	2.4063	8.456
10	64QAM, 466	2.7305	10.266
11	64QAM, 567	3.3223	12.218
12	64QAM, 666	3.9023	14.122
13	64QAM, 772	4.5234	15.849
14	64QAM, 873	5.1152	17.786
15	64QAM, 948	5.5547	19.809

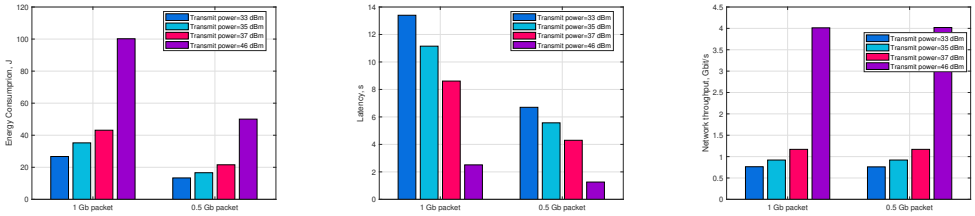
by creating configurations according to the channel conditions of UEs. In particular, in the case of $D2D_{thr}$ equal to 30 m, 40 m, and 45 m, uniformly distributed UEs are too far from each other but, along with it, cannot satisfy SNR requirements of the multicast group. In such a scenario, *bad* UEs have to be deleted from a multicast group and served via unicast links that may deteriorate the performance due to the sequential transmissions. Recall that we consider a single beam system and that higher CQI puts more strict requirements for multicast group channels. Hence, more sidelink communications can be established out of multicast ones. This general trend, while varying the SNR threshold, can be tracked only for $D2D_{thr}$=50 m, where UEs can freely establish sidelink connections.

In Fig. 3.9, we further investigate the performance of the proposed Heuristic compared to MGF and both SA3M algorithms. In particular, we show the mismatch between the energy consumption required by the Heuristic with respect to the analyzed solutions. It can be noticed that the SNR threshold that corresponds to CQI 8 provides the best Heuristic performance. The reason is that in this case, more multicast groups will be unclustered, and sidelink transmissions will be in priority, whereas chosen $D2D_{thr}$=50 m allows UEs to launch sidelink transmissions. Furthermore, regarding the performance vs. complexity trade-off, it is essential to highlight that Heuristic has lower complexity (by orders of magnitude) compared to SA3M while demonstrating comparable performance, given proper setting adjustments.

Effect of Transmit Power

As a final step, we investigate the impact of transmission power on energy consumption, latency, and network throughput. Obviously, by lowering the transmit power, a decrease in energy consumption is achieved. The same decreasing trend is experimented in terms of

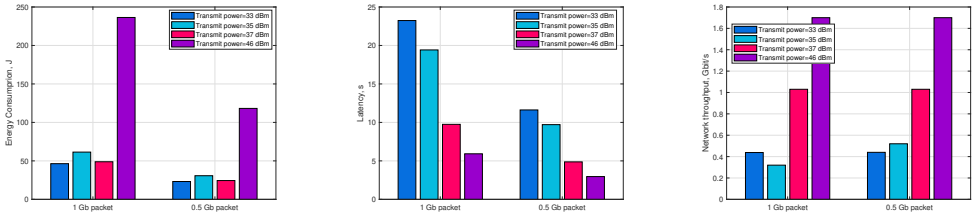
SNR, leading to a throughput degradation which, in turn, causes a delay increase. The raised delay affects energy consumption. Therefore, there is a trade-off between transmit power, P_T, and delay. Moreover, packet size and available bandwidth at a transmission link also impact energy consumption.



(a) (b) (c)

Fig. 3.10. Effect of transmit power on (a) energy consumption, (b) latency, and (c) network throughput. Heuristic with $\mathrm{D2D_{thr}}$=50 m, MCS 8, and $W = 1$ GHz.

To this aim, we vary P_T as shown in Fig. 3.10. Observe that the rise in transmit power increases energy consumption up to 73% comparing the two extreme cases of 46 dBm and 33 dBm. Conversely, P_T reduction results in higher latency, as demonstrated in Fig. 3.10(b). Depending on the service requirements and hardware on the devices (battery life is important), the choice of the transmit power can be shifted to one of the extreme cases or, differently, to a middle value. By analyzing Fig. 3.10 further, one can notice that packet size does not influence the trend of the curves for $W = 1$ GHz since the available bandwidth in the system allows data delivery. Recall that bandwidth represents the highest reliable transmission rate and is necessary for understanding the amount of traffic a connection can support.



(a) (b) (c)

Fig. 3.11. Effect of transmit power on (a) energy consumption, (b) latency, and (c) network throughput. Heuristic with $\mathrm{D2D_{thr}}$=50 m, MCS 8, and $W = 400$ MHz.

To further analyze the energy consumption/latency trade-off, we now discuss results in Fig. 3.11 in which the transmission bandwidth is set to $W = 400$ MHz. We can see that the latency increases under decreasing transmission power values, independently of the bandwidth and the packet size. However, a different trend from Fig. 3.10 can be noticed in

the case of energy consumption and network throughput. Here, we can observe a nonlinear trend under increasing transmit power for both considered packet sizes (see $P_T = 35\,\text{dBm}$ compared to $P_T = 33\,\text{dBm}$ and $P_T = 37\,\text{dBm}$), revealing the existence of the trade-off between the energy consumption and energy efficiency (or, equally, latency).

As a result, one can deduce that bandwidth together with the transmit power can be adjusted by means of scalable numerology to reduce the total power consumption in the network.

3.2 5G NR Sidelink Multi-Hoping in Public Safety and Factory Automation Scenarios

3.2.1 Motivation

3GPP sidelink transmissions in LTE-A systems have already proven to play a crucial role in supporting public safety and Vehicle-to-Everything (V2X) services, among others, by featuring direct communications between two user devices without any BS involvement [96]. Fostered by the successful evolution in LTE-A, the 3GPP sidelink developments are going on in NR based systems, wherein sidelink transmissions become an essential component complementing the Uu communication between UE and BS. 5G wireless communication systems utilize NR sidelink for D2D based proximity service (ProSe) communications [97] that can operate in both lower (up to 7.125 GHz) and higher (up to 52.6 GHz) frequency ranges. Focusing on providing low-latency, high-reliability, and high-throughput services, NR supports a number of new sidelink communication features not provided in LTE-A. These include, among others, support for unicast and groupcast in the radio layers (Long Term Evolution (LTE) only supported broadcast) and Hybrid Automatic Repeat Request (HARQ) operation at the MAC level.

On the application side, we observe that public safety organizations have already begun to shift from traditional land mobile radio to cellular communications systems, leveraging a new set of deployed devices to meet mission-critical requirements and target new public-safety broadband applications. Accordingly, 3GPP Rel-16 targets defining the common architecture for public safety and commercial ProSe services [98]. In the case of public safety, maintaining ProSe discovery and communication is especially critical when the UE resides outside the coverage area of the cellular network, e.g., in the case of disaster management in remote areas. Hence, the support for both direct discovery (discovery is integrated into the initial sidelink connection establishment message) and unicast and groupcast communication (one-to-one and one-to-many communication) was introduced. Moreover, out-of-network-coverage discovery is already feasible in Rel-16.

Note that *public safety* service reliability can be achieved by using either multi-hop[1] D2D communications (e.g., in out-of-coverage scenarios) or through the flexible use of radio resources provided by the multi-connectivity and multi-radio access technologies (multi-RAT). Technologies such as Mobile Edge Computing (MEC) and Software-Defined Networks (SDN) can improve latency and security in public safety services [99]. Also, Network Function Virtualization (NFV) and network slicing can manage various use cases with varying priorities in cellular networks [100]. In this study, we investigate the main *advantages and disadvantages of D2D ProSe transmissions.*

ProSe support can be also beneficial to commercial use cases and services. In the realm of *factory automation*, for example, it can provide new possibilities for discrete manufacturing and help producers accomplish efficient operations. Nowadays, as mentioned above, factory automation is based mainly on wired connectivity, which bounds the degree of freedom for functionalities, especially for mobile terminals. Hence, robust wireless connectivity can improve the location flexibility of a large number of machines, such as sensors, actuators, and programmable logic micro-controllers. Furthermore, as factory automation use cases usually (but not always) belong to the class of URLLC and the existing technologies operating over the unlicensed spectrum are not capable of guaranteeing the required Quality of Service (QoS) in the considered scenario, NR sidelink has the potential to offer interesting opportunities.

Note that the NR sidelink can be easily deployed on the licensed spectrum. Suppose the mobile network operators spectrum is to be used for the Industrial Internet of Things (IIoT). In that case, the NR sidelink may have an advantage over other D2D technologies (because the operator presumably uses NR cellular). Otherwise, in case the spectrum is unlicensed for industrial use, other D2D technologies can be utilized. However, in terms of power consumption, devices should only maintain the cellular interface active to save power [101].

In summary, as proven by several research works focused on direct links between UEs, deploying D2D communications in cellular networks benefits from proximity and spatial reuse gains. However, two main NR sidelink aspects **have not been sufficiently investigated and standardized yet**: (i) *multi-hop transmission (relaying)* and (ii) mobility. In this article, we focus on the first feature – multi-hop transmission – by considering a static scenario. More precisely, while other existing studies only investigate the special case of one relay node (two-hop) sidelink operation, we investigate the case of an arbitrary number of hops. Thus, our article aims to partially fill the mentioned research gap by elaborating on the concept of D2D ProSe communications while referring to public safety and factory automation sample use cases. We first discuss the pros of NR sidelink, including the comparison with LTE ProSe communications. We then review the NR sidelink applicability for public safety and factory

[1] Note that the NR sidelink (Rel-16 and Rel-17) does not support multi-hop (i.e., UE1-UE2-UE3) at radio layers.

automation applications and conduct a preliminary simulation study. Finally, we offer future directions for the NR sidelink development from a standardization perspective.

3.2.2 NR Sidelink in a Nutshell: Why NR Sidelink?

LTE sidelink (or D2D) was introduced for the first time as a part of 3GPP Release 12, aiming at covering public safety scenarios and supporting two operation modes. In mode 1, eNB[2] assists UEs and allocates dedicated transmission resources, whereas, in mode 2, UEs randomly select the radio resources from the pool that was previously sent by eNB. Both modes have the same pool of resources, wherein the transmission is scheduled during the so-called Physical Sidelink Control Channel (PSCCH) [102]. Later, in LTE sidelink Release 14, 3GPP added several enhancements to the Mission-Critical Push-to-Talk (MCPTT) standard and upgraded the functionalities of public safety applications by introducing Mission-Critical Data (MCData) and Mission-Critical Video (MCVideo) [103].

At the radio level, in terms of backward compatibility the following aspects define new and old specifications: LTE sidelink Rel-13 is compatible with LTE sidelink Rel-12; LTE sidelink Rel-14 is not compatible with earlier LTE sidelink; LTE sidelink Rel-15 is compatible with LTE sidelink Rel-14; NR sidelink is not compatible with any LTE sidelink; NR sidelink Rel-17 will be compatible with NR sidelink Rel-16. Here, a new model was introduced, where each set of services is mapped onto a single release of the specification (e.g., safety services are mapped onto Rel-14, whereas advanced driving services are mapped onto Rel-15). That is, each release is aimed at supporting a certain set of services [104].

NR sidelink (Rel-16) operates more efficiently and is designed so as to utilize both licensed and unlicensed frequency bands. More specifically, both LTE and NR sidelink support communications in the licensed spectrum as well as in the unlicensed ITS spectrum (essentially the 5.9 GHz band). However, neither LTE nor NR sidelink support communications in different unlicensed spectra such as the 2.4 and 5 GHz bands. In view of this, various NR protocols facilitate the coordination and control of the sidelink transmissions within the network coverage, which ensures that the D2D communications effectively coexist with cellular data traffic in shared frequency bands.

The direct mode interface (PC5 or sidelink), which complements the cellular interface by introducing new flexibility to the NR technology, has been presented in Rel-16 [105]. PC5 or sidelink operates in in-, out-of-, and partial-coverage scenarios[3], leveraging NR frequency bands and supporting unicast, multicast, and broadcast communication, where members interact via groupcast transmissions. This option is useful in the transmitter-receiver close proximity scenarios and in the intermittent network coverage ones. Release 16 sidelink transmissions solely involve V2X scenarios though, the 3GPP is planning further sidelink-related

[2] E-UTRAN Node B, also known as Evolved Node B (abbreviated as eNodeB or eNB), is an LTE BS.

[3] The LTE sidelink could also operate in-coverage, partial-coverage and out-of-coverage.

features in Rel-17 that are expected to play a decisive role in expanding the applicability of 5G NR to a wide variety of new use cases in both industry and public services, such as public safety, factory automation, enhanced V2X, advanced relay, and XR interactive games, among others.

However, energy efficiency - a crucial feature for pedestrian/drone UE in terrestrial/aerial V2X, wearable UEs in interactive games, or mobile UEs in public safety - is not the primary concern in the Rel-16 sidelink transmission design. In this regard, a high degree of energy efficiency at both the network and device sides must be ensured. In Release 16, the blind decoding of the PSCCH appears to be one of the significant causes of energy consumption in Modes 1 and 2. The transmission and reception procedures of PSCCH and Physical Sidelink Shared Channel (PSSCH) may be further advanced to save power at the UE side. Thus, within NR sidelink Rel-17, a work item on sidelink enhancements is targeting energy efficiency improvements.

To summarize, four new features are introduced in NR sidelink to meet the service requirements of the use cases that demand high reliability, low-latency, high-throughput transmissions, and high connection density. First, point-to-multi-point and point-to-point transmissions are supported in addition to broadcasting. Second, ultra-reliable and low-latency NR uplink communications are achieved thanks to grant-free transmission, a promising multiple access protocol. Finally, the channel sensing and resource allocation procedures are improved to facilitate collision mitigation among different sidelink transmissions initiated by various UEs.

LTE, LTE sidelink, and NR transmissions can also be used for public safety and factory automation scenarios. However, these are still less efficient in supporting mission-critical services. In the case of LTE, even though there are solutions, such as portable eNB on trucks, to address disaster and emergency situations, most of the time, the coverage and robustness of such solutions are somewhat limited and may not guarantee the requirements requested by modern applications. LTE sidelink has a public safety focus, but its main drawback is that it operates only in broadcast mode and only in unlicensed spectrum, thus meaning that there is no support for public safety bands in LTE. Moreover, LTE sidelink has very little support from chipset vendors, implying that there are no real UEs on the field so far. When touching NR, the initial focus of NR was Enhanced Mobile Broadband (eMBB) (Rel-15), then IoT and URLLC (Rel-16 and Rel-17). This means that there is slight support for public safety features, but the technology is not mature yet since more work is expected to be done in the following Rel-18 and Rel-19.

3.2.3 NR Sidelink as a Tool to Support Public Safety and Factory Automation Use Cases

Several new use cases are still expected to be supported in Rel-17. These use cases are related to V2X and public safety. Then, the Rel-17 NR sidelink can also be used for industrial

communication, such as sidelink between robots, machines, and industrial sensors (even though IIoT is out of the scope of Rel-17). To fully cover new and already existing use cases, we consider public safety and factory automation applications that differ in requirements.

Public Safety

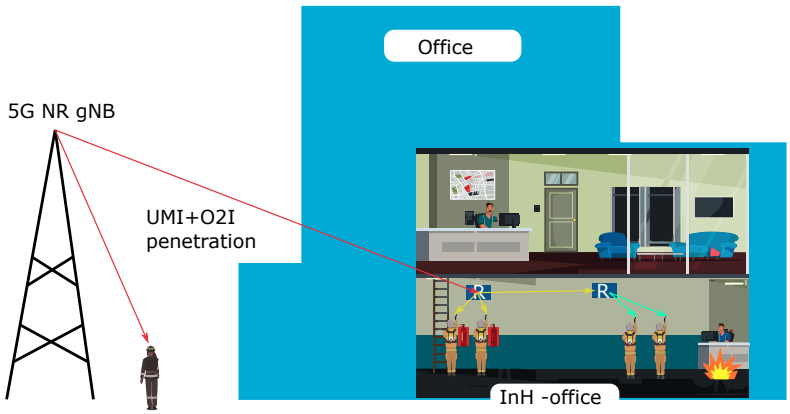


Fig. 3.12. Public safety use case illustration. "R" stand for a relay. "R" can transmit data via unicast and groupcast. In this work, we use only unicast. Note that broadcast is only used for LTE.

Public safety organizations are responsible for providing services that ensure the people's and properties' safety thanks to first responders, such as firefighters, emergency medical service staff, etc., equipped with devices exchanging time-sensitive and critical information via typically wireless communication links. To support the mission-critical requirements of public safety services, these organizations have begun to move from traditional land mobile radio to cellular communications systems with a new set of terminals.

Reliability in public safety services can be achieved via multi-hop relaying, which is considered to be one of the key technologies facilitating enhanced system performance in future 5G+ systems. For example, it allows establishing direct connections between devices in scenarios outside the coverage area, thus ensuring first responders with the connectivity they need, especially in hazardous situations. For instance, in [106], the fire brigade use case is already under consideration to enhance indoor coverage.

Furthermore, public safety use case introduces potential new requirements [106], such as, among others, the following ones: *(i)* the 5G system shall support the relaying of MCPTT, MCVideo, and MCData services between remote UEs and a network using multi-hop relay UEs; *(ii)* the 5G system shall support service continuity when a remote UE moves into an area within the coverage of a different multi-hop relay UE; *(iii)* the 5G system shall allow the user to decide when to deploy additional multi-hop relay UEs to maintain a reliable communication path, etc. We emphasize that new public safety services with their mission-

critical requirements call for cellular communication systems that support a D2D ProSe (see, e.g., Fig. 3.12).

Factory Automation

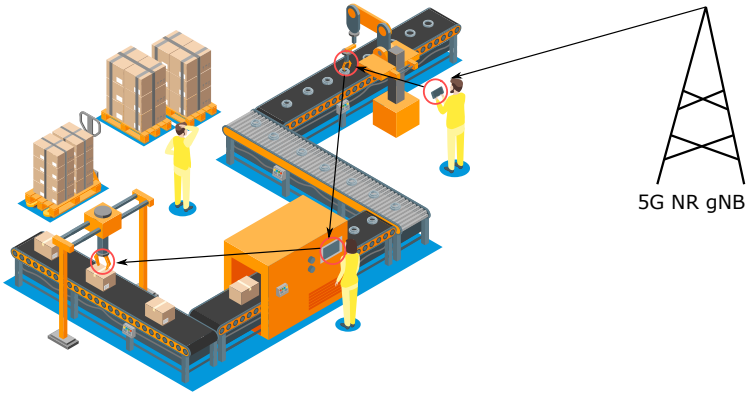


Fig. 3.13. Factory automation use case illustration.

The scenario that is expected to be investigated in Rel-18 is multi-hop sidelink for *factory automation scenarios* (see, e.g., Fig. 3.13). Here, complementary to deploying multiple gNBs[4], the factory owner may leverage UEs with multi-hop relay capabilities to relay messages between, e.g., remote UEs and gNBs [106].

IIoT scenarios, such as factory automation use cases, usually demand URLLC and include communication between automation devices, such as industrial robots, controllers, and sensors. Even though different technologies are designed to support industrial communication, e.g., IWLAN, WISA, WirelessHART, these standards do not satisfy the flexibility and real-time-response requirements of control loops. Moreover, 3GPP defined the target packet error rate at $10^{-}6$ (known as "six nines" reliability) [107].

To this end, factory automation services can benefit from D2D ProSe since direct communication between industrial terminals helps reduce communication latency. Further, it results in reduced resource usage compared to the legacy centralized traffic stream through the BS or, equally, gNB.

Performance Indicators

In factory automation and public safety scenarios, strict requirements[5] on the following concepts have to be considered:

[4] Next Generation Node Bases (gNBs) is the NR term for a BS).
[5] Note that not all requirements have to be satisfied at the same time.

- **Latency**. NR sidelink mitigates the end-to-end latency determined as the time taken for devices to respond to each other over the wireless network and is crucial for remote controlling, URLLC applications, etc.
- **Power consumption**. The terminals that operate in D2D fashion experience a reduced energy consumption primarily due to the mutual proximity and reduced latency, since the total energy consumption of transmissions in the network is calculated by multiplying power in watts by time.
- **Service continuity**. When a link is broken, the terminal should be able to select another link as soon as possible not to lose too many packets, which can be achieved, e.g., through multi-hop D2D communication.
- **Reliability**. D2D technology can improve reliability, expressed as the fraction of sent network layer data units that are successfully delivered to a given node within the time constraint required by the targeted service.
- **Service availability**. A further metric that can benefit from the deployment of multi-hop D2D (among other technologies, such as MEC, multi-RAT, Integrated Access and Backhaul (IAB)) is the communication service availability. It is defined as the time interval during which the end-to-end communication service is delivered in compliance with an agreed QoS, divided by the total time interval the system is expected to provide the end-to-end service in a given area [108].
- **Energy efficiency**. Efficient energy use aims at a reduction in the consumed energy while providing the service; it is defined as the achieved network throughput divided by the consumed energy.
- **Network throughput**. Network (or aggregate) throughput is the total data transfer rate delivered to all devices in the network.

A more detailed description of public safety or factory automation-related requirements from the communication and other perspectives is available in 3GPP specifications [106, 108]. For example, 30 ms end-to-end latency (all hops are included) should be ensured for public safety scenarios from the communication vision, whereas an example of functional requirements associated with public safety can be that the maximum operating range of a UE-to-network relay UE in an indoor scenario with unobstructed view shall exceed 50 meters. For industrial settings, end-to-end latency in the range between 5 ms and 1 s might be enough, while the 5G system shall support the relaying of UE traffic through multi-hop relay UEs in a manner that minimizes the impact of the relaying on system performance. Otherwise, we refer readers to Table 6.7.4-1 for public safety and Table 8.3-1 for detailing the factory automation use case [106, 108].

3.2.4 Performance Assessment

To analyze the gains and downsides deriving from NR sidelink multi-hop relaying in the use cases described above, we developed a MATLAB simulation environment (considering trans-

Table 3.6. Simulation parameters factory automation scenario

Scenarios	**Factory Automation**	**Public Safety**
Area	100 m x 70 m	500 m x 1000 m
NR carrier frequency	28 GHz (FR 2)	700 MHz (FR 1)
LTE carrier frequency	2.1 GHz	2.1 GHz
Total NR bandwidth	100 MHz	100 MHz
Total LTE bandwidth	100 MHz	100 MHz
Height of BS	3 m [110]	10 m (UMi) [110]
NR subcarrier spacing	120 kHz	60 kHz
LTE subcarrier spacing	15 kHz	15 kHz
NR transmitter processing delay	0.0357 ms	0.0179 ms
LTE transmitter processing delay	1 ms	1 ms
NR transmitter processing delay	0.0357 ms	0.017 ms
LTE transmitter processing delay	1 ms	1 ms
NR frame alignment time	0.0179 ms	0.0089 ms
LTE frame alignment time	0.5 ms	0.5 ms
NR transmission time	0.0357 ms	0.0179 ms
LTE transmission time	1 ms	1 ms
NR receiver processing delay	0.0536 ms	0.0268 ms
LTE receiver processing delay	1.5 ms	1.5 ms
NR one way latency	0.1429 ms	0.0715 ms
LTE one way latency	4 ms	4 ms
NR HARD RTT	0.2143 ms	0.1074 ms
LTE HARD RTT	8 ms	8 ms
Height of UE	1.5 m [110]	1.5 m [110]
Number of BSs	2 BSs	1 BS
Number of UEs	10 UEs	6 UEs
SNR threshold	-9.478 dB	-9.478 dB
Transmit power	20 dBm [111] for both BS and UE	46 dBm [BS]/ 23 dBm [UE] [112]
Fading margin	4 dB	4 dB
Interference margin	3 dB	3 dB
Path loss model	Heavy industry [111]	UMI+O2I penetration loss (low-loss model) [110] / InH - office
Antenna array	32x4 URA	16x4 URA
Packet size	10-300 byte	10-300 byte

mission part only) dimensioned according to the parameters' values listed in Table 3.6. We emphasize that the NR sidelink frequency of operation can be FR1 that contains frequencies from 410 MHz to 7.125 GHz and FR2 (mmWave) that covers the range between 24.25 GHz and 52.6 GHz [109]. This subsection compares the performance of the NR and LTE sidelink-enabled systems with NR and LTE. Note that LTE and NR benchmarks exploit sequential BS-UE transmissions scheduled one by one, whereas multi-hop transmissions are used only for NR sidelink use cases.

Factory Automation Scenario

For the factory automation use case, we use 5G FR2 that has been allocated to 5G in the mmWave region. We deploy a private network with two BSs within the factory that is

separated from the global network. Nodes (10 UEs) are uniformly distributed in an area of 100 m x 70 m. From the application point of view, this scenario corresponds to URLLC IIoT. In our performance evaluation, we use a bandwidth of 20 MHz and 100 MHz per single LTE and NR carrier, respectively. Then, to fairly compare the performance of NR and LTE, we use LTE carrier aggregation to 100 MHz. The path loss model is adopted from the heavy industry [111], and the received rate is computed using the Shannon Theorem.

We consider the following traffic model: the number of users in the cell (or area of interest) is constant, and each user is assigned a finite payload to receive. When a UE receives the packet, it can transmit the data to the other UEs in the network by establishing multi-hop communication. We assume that multi-hop communication can be established as follows:

- **Case 1:** multi-hop communication is established as a chain of sequential unicast transmissions (see Fig. 3.14(a)). For instance, the following chain of transmissions can be determined: BS/AP → relay (1) → relay (2) → relay (3) → relay (4) → relay (5), etc. Here, each next relay is selected based on the best channel quality (e.g., SNR, RSS) between the last relay and the devices that have not received the data. We note that thanks to the better channel conditions between hops, the latency can be reduced. However, relaying may also add a delay (compared to Multimedia Broadcast Multicast Services (MBMS)) due to the hops when one considers HARQ, etc.
- **Case 2:** multi-hop communications with concurrent unicast transmissions can be established. For example, both pink links marked as (3) in Fig. 3.14(b), are concurrent links, and transmissions (2) and (3) coming from the first relay are performed one after the

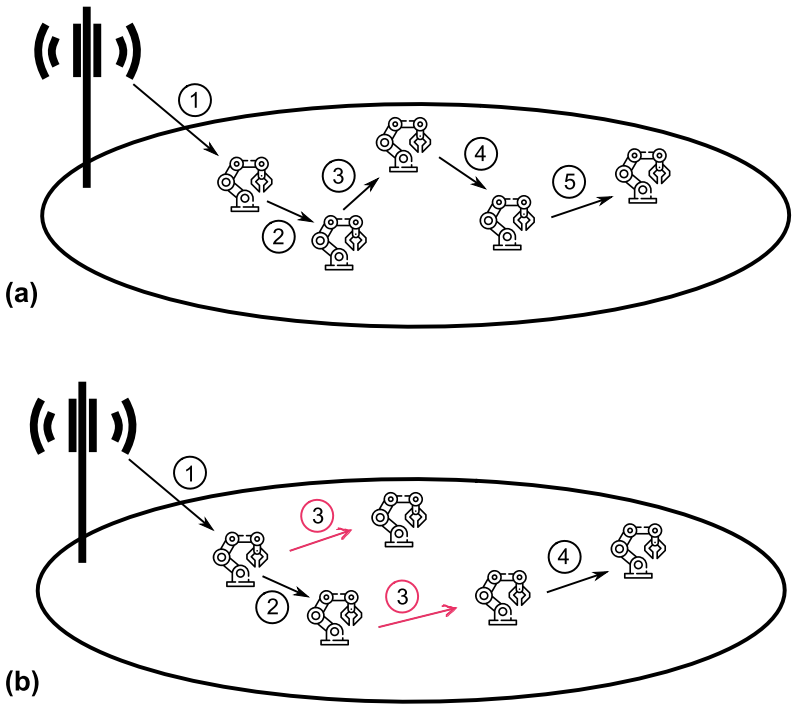


Fig. 3.14. Multi-hop establishment illustration: (a) chain, (b) concurrent transmissions.

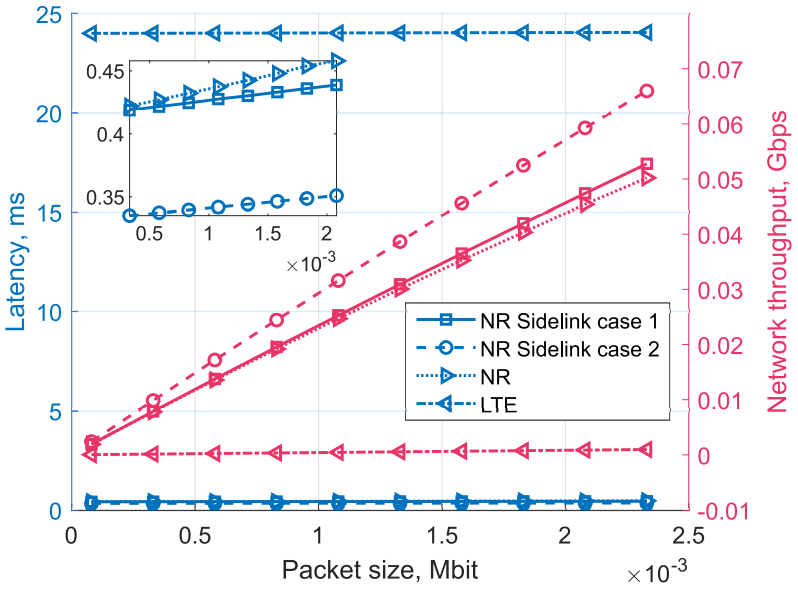


Fig. 3.15. Factory automation: latency as a function of packet size (blue); network throughput as a function of packet size (red).

other. Here, similarly to case 1, each next relay is selected based on the best channel quality, but at the same time, several relays (that already received packet) can forward data to those who are still waiting to be served.

For the *factory automation scenario*, we consider two main metrics that are critical for the use case: *latency* and *network (aggregated) throughput* (see Fig. 3.15). Note that latency is calculated as the time required for the data packet dissemination to all the terminals in the network, whereas throughput can be described as the total data rate delivered to all UEs and is calculated as the product of the packet size times the number UEs UEs divided by latency.

One may notice that in our simulation settings, NR sidelink improves the system parameters compared to other systems in terms of end-to-end transmission delay and network throughput thanks to the gains obtained from reuse, relay, and proximity. We recall that latency in the case of the URLLC scenario has to be minimized, whereas network throughput (or, equally, the sum of data rates) benefits from the reduced end-to-end latency. Recall that the sum of data rates varies with the channel variations. Then, an increase in packet size may lead to a rise in latency to some extent. Note that the superior behavior of the NR sidelink comes from more flexible scheduling and scalable numerology. Also, sidelink can provide better energy efficiency and latency since transmissions generally occur over short distances, leading to the fact that the modulation and coding scheme selected during a sidelink transmission is generally high.

Most importantly, one may observe that the relay selection mechanism plays an important role in improving the system performance. More precisely, the possibility of utilizing concurrent transmissions and the sophisticated selection of the next-hop relay according to

the channel quality between devices (case 2) reveals the best performance. Hence, we can deduce that designing algorithms that are able to take fast and intelligent decisions on relay discovery and selection is of particular interest.

Further, we emphasize that cooperation with several relays may introduce sufficient macro-diversity and system reliability in conditions of a high probability of LoS paths being blocked. Hence, without sidelink features (or IAB, among other technologies), the system would experience difficulties guaranteeing reliable communication in factory automation setups. Then, this problem worsens for the dynamic scenario.

Public Safety: Fire Brigade Scenario

To investigate the impact of NR sidelink on system performance considering the public safety use case, we consider uniformly distributed points with coordinates (x, y) within an area of 500 m x 1000 m. Then, we deploy uniformly 6 UEs in a radius of 20 m with (x, y) as a center. The BS is located at the center of 500 m x 1000 m rectangular. The Head of the team is assumed to be out of the building and is controlling the rescue operation. One of the UEs is the first relay device from outdoor to indoor environments (the one with the best channel). We use the UMi 3GPP path loss model for the link between the relay device and the BS as well as consider Outdoor-to-Indoor (O2I) penetration loss. The path loss incorporating O2I building penetration loss is modeled as described in [110]. Then, for indoor multi-hop relaying, we use the indoor 3GPP model (InH - office).

Regarding the way of forwarding the data to each device from the relay, there are two possible approaches: *(i)* relay UE forwards the data to each device independently, and *(ii)* relay UE forwards the data to one device, and then this device forward the data to another device and so on (i.e., a chain of transmissions). In our simulation, we exploit the second option, wherein gNB is connected to the first relay device, and the relay device forwards this information to the rescue team (according to the multi-hop case 2). To simulate antenna arrays, we use MATLAB Antenna Toolbox. We consider FR1 envisaged to carry much of the traditional cellular mobile communications traffic.

Differently from the previous use case, public safety focuses on wider coverage and power saving for battery-based UEs. As one may learn, Fig. 3.16, presents the results for the *fire brigade scenario* in terms of *energy efficiency* and *power consumption* as a function of the packet size. We estimate the total power consumption in the system by multiplying transmit power in watts by the time required for the packet delivery, whereas efficient energy use is defined as a division of network throughput by the consumed energy. That is, energy efficiency is defined as the obtained network throughput divided by the used energy in bit/s/J, which assesses how effectively energy is utilized to get the network throughput.

We highlight that both transmit power and transmission delay impact on energy use. Then, the energy consumption can be reduced by lowering the transmit power. Hence, in our setup, the transmit power from the gNB and between relays are set to 46 dBm and

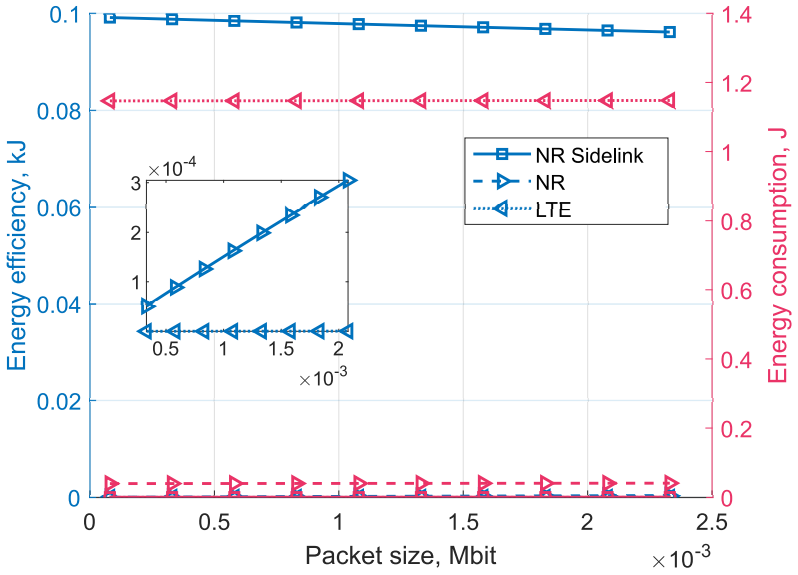


Fig. 3.16. Public safety (packet size ranges from 10 to 300 byte): energy efficiency as a function of packet size (blue); power consumption as a function of packet size (red).

23 dBm, respectively. However, there is a trade-off between lower transmit power (less energy consumption) and delay (which also causes energy consumption).

From the obtained results, one may infer that NR sidelink brings its advantage in terms of proximity and relaying, affecting propagation properties as we deal with mixed outdoor-indoor environments. NR ProSe allows for high reliability, high bit rates, low power consumption, and low latencies. Note that reliability in public safety services can be satisfied by using different tools, such as multi-hop relaying, multi-connectivity, multi-RAT, etc. Relaying has many advantages, including the possibility to ensure extended coverage and reliability in the case of network failure. However, several challenges need to be addressed to provide a robust solution. For example, the security of D2D communications has to be guaranteed. Then, the question is when to do relaying, and on what parameters the source node decides to relay via nearby nodes. Finally, latency constraints: relaying may also add a delay (compared to MBMS) due to the hops.

3.3 Conclusions

This chapter introduced the concept of D2D communication. More precisely, we developed a two-step framework for optimal sidelink-assisted multiple modes mmWave scheduling. Its complexity is reduced by exploiting an existing unsupervised learning algorithm to form multicast clusters. The resulting proposed solution leverages both optimization and machine learning techniques to deal with different types of users' mobility, user distribution, and network-side parameters, such as transmit power and bandwidth. To face complexity issues, a heuristic, which tracks channel conditions of multicast users, is also designed.

A thorough analysis of the system behavior has revealed crucial quantitative trade-offs to handle with. Specifically, we elaborated on the complexity/performance trade-off connected with users' mobility. In particular, we recommend rerunning the multicast group formation algorithm, which allows for achieving better performance in mobile scenarios at the expense of low computational complexity. We then evaluated the energy consumption reduction as a consequence of a decreased transmit power and its impact on the total network latency. We emphasize that in 5G NR, the network's overall power consumption can be reduced by adjusting both bandwidth and transmit power. By combining the achieved results, we may conclude that *multicast and D2D technologies are powerful tools to improve the performance of mmWave directional systems in the presence of dynamic users.*

Further, we provided an overview of the main functionalities and features of NR sidelink compared to the LTE-A. NR sidelink is believed to become an essential technology to ensure both mission-critical and ultra-reliable low-latency communications. Then, we elaborated on NR sidelink as a solution for public safety and factory automation, outlining their principal requirements and use case flow. Throughout the simulation study, we raised the possibility of using NR sidelink communication for public safety and factory automation scenarios, demonstrating noticeable end-to-end latency and energy efficiency performance improvement compared to LTE, LTE sidelink, and NR transmissions.

Materials of this chapter were published in [113].

4

Cooperative Positioning

This chapter addresses the problem of cooperative localization in wireless systems. In particular, we attempt to provide a holistic overview of the collaborative localization methods using D2D technology and integrating the transmission and localization components of wireless systems. Further, by relying upon the path loss models, we offer a theoretical model based on RSS ranging for D2D- and RIS-aided cooperative positioning to provide a comparison between them in terms of positioning accuracy. We also analyze the performance of joint localization integrating cellular, D2D, and RIS cooperative systems implemented by using statistical tools.

4.1 D2D-based Cooperative Positioning Paradigm for Future Wireless Systems

4.1.1 Motivation

A wide variety of emerging 5G and beyond 5G (B5G) applications demand high positioning accuracy, which, in turn, affects the performance of various location-based applications. The recent advances in the Global Navigation Satellite System (GNSS) area, such as Precise Point Positioning (PPP) or Real-Time Kinematic (RTK), are enabling positioning accuracies below 10 cm outdoors. Despite positioning is more or less solved outdoors, positioning accuracy still requires an improvement indoors. For instance, for indoor localization or positioning in dense urban settings, Wireless Local Area Network (WLAN) fingerprinting techniques can provide the accuracy of 3 m to 4 m, while to achieve this result, one needs to maintain a large fingerprint database. Meanwhile, next-generation high accuracy positioning will require the accuracy of less than 1 m in more than 95% of the service area, including urban, outdoor, and indoor environments [114, 115]. It is already shown in [116, 117] that 5G NR technology can facilitate high-accuracy continuous tracking and positioning and confirmed that it is possible to achieve sub-meter localization accuracy over 99% of the time with 3GPP-specified 5G NR parametrization.

Key prospects of 5G networks, such as increased bandwidth, small cells and high MT density, D2D communication capability, and multiple radio access technologies, are favorable for localization [118]. Specifically, 5G systems envisage scenarios, where MTs may cooperate for achieving accurate positioning by directly exchanging necessary data through the D2D

links. More technically, MTs transmit their physical layer estimates, thereby accelerating local decisions. Then, higher layer information (e.g., position estimates between terminals or mobility information about the neighboring nodes) is exchanged between MTs. This approach, for example, can be applied to use cases where the positioning accuracy is around a few meters to lessen the accumulative positioning error via cooperative positioning and, hence, improve the accuracy [119, 120].

Notably, in 5G delay-sensitive networks, a distributed D2D cooperative localization helps in reducing the delay to discover nearby nodes' locations, which is crucial, e.g., to perform specific intelligent control, such as self-driving [118]. A further feature paving the way for cooperative positioning applications in 5G networks is the availability of high-density small cells, favoring D2D communication between MTs. For example, in view of future V2X scenarios, D2D communications are already considered in 5G NR Release 16 [4] as a valued means to enable ultra-dense cooperative localization. Due to this, D2D-aided cooperative positioning is expected to achieve the ubiquitous positioning of below one-meter accuracy [121], thereby fulfilling the 5G requirements [114]. Motivated by the aforementioned emerging interest in the subject, this section focuses on integrated D2D positioning systems.

Differently from the previous studies on D2D and/or 5G technologies that fall into communication or multimedia sectors, this section is devoted to positioning aspects. In this section, we attempt to provide a holistic overview of the collaborative localization methods using D2D technology and the integration of the transmission and localization components of wireless systems. The section investigates the state-of-the-art on D2D communication capabilities used for localization, thus acting as a glue between these two promising technologies for satisfying the ever-growing requirements of 5G and B5G networks. Summarizing, in this section of the thesis, we concentrate on answering the following questions: *(i) can D2D technology help positioning by increasing its accuracy?; (ii) can the cellular network provide effective support to positioning?; and (iii) what are pros and cons of D2D-based cooperative localization?*

The section's key takeaway points are: *(i) there are two ways to exploit D2D for cooperative localization. (ii) cooperative D2D positioning provides increased accuracy and extends coverage compared to non-cooperative methods. (iii) Cooperative D2D positioning allows for relative position estimation, even in the absence of reference stations. (iv) 5G mobile networks have enhanced positioning capabilities in comparison with previous generations. And (v) cooperative D2D-aided cellular positioning can facilitate sub-meter localization accuracy demanded by 5G applications.*

4.1.2 5G Positioning Applications and Enabling Technologies

As specified in [5, 115], 5G positioning services intend to support applications and verticals, such as Intelligent Transportation Systems (ITSs), eHealth, Industry 4.0, with sub-meter positioning accuracy. Thus, many 5G verticals and applications (including regulatory needs

services) pose ambitious system requirements for positioning accuracy (see Table 4.1). For instance, in Location-based Services (LBS) and ITS scenarios, high accuracy is critical to new services and applications for both outdoors and indoors.

Regarding the Industry 4.0 use case, it is critical to locate assets and other moving objects (e.g., forklifts) on the factory floor. Similarly, in the logistics and transport industry, for example, in the use of drones, rail, and road transport, the need for object location determination exists. Further, in use cases with guided vehicles, Unmanned Aerial Vehicle (UAV), and objects involved in safety-related functions, also a high resilience in position availability is a key issue [5].

A further category, which demands high precision positioning, is mission-critical services. For instance, first responders may be located at all times during usual and critical operations, indoors as well as outdoors. Here, the level of positioning accuracy (and other Key Performance Indicator (KPI)[1] such as horizontal accuracy, vertical accuracy, positioning service availability, heading [122], latency for UE position estimation, corresponding positioning service level, UE speed, update rate, and Time to First Fix (TTFF)) required is much more stringent than that demanded by local and regional regulatory requirements for commercial 5G users [5]. We summarize the main requirements for vertical use cases in Table 4.1, partial information of which is extracted from [4] (Section 6) and [5] (Section 9, Annex B).

Moreover, in particular applications and services, the network operator has to provide a customized localization service for various users requiring different performance levels. Hence, the support of several localization services is considered as a separate use case, which can be managed by relying on multiple technologies, for instance, 3GPP and non-3GPP technologies, as well as a combination of both 3GPP with non-3GPP positioning technologies. It is important to mention that different localization methods are capable of offering different accuracy levels [123].

Non-3GPP Technologies

The exploitation of existing standards designed only for communications: currently, numerous wireless communication technology standards are available for WLAN, Wireless Sensor Network (WSN), and IoT applications. Examples can be Wireless Fidelity (Wi-Fi), ZigBee, Radio Frequency Identification (RFID), and Bluetooth Low Energy (BLE). These technologies do not provide any particular positioning capabilities. Instead, their transmitted signals are used and mixed to offer different levels of localization accuracy.

While RFID and BLE are typically used with proximity methods due to their limited transmission range, Wi-Fi technology has been successfully adopted in several positioning systems, usually leveraging fingerprinting techniques where accuracies at meter-level can be

[1] For example, TS 22.104 Clause 5.7 provides positioning requirements for horizontal and vertical accuracy, availability, heading, latency, and UE speed in an industrial use case scenario.

Table 4.1. Vertical Use Cases for Localization and Requirements [4, 5]

Vertical	**Regulatory and mission-critical**	**Location-based**	**Industry**	**eHealth**	**Transport and logistics**
Use cases	Public safety **First responder** Emergency call service	Augmented reality **Wearables** (tracking, activity monitoring, and emergency messages) Advertisement push Collaborative activities (bike sharing, guidance, and flow management)	**Waste collection and management** Connected enterprises Smart retail	Connected healthcare **Patient tracking** and surveillance inside or outside Hospitals Location of emergency equipment outside Hospitals (public spaces, offices, etc.)	Traffic monitoring, management, and control Road-user charging (RUC) Asset and freights tracking **Drone tracking** Supply chain visibility Smart retail Autonomous driving, V2X
Environment of use	Both indoor and outdoor	Both indoor and outdoor in the 5G service area.	Outdoor	Both indoor (Hospitals, housing, offices, etc.) and outdoor (5G service area)	Outdoor
Accuracy	<1 m horizontal, <2 m vertical (indoor for floor detection) and <0.3 m vertical (relative) outdoor	2 m horizontal, <3 m vertical	3 m horizontal	3 m to 10 m horizontal, <3 m vertical	0.1 m to 0.5 m horizontal, 0.1 m to 0.3 m vertical
Availability	> 95% (98% outdoor)	99%	99%	from 90 to 99%	99 − 99.9%
Velocity	-	-	-	-	velocity < $3\,\mathrm{m\,s^{-1}}$
TTFF	10 ms	10 s	-	-	10 s
Latency	1 s (5 s outdoor)	1 s	60 ms	-	150 ms
Other KPI	MCX confidence event-triggered report	Normal mode	Very low energy (15 years)	-	Low energy, anti-spoofing, antitampering

achieved in many conditions [123]. Wi-Fi and BLE have already been considered as complementary technologies in LTE Release 13 to enhance positioning in indoor environments [124], chiefly thanks to their wide diffusion.

The ad-hoc solutions: Ultra-Wide Band (UWB) is known to be the most promising ad-hoc technology, which is able to achieve high-accuracy positioning in an indoor environment. This is thanks to the fact that a larger signal bandwidth ensures a higher time measurement resolution and, thereby, providing high-accuracy positioning estimations [123].

3GPP Technologies

From a standardization perspective, positioning in 5G systems is discussed within the dedicated task in Release 16 [4]. Here, localization will be performed based on both *(i)* the NR uplink and downlink signals characteristics and *(ii)* the new network configurations and technologies such as sensors, Bluetooth, RFID, WLAN, GNSS, and Terrestrial Beacon Systems (TBS) [123]. Further, the main 5G technology enhancement consists in the employment and use of massive MIMO, beamforming, and mmWave directional transmissions. The mmWave frequency band exploitation yields a two-fold advantage, which is the large available bandwidth and the possibility to pack a large number of antenna elements even in small spaces (e.g., in a smartphone) [123]. Thanks to the wide bandwidth transmission, the former advantage makes it possible to enhance multipath robustness and time resolution performance of, e.g., Observed/Uplink Time Difference of Arrival (OTDOA/UTDOA) localization approaches [125, 126].

Downlink OTDOA is assisted by UE, which means that UE receives downlink signals from the serving cell and multiple neighbors. Then, the UE determines the time difference between serving and neighbor cells to conclude on the position. Hence, it is a handset or UE-based method, where UE is responsible for measuring time difference, and it requires a specific implementation at the UE side. Alternatively, UTDOA is a network-based location estimation method, and it does not require any UE interaction for position determination [127]. Here, the uplink transmissions from UE are received by highly sensitive receivers, which will determine the time differences of arrival and, hence, UE position. The latter advantage facilitates single-anchor methods offering cm-level positioning accuracy, thereby addressing the bottleneck in indoor localization, which is a redundant ad-hoc infrastructure deployment.

From the application perspective, cellular technologies have been utilized for rough positioning for decades (e.g., second-generation mobile telecommunications (2G), third-generation mobile telecommunications (3G), fourth-generation mobile telecommunications (4G)) for those cases where accuracy is less important. Such a positioning relies on the existing communication infrastructure and, hence, does not require dedicated deployments and significant maintenance costs. However, none of these previous cellular generations can meet future networks' positioning requirements (see Table 4.1). Conversely, 5G NR systems support several novel features, such as D2D communication and network densification, high carrier frequencies and large antenna arrays, large bandwidths, that make them favorable for the positioning of sub-meter accuracy. This means that 5G NR can provide positioning services with accuracy exceeding GNSS with no additional cost [128, 129].

4.1.3 Positioning methods

There are multiple dimensions to be considered when developing a positioning system. In this subsection, we overview the existing localization method classifications. For example, Global Positioning System (GPS) can be included in such categories as distributed, absolute,

and non-cooperative localization, whereas passive RFID tags coupled with RFID readers correspond to centralized, relative, and non-cooperative localization approaches.

Indoor Localization vs. Outdoor Localization

First and foremost, positioning can be divided into two types, depending on the positioning environment: outdoor and indoor. To this end, we also specify infrastructure-based and infrastructure-less indoor positioning systems as follows.

Indoor positioning is referred to as the *last kilometer problem* since the GNSS cannot work indoors [130]. *Infrastructure-based Indoor Positioning System (IIPS)* is usually referred to as those systems that need the environment to be sensed to have indoor positioning (e.g., deploying BLE beacons, ultrasound receivers, etc.) Alternatively, *Infrastructure-less Indoor Positioning* is known as device-based positioning, wherein no additional infrastructure is required to operate, e.g., an indoor positioning system based on inertial measurements or magnetic field [131]. **Outdoor positioning** capabilities require regional or even global coverage compared to indoor environments, which are limited in size to rooms and buildings.

Indirect Localization vs. Direct Localization

Localization methods can be divided into two categories depending on the process of location estimation. It can be performed without relying on any intermediate parameters, i.e., directly from the received signal (direct) or relying on the first estimated intermediate parameters (indirect).

Multiple channel intermediate parameters' measurements of the multipath environment, e.g., RSS, Angle Of Arrival (AOA), Time Of Arrival (TOA), Angle Of Departure (AOD), Time Difference Of Arrival (TDOA), phase, or combinations of them, are used for the positioning algorithms depending on the radio technology [132, 133]. These methods are referred to as the **indirect localization** (also known as two-step localization) where first the intermediate parameters are estimated, and then the user's position is obtained by using geometrical or triangulation/multilateration manipulations [134]. The measurement phase is affected by uncertainty due to environmental changes, such as channel noise, interference, multipath, blockage, clock drifts, among others [135], that influence positioning estimates differently depending on the chosen underlying technology [135]. For example, the low penetration capabilities of the GPS signal does not allow consumer-grade GPS receivers to make reliable measurements, leading to inadequate position information [135]. Localization performance also varies depending on the specific algorithm used in the localization-update phase. An example of indirect localization is Uplink TDOA (UTDOA) used in LTE systems. Here, first, sensors estimate TDOA (UTDOA) from all incoming signals. Second, they transmit such estimates to a central node that determines the sources' position by multilateration.

Multiple **direct localization** methods, where the user's location is determined directly from the received signal, have been developed to combat the uncertainty caused by the

environmental changes [134, 136]. Here, contrary to the traditional two-step localization, the location of the source is estimated directly from the data, without estimating intermediate parameters, such as the AOAs of the LoS paths. Thus, they avoid location information loss and have higher location accuracy. Direct localization requires that the signals from multiple BSs, or a function of them, have to be obtained by a fusion center to perform the user location estimation. For example, in [134], instead of performing triangulation with the strongest signals, all received multipath components are processed by a fusion center that determines the LoS directions, leading to an estimated position through triangulation.

Active Localization vs. Passive Localization

Localization systems could be divided into active and passive systems depending on whether the users have to carry the measurements on their devices such as smartphones, smartwatches, etc. More specifically, for **active localization** systems (also known as device-based active localization), users carry specific measuring devices [137]. It means that the entity being detected and tracked carries the tag or any device attached. Alternatively, **passive localization** (also known as device-free passive localization) aids the device to estimate its location in the environment. This method provides the capability of tracking and localizing entities not carrying any devices nor participating actively in the localization process. However, in real-life applications, the accuracy of passive localization is limited due to such effects as the environmental noises, multipath, among others [138].

Centralized Localization vs. Distributed Localization vs. Decentralized

Localization schemes can be classified depending on the place where the computation is performed [119]. According to this classification, distributed and decentralized localization is sometimes referred to as self-localization, whereas centralized localization is also known as remote localization. In **centralized positioning**, the central processor collects all location and measurement data from the anchors to calculate unlocalized nodes' positions jointly. Moreover, centralized approaches are supposed to offer a more accurate location performance since the information about the entire network is available in this case [139]. Inversely, only local information exchange is required in **distributed systems** as the computation is spread over the entire network [140]. The advantage of the distributed approach consists in scalability and robustness to node failures. In **decentralized positioning**, there are different devices connected to the network, and each device can perform positioning independent of the other [141]. For example, scenarios such as a smartphone receiving signals from BSs, from which it may infer a position estimate about its actual position fall in the decentralized category.

The previous general classification leads to different computation paradigms for positioning that include Cloud Computing, Mobile Cloud Computing, Fog Computing, Mist/Things Computing, and Edge Computing. However, the computation paradigm is not usually fully

described when introducing a new regular positioning system. The main focus is the algorithm description and assessment rather than its practical implementation, which can be server-based or on-device. In contrast, for cooperative/collaborative systems, the trends suggest that the decentralized approach is the most popular in the recent few years [131].

Absolute Localization vs. Relative Localization

Depending on the way the positioning systems provide location information, absolute and relative localization methods can be defined. **Absolute localization** refers to localization in a single predetermined coordinate system. The coordinate system is usually given by a geographic coordinate system, such as latitude, longitude, and altitude in GPS localization or implied in anchor locations [142]. **Relative localization** refers to the localization method where the coordinate system may vary from node to node, e.g., in the case of one's neighbors or the local environment. Relative location without a given coordinate system is also known as a relative map or a relative configuration [143]. An absolute or a relative position estimation can be further specialized into a vertical and a horizontal position. The former term refers to the position estimation on the vertical axis or altitude, whereas the letter means positioning in a horizontal plane or two-dimensional (2D) reference [123].

Non-cooperative Localization vs. Cooperative Localization

The locations of uses can be determined either based on only the measurements between MTs and BSs, i.e., without the internode measurements between MTs or by using the inter-agent communication. In a **non-cooperative approach**, communication is established only between agents and anchors (no inter-agent links). Thus, every agent has to communicate with multiple anchors (BSs), demanding either long-range anchor transmissions or a high density of anchors [144]. In **cooperative localization**, agents are no longer required to be under the coverage of multiple anchors as inter-agent communication (though D2D) removes the need for high anchor density and long-range anchor transmissions [135]. Here, long-range BS transmissions are replaced with multi-hop communications among the densely located MTs.

In cooperative localization, agents can obtain information from both anchors and other agents within the communication range (see Fig. 4.1). Moreover, D2D links have higher SNR and lower probability of the NLoS path. Hence, cooperative localization increases the accuracy and extends the coverage compared to non-cooperative localization, which is the focus of this section.

4.1.4 D2D-based Cooperative Positioning

In this subsection, we first provide a comparison of the technologies and standards used for D2D communication. We then review the current trends in D2D-based cooperative positioning for both indoor and outdoor environments.

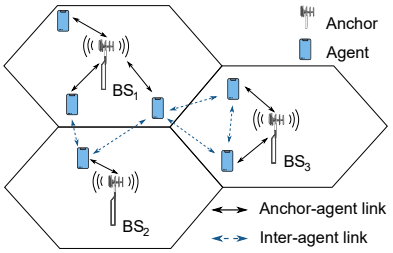


Fig. 4.1. D2D-aided positioning illustration.

D2D Technologies Comparison

D2D communications help in disseminating users' identification data, thereby facilitating direct interaction between mutually close devices, which require respective discovery and identification. Further, collective content creation and exchange enable users located in proximity to share and receive the desired content opportunistically. D2D-based interaction facilitates nearby users to participate in collaborative activities and communicate with each other's devices, thus emphasizing socialization and leisure, among many other LBS [145].

D2D technologies may be classified into two general categories: sharing licensed cellular spectrum or using dedicated resources. The former one tends to be limited by power and spectrum management perspectives and by usage cost. In contrast, the latter one suffers from uncontrolled interference and offers no QoS guarantees due to the random access behavior of, e.g., Institute of Electrical and Electronics Engineers (IEEE) 802.11 protocol stack. On the contrary, Wi-Fi outperforms cellular technologies in data rates and energy efficiency and is considered as the dominant solution for D2D connectivity, supporting *wearable aggregation nodes* [145].

Hardware costs and power consumption profiles in WLAN and Bluetooth specifications are more suitable for wearables. However, LTE and 5G road maps provide a gigabit-per-second experience, although some wearable devices may not be able to entirely derive the benefit of LTE and 5G due to potential cost and hardware complexity [146, 147].

A practical protocol for supporting D2D communications in cellular networks by jointly using Wi-Fi Direct and LTE is proposed in [148]. Here, the communication between the cluster head and cluster clients is performed over Wi-Fi Direct. Proposed D2D architecture shows good performance in terms of delay and traffic load to be supported by D2D connectivity, and the scheduler minimizes LTE packet delays, which leaves room for relaxed Wi-Fi operations at reasonable transmission rates.

Most recently, authors in [101] proved that the best option for cooperative data delivery in terms of energy consumption is to select a relay and perform D2D transmission over LTE. More specifically, the power consumption analysis from both the infrastructure and user device perspectives conducted in [101] indicates that the minimization of active interfaces and sending the data with the best possible data rate is required to save energy consumption. Since Wi-Fi is a good option only for high data rate transmissions, if there is no need for that,

communicating devices should keep only the LTE interface active to save power. Similarly, in [149], authors advocate the use of LTE sidelink transmissions to address mission-critical requirements and target new broadband public safety applications.

It is important to mention that, for example, for scenarios and applications such as emergency management (i.e., rescue and critical applications), multiple standards are required [150]. For instance, low-power Wireless Personal Area Network (WPAN) standards such as IEEE 802.15.6 are more suitable for on-body communication, they support a variety of real-time health monitoring and consumer electronics applications. However, these standards are not designed for body-to-body communication. For that case, the use of low-power WPAN IEEE 802.15.4 (Zigbee), Wi-Fi, 4G/LTE D2D is required, which can effectively extend the network connectivity and coverage. Further, for off-body communication, one of the end-devices should communicate through cellular networks or infrastructure-based networks (e.g., 4G/LTE). To summarize, even if existing devices can already support multiple standards, existing protocol stacks are not smart enough to provide connectivity or routing between different network technologies, which poses one of the critical challenges for future wireless networks [150].

5G wireless communication systems utilizes NR sidelink for D2D communication [97]. The central scenario of Release 16 NR sidelink transmissions targets V2X, and NR services are no longer limited to the Cooperative Awareness Message (CAM) and the Decentralized Environmental Notification Message (DENM) compared to LTE. In addition, by using a wider bandwidth, flexible massive antenna systems, and beamforming, NR-V2X will provide more precise timing and accurate measurement of equivalent signal techniques in LTE-V2X [152]. The new NR sidelink use cases require low-latency, high-reliability, and high-throughput transmissions, as well as a high connection density. To this end, four new designs are introduced to NR sidelink: *(i)* in addition to broadcast, also unicast and multicast are supported; *(ii)* the performance in terms of latency is improved by grant-free transmissions adopted in NR Uplink (UL) transmissions; *(iii)* it improves the channel sensing and resource allocation procedures to mitigate collisions among different sidelink transmissions initiated by various MTs; *(iv)* high connection density is achieved by supporting congestion control and QoS management. For more detail, interested readers are referred to [151]. We summarize our main findings on D2D technology comparison in Table 4.2.

Review of D2D-based Cooperative Positioning

As discussed above, 5G supports D2D communication via sidelink which can be controlled by the BS. The UL, Downlink (DL), and sidelink communications can operate over two frequency bands: sub-6 GHz and above 24 GHz (known as mmWave band). Whereas much effort has been done in the sub-6 GHz band [153], mmWave frequency spectrum is of special interest for localization [128, 154] thanks to the large available bandwidth, sparse channels, and large antenna arrays. In [155], authors investigate a real-time positioning based on mmWave

Table 4.2. D2D Technologies Comparison

Work	Technology	Advantages / Features	Device Type	Application
[101, 145, 148]	Wi-Fi	Wi-Fi provides higher data rates and energy efficiency (due to lower transmission delay) compared to cellular technologies. The devices are recommended to reduce the number of active radio interfaces and transmit with the best possible rates to minimize total power consumption. Thus, Wi-Fi is used only for high data rate services and is suitable for body-to-body communications	Wearables	Collective activities, socialization and leisure, various LBS
[101, 149, 150]	LTE	If high data rates are not required, devices should keep only the LTE interface active to save power. 4G/LTE D2D are required for wearable off-body and body-to-body communication to extend the network connectivity	IoT, wearables	Vary
[150]	IEEE 802.15.6	Low-power WPAN standards such as IEEE 802.15.6 is more suitable for on-body communication	Wearables	Public safety and critical applications
[150]	IEEE 802.15.4	For body-to-body communication, using IEEE 802.15.4 D2D is required, which can extend the network connectivity in an effective manner	Wearables	Public safety and critical applications
[146]	Bluetooth	Bluetooth (and Wi-Fi) technologies are more suitable for wearables in terms of hardware cost and power consumption. Depending on the wearable device type, devices may not be able to take full advantage of LTE and 5G due to the cost and hardware complexity	Wearables	Vary
[151]	5G NR	NR sidelink ensures low-latency, high-reliability, and high-throughput transmissions, as well as high connection density services	Mostly vehicles	Advanced and remote driving, platooning, sensors

D2D links, which will ensure reliable communication in 5G cellular networks. Authors state that mmWave signals can provide up to centimeter-level accuracy. Similarly, in [156], authors examine vehicle positioning using 5G D2D mmWave signals. *Obviously, what clearly emerges is that if all MTs' positions are unknown, the absolute position will not be obtained, which means that D2D itself can provide only relative measurements. Hence, the main interest shifts* towards integrated positioning systems.

Indoor Positioning

We first focus on the indoor environment. Recall that cooperative positioning is of extreme importance for indoor scenarios where GPS signals are usually weak to provide sufficient position information. In [157], a cooperative localization scheme for WLAN fingerprinting is proposed to improve the accuracy of fingerprint-based location estimations affected by random environmental changes. More precisely, instead of regenerating the radio map (typical for fingerprinting), users cooperate by exchanging both their real-time RSS measurements

Table 4.3. The State-of-the-art on D2D-based Indoor Positioning

Work	Methods	Approach	D2D approach	Accuracy
[119]	IoT and 5G localization and D2D	In IoT and 5G localization, the value of inter-device measurements helps to surpass current accuracy and coverage levels by at least an order of magnitude	Pseudo-range estimates	sub-meter accuracy
[120]	RF Fingerprinting and D2D	Verification of the fingerprint positioning response by TOA-based distance transmitted over D2D to nearby devices to rectify the erroneous response of BS	Pseudo-range estimates	5.34 m (static) 5.59 m (dynamic)
[135]	UWB and D2D	The cooperation between nodes is used to increase the accuracy and robustness of UWB systems	Pseudo-range estimates	accy. improvement against non-coop.
[157]	WLAN Fingerprinting and D2D	Iterative update of the initial estimated location information from WLAN fingerprinting by exchanging both real-time RSS measurements and location estimations between MTs	Location information exchange	accy. improvement against non-coop.
[158]	Dead reckoning and D2D (Wi-Fi ranging exchange)	Cooperative positioning information exchange among multiple pedestrians to reduce the accumulative error of dead reckoning carried out with only the sensors and Wi-Fi in smartphones	Pseudo-range estimates	less than 5 m
[159]	Wi-Fi Fingerprinting and D2D	Collaborative D2D-based method is used to eliminate redundant fingerprints from a crowd-sourced database	Location information exchange	less than 25 m and 50 m
[160]	RSS of Wi-Fi and D2D over Bluetooth	RSS of Wi-Fi contains information about users' relative distances, and Bluetooth signals are exchanged between them to improve the performance of existing indoor positioning methods	Pseudo-range estimates	1.34 m (average), 6.68 m (maximum)
[54, 161]	LTE (OTDOA) and D2D	OTDOA positioning is enhanced with D2D cooperative techniques	Pseudo-range estimates	accy. improvement against non-coop.

and their location estimations and processing them with the aid of the Self-Organizing Map structure. Allowing collaboration among users increases the accuracy of the non-cooperative fingerprinting approach, and the performance enhancement is higher for denser user populations.

In [158], authors exploit cooperative positioning among multiple pedestrians to reduce the accumulative error of dead reckoning carried out only by means of the smartphone low-cost sensors (infrastructure-less system) and Wi-Fi interface by exchanging RSS measurements. Here, the participants communicate and measure the range (Wi-Fi ranging) between each other and then correct each of their positions to make them consistent with the range information. The results demonstrate an accuracy comparable to GPS when a high number of pedestrians cooperate in collaborative positioning. The proposed cooperative scheme solves the typical problem of the indoor positioning system, which is the cost (as positioning

is performed on a device) and accuracy (improved by means of D2D technology). Similarly, in [135], cooperation between nodes in UWB system results in high accuracy and robustness.

Table 4.4. The State-of-the-art on D2D-based Outdoor Positioning

Work	**Methods**	**Approach**	**D2D approach**	**Accuracy**
[118]	5G and D2D	The inter-MT communications provide relative location information between the MTs that serve as a supplement to the BS–MT links	Pseudo-range estimates	1 m and below (suggested)
[154]	5G and D2D	The exchange of data between MTs and the network or between the MTs in 5G to increase the accuracy of, e.g., TDOA-based method	Pseudo-range estimates	sub-meter accuracy
[128, 162]	5G mmWave and D2D	D2D is used for disseminating and computing location information between vehicles	Pseudo-range estimates	accy. improvement against non-coop.
[163]	Radio Frequency (RF) fingerprinting and D2D	MTs communicates not only with BS to capture a series of RF measurements but also with other MTs by leveraging the D2D communication protocol	Pseudo-range estimates	accy. improvement against non-coop.
[164]	GNSS and 5G D2D	Any-time and any-where seamless positioning by using the integrated methodology of GNSS and D2D measurements	Pseudo-range estimates	56.2% better than for non-cooperative
[165]	Multi-Kalman Filter (MKF) approach and D2D	The combination of the interacting multiple model estimations with the Multi-Kalman Filter (MKF) approach based on GPS and Geographic Information System (GIS) big data	Pseudo-range estimates	-

A collaborative method based on D2D communication to enhance the indoor positioning accuracy using only direct communication to nearby devices and fingerprinting is introduced in [120]. The main idea behind this approach consists in verifying the fingerprinting positioning by TOA-based distance passed through the D2D link to rectify the erroneous response of BS. Communication with other devices is repeated several times within a short period to increase the confidence level in the verification process. However, power consumption restrictions have not been considered, while there is a growing interest in developing more energy-efficient algorithms and protocols to support Green Communications, thereby reducing the environmental and economic impact. Likewise, the benefits of direct interconnection between nodes in terms of localization accuracy (sub-meter) and coverage improvement for 5G IoT applications are demonstrated in [119].

In [159], a collaborative D2D-based method is used to remove the redundancy from the crowd-sourced Wi-Fi fingerprint database. Moreover, D2D is also used to provide privacy to the users by breaking the links between them and the data. Similarly to [158], authors

deploy a localization system on devices rather than on infrastructure, thereby solving the problem of the infrastructure cost. Yet, as we see from Table 4.3, the accuracy of Wi-Fi fingerprints localization still has room for improvement. In [160], a framework for improving the performance of existing indoor positioning methods for smartphones with the help of information exchange between users is designed. More precisely, the RSS of Wi-Fi contains information about users' relative distances, and Bluetooth signals are exchanged among them for the purpose of assessing the probability distribution functions of users' states. The average error of the position estimates of the proposed system is 1.3357 m, which can be considered as good performance for indoor localization performed on the device. The accuracy of the cooperative D2D-based localization scheme in [160] significantly outperforms the solutions proposed in the literature for low-cost indoor infrastructure-less positioning [158, 159].

As discussed in [54, 161], LTE cooperative localization technique, wherein the MT communicates with both eNodeBs[2] (OTDOA positioning method supported by 3GPP) and other MTs in proximity, can significantly improve the localizability in the network and enhance the accuracy which is highly beneficial to some applications, e.g., E911. It is demonstrated that cooperative localization is undoubtedly able to overcome problems of non-cooperative positioning (e.g., bad geometry, etc.) and, by increasing the number of collaborators, significantly improve localization accuracy. Note that the synchronization problem between MTs that may appear in distributed cooperative localization systems and cause high accuracy error rate can be handled by inband D2D (that is, LTE sidelink) communication through the primary sidelink synchronization signal (for those MTs that are under the eNodeBs coverage). However, synchronization is a challenging problem when MTs are out-of-network coverage.

Outdoor Positioning

Similarly to the indoor environment, D2D communications can enable performing cooperative localization in cellular networks, where the BS–MT links are supplemented with the inter-MT links that provide relative location information between the MTs [118].

In [154], authors investigate D2D communication capabilities to empower cooperative positioning in 5G for high-density scenarios, which have the *potential* to enable centimeter-level accuracy positioning estimates. More specifically, MTs receive and exploit signals from other MTs in proximity. The pseudo-range estimates (e.g., TDOA) between MTs are then used to increase the positioning accuracy. Authors demonstrate that at densities greater than 1,100 MTs per square kilometer, sub-meter positioning accuracy can be provided with the outage probability converging to zero.

In [163], a cooperative localization technique using RF Pattern Matching for LTE systems is proposed, wherein, by leveraging the D2D communication protocol, the MTs communicate with both the BS to capture a series of RF measurements and other MTs. Simulation results

[2] E-UTRAN Node B, also known as Evolved Node B (abbreviated as eNodeB or eNB), is an LTE BS.

of the proposed cooperative algorithm testify to a significant improvement of the positioning accuracy in cellular LTE networks.

In [164], authors focus on the integrated methodology of GNSS and D2D measurements in 5G communication systems to achieve a high-level accuracy. In addition, authors state that the high-dense property of 5G networks also eases the process of obtaining sufficient D2D measurements to achieve any-time ubiquitous positioning as D2D benefits from the high density of connected MT. The positioning accuracy of GNSS system is improved up to 56.2% compared to the non-cooperative approach.

The collaboration among the communication, signal processing, and control sub-systems in ITS systems is considered in [162], while [128], with reference to ITS, discusses the key characteristics of 5G mmWave positioning for vehicular networking that can benefit from 5G technologies, such as D2D. Moreover, even in the absence of reference stations, D2D cooperative localization allows for relative positioning.

In [165], authors investigate content distribution problems in D2D-based cooperative vehicular networks. The same authors propose the algorithm to achieve dependable content distribution through highly dynamic and unreliable D2D-Vehicle-to-Vehicle (V2V) links by combining big data-based vehicle trajectory prediction with coalition formation game-based resource allocation.

By studying Table 4.3 and Table 4.4, one can learn that *D2D-aided cooperative positioning provides increased accuracy compared to non-cooperative methods*. It is important to emphasize that D2D also helps extend the coverage since it allows for relative position estimation, even in the absence of reference stations. Further, according to Table 4.3 and Table 4.4, *5G mobile networks have enhanced positioning capabilities in comparison with previous generations and, with the help of D2D-aided cellular positioning, can ensure sub-meter accuracy needed by 5G applications and have the potential to enable centimeter-level positioning accuracy.*

We note that *D2D provides two advantages for positioning accuracy improvement*. First, the direct exchange of necessary data between MTs can be performed. In this case, both common physical layer estimates (to speed up the local decisions), as well as **position information**, are exchanged over D2D links to enhance the accuracy of the localization system. Such information, including uncertainties of estimates, is relevant to improve the convergence time of estimation processes in the MT. Second, with the implementation of D2D communication capabilities, MTs are inherently receiving signals from each other. These numerous links contain the additional signal observations and, hence, can be used to determine the **pseudo-range** estimates between MTs.

Analyzing Table 4.3 and Table 4.4, one can deduce that the cooperative positioning approach, where additional pseudo-range observations from D2D links are available, is more commonly used by the research community. With this principle, it is possible to estimate distances between MTs in the form of pseudo ranges that come on top of MTs-BSs ranging and are used in today's mobile communications systems. Thereby, seamless positioning at

a sufficient accuracy level can be achieved. In the case of location information exchange, benefits are in terms of improvement in the location process convergence time. In any case, cooperative D2D-aided positioning in 5G requires the exchange of data between the MT or between MTs and the network, which can be achieved by an appropriate position information exchange protocol.

4.1.5 Lessons Learned

This subsection summarizes the main lessons learned while exploring the existing research on D2D-based cooperative positioning. In summary, we have learned that:

1. Ultra-dense network deployments may pave the way to provide accurate collaborative positioning. D2D communication between mobile devices together with a high density of small cells in 5G NR networks paves the way for cooperative positioning applications. Moreover, sophisticated resource allocation between users can improve localization performance, as in [166]. Thus, cooperative D2D-aided positioning in ultra-dense 5G networks is foreseen to provide continuous high-level accuracy positioning estimations.

2. There is a lack of efficient protocols for cooperative positioning, which needs to be improved. Compared to previous generations of mobile networks, in the upcoming commercial 5G deployments, significantly higher positioning accuracy is anticipated. This is especially due to, among others, the capability of D2D communication to enable cooperative positioning. Cooperative positioning protocols could exploit the high density of asynchronous and synchronous nodes in future heterogeneous networks independently of the location method. Although preliminary studies have been launched within D2D communications, the exchange of location information among mobile devices, small cells, and BSs needs to be improved to enable the deployment of cooperative positioning techniques within the whole network. Thus, the standard should contain a cooperative protocol within the network elements to improve location information sharing for both communication and positioning purposes. Moreover, recent results in this research direction have already demonstrated the potential for centimeter-level localization accuracy.

3. Cooperative positioning may become complex as the number of collaborators/actors increases. While the positioning accuracy becomes higher when the number of collaborators grows, the complexity also increases. For example, in a fully connected mesh network, the number of links established among MTs is $N_{\mathrm{mt}}(N_{\mathrm{mt}}-1)$ that grows quadratically with a number of MTs N_{mt}. Therefore, the performance-complexity trade-off is a key issue that still has to be investigated by the research community. Moreover, when the number of possible collaborators exceeds the number of needed collaborators to perform the accurate positioning, the procedure for collaborators selection has to be established. For example, random selection can be determined.

4. There is a trade-off between energy consumption and high transmission rates/low latency. From an energy consumption perspective, minimizing the active interfaces and sending

data at the best possible data rate will bring the lowest energy usage on the MT side. Thus, for D2D transmission, MTs should keep only the cellular interface active to save energy unless there is a need for very high data rate transmissions, which depends on the application. However, high transmission rates guarantee low latency services. Therefore, finding an optimal power consumption vs. latency trade-off presents itself as an important research direction, especially for high-accuracy positioning. In addition, different from previous cellular technology generations, 5G NR is able to meet the future networks' positioning requirements of less than one-meter accuracy.

5. D2D technology reduces location estimation error under NLoS environments. The localization accuracy sharply declines once the receiver enters NLoS environment. Cooperative localization also combats this problem, increases the accuracy, and extends the coverage compared to non-cooperative localization. Moreover, it can provide resilient relative localization estimates even when the network infrastructure is not available.

4.2 A Theoretical Model for Cooperative RSSI-based Localization with D2D versus RIS

4.2.1 Motivation

While the 5G cellular system is spreading around the world, researchers are beginning to work on the 6G mobile communication networks [167, 168]. 6G is expected to provide intelligent and ubiquitous wireless connectivity over 3D network coverage at Terabits per second data rates and sub-millisecond latency. Therefore, acquiring accurate location information from MTs is becoming increasingly crucial for achieving the mentioned objectives. And this is not only true for location-based services but also for improving wireless communication performance in various ways, including channel estimation, beam alignment, medium access control, routing, and network optimization [169,170]. In this regard, for accurate localization and sensing systems, we can expect localization of end devices using 5G/6G technology to exploit D2D [171] and RIS *cooperative techniques* [172].

As detailed in section 4.1, cooperative localization is a promising technology that offers additional information for positioning by exploiting cooperative links between multiple MTs. In D2D-aided cooperative localization, D2D links loose the requirements for all MTs to be connected to any of the available BSs, allowing multi-hop communications among densely located MTs to replace long-range BS-MT communications. Cooperative localization can provide more accurate positioning and enhanced coverage than noncooperative localization, thanks to the D2D links that have better SNR due to operation over short distances and a lower probability of blocked LoS [118].

Indeed, the NLoS problem is one of the most challenging problems for 6G, which can drastically reduce localization accuracy. While very high carrier frequency (e.g., mmWave and Terahertz (THz)) and large antenna arrays aim at providing outstanding opportunities

for accurate wireless localization, millimeter-scale wavelength usage results in severe path loss, primarily in NLoS scenarios. Hence, the blockage is a major obstacle to the widespread use of accurate wireless localization systems operating in the mmWave and THz ranges [173]. RIS, known as an energy- and cost-efficient technology, can be used to establish a LoS link between the transmitter and the receiver even in the presence of obstructions or when the received power from the direct path does not allow for a reliable connection [174]. Thus, RISs are regarded as one of the foremost technologies capable of controlling the physical propagation environment wherein they are embedded by passively reflecting radio waves in desired directions and actively sensing this environment in both receive and transmit directions [175].

RIS as a reflector can work in two ways: *(i)* as part of the passive environment, operating like any other scatterer or reflector, and *(ii)* as part of the infrastructure, acting as a global reference or anchor point [176]. The potential of RIS for localization has received only limited coverage in the literature [176]. A multi-user RIS-enabled localization problem is studied in [175], where the users' position in 3D space is estimated by calculating the TOA of the LoS and NLoS paths at multiple receivers. In [177], RIS-aided mmWave MIMO systems for joint positioning and communication. Conversely, in [178], RIS is exploited as a means for blockage mitigation and a channel state information acquisition for a RIS-aided mmWave system is performed via 3D positioning.

In general, RIS technology provides many technological advantages over existing mainstream technologies [179]. However, the central question that can stimulate the practical RIS development of "what is a convincing use case for RIS?" remains unanswered [180]. Coverage expansion is one possibility, but traditional half-duplex relaying is a viable alternative, and full-duplex regenerative relays are emerging [181]. RIS technology has a competitive disadvantage over wideband channels since each RIS element must be programmed identically over the entire frequency range. Another possible use case is improved spatial multiplexing and interference reduction, but it must compete with cell-free massive MIMO. Perhaps the RIS technology will be most effective in the THz regions, where the development of coherent transceivers is challenging, and the sparse channels make additional propagation paths important even if they are poor. So far, there is no evidence, only guesses [180]. In this section, we strive to fill this gap.

Radio positioning aims to solve a set of nonlinear equations given a collection of location-related data, such as TOA and AOA. As a result, the precision of the measurements used directly impacts the ultimate positioning performance. The overall number of available measurements in the system is determined by the number of participating target nodes, the deployed anchor nodes, and their geometry, whereas the measurement accuracy is determined by signal characteristics and specific estimation methods.

Several location estimation schemes have been investigated so far that are characterized by particular benefits and shortcomings. For example, TOA and TDOA require strict synchronization, while the AOA method needs complicated receivers [182, 183]. These systems

all suffer from complex receiver structures. Many research works still focus on the simple trilateration that uses RSS to address this issue and simplify the receiver structure [184–186]. RSS is one of the most popular outdoor and indoor localization techniques based on signal propagation measurement that may be converted into distance measurements subject to suitable path loss models.

In this section, we study cooperative RSS-based localization using D2D and RIS technologies to increase the localization accuracy in cellular systems by offering a high probability of LoS links. Although much effort has been devoted to evaluating the performance of D2D and RIS-assisted localization systems separately, a comparison of these technologies from a localization perspective and the synergies resulting from their *rendezvous* have not been sufficiently investigated despite the fact that they may produce a new perspective for research and industrial communities. This serves as a motivation for this study.

4.2.2 Cellular BS-MT and Relay Model

In this subsection, we introduce D2D- and RIS-aided localization models based on RSS ranging while considering a general path loss model. Notations used throughout this chapter are summarized in Table 4.5.

Assumption 1. We consider an environment that is not friendly for localization purposes (scattered deployment of beacons). Therefore, cooperation among peers can improve accuracy.

Assumption 2. One user has no access to the localization system. Here, he can estimate the position based on information received from peers/RIS as a part of the infrastructure.

D2D-aided Localization Using Power Measurements

The purpose of localization is to find the 3D coordinates of specific targets. However, a necessary condition for positioning is that there exist at least four beacon nodes in wireless communication networks. The positioning algorithm typically depends on range measurements [187]. Generally, four ranging techniques are utilized, such as RSS, TOA, AOA, TDOA [188], frequency of arrival (FOA), and frequency difference of arrival (FDOA).

Most commonly, positioning-related measurements can be divided into four classes:

- Time-based (TOA, TDOA, round-trip time);
- Angle-based (AOA, angle of departure);
- Power-based (RSS, backscattered power) using path loss models;
- Frequency-based (FOA, and FDOA).

Given that RSS decreases according to the path loss models, a general model to convert power measurements m into the distance can be written as in [189]

$$m = f(d) + n, \text{hence}, d \approx f^{-1}(m), \tag{4.1}$$

Table 4.5. Main Notation for Cooperative RSSI-based Localization

Communication	
h_U	Height of MTs, m
h_A	Height of NR BS, m
h_B	Height of blockers, m
r_B	Radius of blockers, m
W	Available bandwidth, Hz
W_{GHz}	Operating bandwidth in GHz
f_c	Carrier frequency, GHz
d_{2d}	Two-dimensional distance between MT and NR BS, m
d	Three-dimensional distance between MT and NR BS, m
γ	Signal-to-noise ratio, SNR, W
P_T, P_R	Transmit/received power, W
G_T, G_R	Transmit/received antenna array gains, dBi
N_0	Power spectral density of noise, dBi/Hz
$L(d)$	Path loss in linear scale
$L_{dB}(d)$	Path loss in decibel scale
A, ζ	Propagation exponents
$p_B(d)$	Distance-dependent blockage probability
w	Shadow fading, dB
σ_{SF}	Standard deviation of noise
a^{bs}, a^{mt}	Flat fading coefficients
p_L	LoS probability
Localization-specific parameters	
m	Measurements
n	Noise
α_i	Constant term which takes into account the transmission power of the node to be localized (propagation coefficient)
$L(d_0)$	Path loss at the reference distance d_0
d_0	Reference distance, m
(x, y, z)	Unknown coordinates
(x_i, y_i, z_i)	Position coordinates of anchor i
$\hat{d}_i$	Calculated distance from anchor i
RIS-specific parameters	
d_{SR}	Distance from BS to RIS, m
d_{RD}	Distance from RIS to MT, m
$L(d_{\mathrm{SR}})$	Path loss of sub-path from BS to RIS
$L(d_{\mathrm{RD}})$	Path loss of sub-path from RIS to MT
Γ_k	Reflection coefficient of the k-th RIS element
L_{TOT}	Total path loss of a RIS

where $f(.)$ is the model function, d is the distance, n is noise, and m are measurements.

Among several channel models proposed for outdoor and indoor environments (Nakagami, Rayleigh, Ricean, etc [190]), the most popular one for RSS-based localization is the simple lognormal shadowing path loss model, which expresses the following relationship between the received power and the transmitter-receiver distance [189, 191, 192]:

$$RSSI(dBm) = \alpha_i(dBm) - 10\zeta \log(d) + w, \tag{4.2}$$

Table 4.6. Propagation Models [6–8]

Environment	**PLE**, ζ	**SF**, σ_{SF}	$L_{db}(d_0)$, non-blocked	$L_{db}(d_0)$, blocked
3GPP Indoor, InH - Office LoS	1.73	3	$32.4 + 20\log_{10} f_c$	$47.4 + 20\log_{10} f_c$
3GPP Indoor, InH - Office NLoS	3.19	8.29	$32.4 + 20\log_{10} f_c$	$47.4 + 20\log_{10} f_c$
mmMAGIC UMi Street Canyon LoS	1.92	2	$32.9 + 20.8\log_{10} f_c$	$47.9 + 20.8\log_{10} f_c$
mmMAGIC UMi Street Canyon NLoS	4.5	7.82	$31 + 20\log_{10} f_c$	$46 + 20\log_{10} f_c$
5GCM UMi Open Square LoS	1.85	4.2	$32.4 + 20\log_{10} f_c$	$47.4 + 20\log_{10} f_c$
5GCM UMi Open Square NLoS	2.89	7.1	$32.4 + 20\log_{10} f_c$	$47.4 + 20\log_{10} f_c$
3GPP UMi Street Canyon LoS	2.1	4	$32.4 + 20\log_{10} f_c$	$32.4 + 20\log_{10} f_c$
3GPP UMi Street Canyon NLoS	3.19	7.82	$32.4 + 20\log_{10} f_c$	$32.4 + 20\log_{10} f_c$
3GPP UMa LoS	2.2	4	$28 + 20\log_{10} f_c$	$43 + 20\log_{10} f_c$
3GPP UMa NLoS	3	7.8	$32.4 + 20\log_{10} f_c$	$47.4 + 20\log_{10} f_c$

where α_i is a constant term that takes into account the transmission power of the node to be localized (propagation coefficient), d is the distance between the transmitter and receiver, ζ is the Path Loss Exponent (PLE), and w is a zero-mean Gaussian random variable.

$$w[dB] \sim \mathcal{N}(0, \sigma_{\text{SF}}), \tag{4.3}$$

where standard deviation σ_{SF} is specified in Table 4.6 for different propagation environments.

Denote $\alpha_i = P_r(d_0)$:

$$\alpha_i = P_T + G_T + G_R - L(d_0), \tag{4.4}$$

where P_T is the transmit power, G_T and G_R are transmit/receive antenna gains, and $L(d_0)$ is the path loss at the reference distance d_0 (usually $d_0 = 1\,\text{m}$).

By using expression (4.2), we get that the distance between that MT and the BS i (or between the MT and relay i) can be read as

$$\hat{d} = 10^{\frac{\alpha_i - RSSI_i + w}{10\zeta}}. \tag{4.5}$$

The quadratic equation for 3D trilateration localization algorithm is as follows

$$\hat{d}_1 = \sqrt{(x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2}, \tag{4.6}$$

$$\hat{d}_2 = \sqrt{(x_2 - x)^2 + (y_2 - y)^2 + (z_2 - z)^2}, \tag{4.7}$$

$$\hat{d}_3 = \sqrt{(x_3 - x)^2 + (y_3 - y)^2 + (z_3 - z)^2}, \tag{4.8}$$

$$\hat{d}_4 = \sqrt{(x_4 - x)^2 + (y_4 - y)^2 + (z_4 - z)^2}. \tag{4.9}$$

where (x, y, z) is the estimated target MT coordinate, (x_i, y_i, z_i) is the position of the anchors and $\hat{d}_i$ is the calculated distance from target MT to anchor i, $i = 1, ..., 4$. Since the equations contain three unknowns and four equations, any three equations from (4.6)-(4.9) can be used to calculate unknowns. In practical scenarios, $\hat{d}_i$ might not be correct, hence, to have an accurate value, we utilize four reference points to estimate the estimated target MT coordinate.

We consider the human body and building blockage with corresponding LoS/NLoS states, which presumes outdoor scenarios, such as city streets, squares, stadiums, etc. For example,

similarly to [193], one may consider the scenario where the MT can be in one of the four states: (LoS,blocked), (LoS,non-blocked), (NLoS,blocked), (NLoS,non-blocked). Here, NLoS state means that buildings can also block the path between the BS and the MT, whereas the blocked state assumes the human blockage of 15 dBm. We consider both large-size structures, e.g., buildings on the radio path, and small-scale ones, such as the human body and foliage.

Then, the associated path loss measured in dB for the four states (LoS nBl.; LoS Bl.; NLoS nBl; NLoS, Bl.) is given by

$$L_{dB}(d) = L_{dB}(d_0) + 10\zeta \log_{10} d, \tag{4.10}$$

where d is the 3D distance between the mmWave BS and the MT (and between the two MTs), f_c is the carrier frequency in GHz, PLE ζ and $L_{dB}(d_0)$ depend on the propagation environment and are gathered in Table 4.6.

The LoS probability for the 2D distance d_{2D} between the NR BS and the MT (and between the two MTs), p_L, can be obtained by using propagation models as in [6–8] (see Fig. 4.2), whereas the human blockage probability at the 3D distance d is derived in [42].

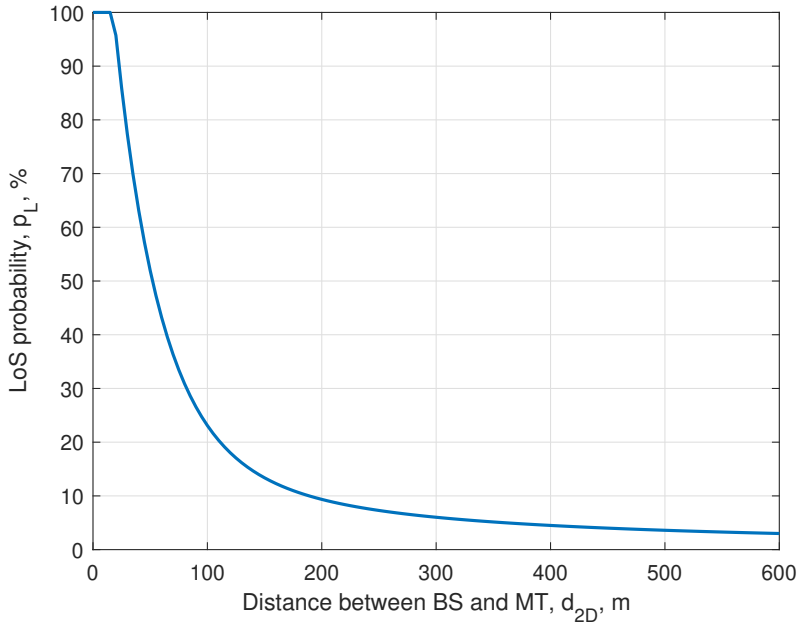


Fig. 4.2. LoS probability as a function of 2D distance between the BS and the MT according to 3GPP UMi Street Canyon model.

The path loss (4.10) in the linear scale can be represented by utilizing the model in the form of $L(d) = Ad^{\zeta}$, where A and ζ are the propagation coefficients, and $A = 10^{\frac{L_{db}(d_0)}{10}}$. Therefore, the flat fading coefficients are

$$a^{bs} = 10^{-\frac{L_{dB}(d)+w}{20}}. \tag{4.11}$$

For D2D links between MTs, we assume free space path loss without shadow fading [154], being the flat fading coefficients:

$$a^{mt} = 10^{-\frac{L_{dB}(d)}{20}}. \tag{4.12}$$

RIS-aided Localization Using Power Measurements

Here, we assume the concept of Reflection Unit Set (RUS) proposed in [178] and calculated $\hat{d}_i$ for each RUS_i, $i = 1, ...4$. Here, given four specific RUSs as anchor nodes with known locations with coordinates are $p_1 = (0, 0, 0)$, $p_2 = (0, a, 0)$, $p_3 = (0, a, b)$ and $p_4 = (0, 0, b)$ (both a and b are real values), the unknown position of the MT is (x, y, z) and the measured distances between each anchor and the MT are d_i, respectively. Then, the actual distances are defined as

$$\hat{d}_1 = \sqrt{x^2+y^2+z^2}, \tag{4.13}$$

$$\hat{d}_2 = \sqrt{x^2+(y-a)^2+z^2}, \tag{4.14}$$

$$\hat{d}_3 = \sqrt{x^2+(y-a)^2+(z-b)^2}, \tag{4.15}$$

$$\hat{d}_4 = \sqrt{x^2+y^2+(z-b)^2}. \tag{4.16}$$

The three equations can be solved to determine the three unknowns (x, y, z). However, as discussed above, whether in terms of clock synchronization or antenna array, the four equations should be strictly solved for accurate estimation.

In the case of power measurements of the cascaded BS-RIS-MT channel, distance can be calculated as per (4.5) with

$$\alpha_i = P_T + G_T + G_R + |\Gamma_i| - L_{TOT}(d_0), \tag{4.17}$$

where Γ_i and L_{TOT} are calculated as described below in this section. Then, the distance away of the MT from the BS i is

$$d = 10^{\frac{\alpha_i - RSSI_i + w}{10\zeta}} - d_{\text{BS,RIS}}, \tag{4.18}$$

where RSSI is calculated as (4.21), $d_{\text{BS,RIS}}$ is the known distance between the BS and RIS.

Assumption 3 [194]. The RIS is composed of $M \times N$ sub-wavelength elements, each with the size of $s_M \times s_N$. Assume that Dm, n and dm, n are the distance between the BS and the (m, n)-th RIS element, and the distance between the (m, n)-th RIS element and the MT, respectively. The distances from different RIS elements to the BS are approximately the same, i.e., $D_{m,n} = d_{\text{SR}}$, where d_{SR} is the distance between the BS (or, the same, source (S)) and the center of the RIS.

The path loss models from Table 4.6 are utilized to express the path loss of sub-paths, which are defined from S to RIS ($L(d_{\text{SR}})$) and from RIS to destination (D) ($L(d_{\text{RD}})$). Total received power in watts at the D through the k-th RIS element is calculated under the plate scattering paradigm [195] as

$$P_{\text{rx},k} = \frac{P_T|\Gamma_k|G_T G_R}{L(d_{\text{SR}})L(d_{\text{SR}})}, \tag{4.19}$$

where $L(d_{\text{SR}}$ and $L(d_{\text{RD}})$ are path losses at distances d_{SR} and d_{RD}, respectively, Γ_k is the reflection coefficient of the k-th RIS element, which is given by

$$\Gamma_k = e^{-j\varphi_k} G_t^e G_r^e \epsilon_b, \tag{4.20}$$

where φ_k is the phase difference induced by k-th RIS element, G_t^e is the gain of the RIS in the direction of an incoming wave, G_r^e is the gain of RIS in the direction of a received wave and ϵ_b is the efficiency of RIS, which is described as a ratio of power transmit signal power by RIS to received signal power by RIS. It is assumed that RIS consists of passive elements and $\epsilon_b = 1$.

The total received power at the receiver including all RIS elements is expressed as:

$$P_{T,k} = \left(\sum_k \sqrt{\frac{P_T |\Gamma_k| G_T G_R}{L(d_{\mathrm{SR}}) L(d_{\mathrm{RD}})}} e^{j\phi_k} \right)^2, \tag{4.21}$$

where ϕ_i represents the phase delay of the signal received through k-th RIS element.

Assumption 4 [196]. For simplicity, we assume that RIS-elements reflect signal with unit-gain reflection coefficients ($|\Gamma_k| = 1$) and in such a way that all the signals coming through different RIS elements are aligned in phase at the receiver ($\phi_k = \varphi_k$).

Then, expression (4.21) becomes

$$P_{\mathrm{rx},k} = \left(\sum_k \sqrt{\frac{P_T G_T G_R}{L(d_{\mathrm{SR}}) L(d_{\mathrm{RD}})}} \right)^2. \tag{4.22}$$

Therefore, the total path loss is given by

$$L_{TOT} = \left(\sum_k \sqrt{\frac{1}{L(d_{\mathrm{SR}}) L(d_{\mathrm{RD}})}} \right)^{-2}. \tag{4.23}$$

The SNR in the presence of a RIS, in this case, can be derived as

$$\gamma = \frac{P_T G_T G_R}{N_0 W L_{TOT}}, \tag{4.24}$$

where N_0 is the power spectral density of noise in the channel, W is the operating bandwidth in Hz.

The data rate in presence of a RIS can be calculated according to the Shannon-Hartley theorem:

$$D[\mathrm{Gbps}] = W_{\mathrm{GHz}} \log_2(1 + \gamma), \tag{4.25}$$

where γ is the SNR in linear scale, W_{GHz} is the operating bandwidth in GHz.

4.2.3 Passive Positioning with RISs vs Relays

When analyzing the current studies on the comparison between RIS and D2D performance, we focus on spectral efficiency, complexity, and energy efficiency, among other relevant metrics (see Table 4.7). First, it is worth highlighting the spectral efficiency gains of RIS (compared to relays) when their size is sufficiently large as compared with the wavelength of the radio waves [180, 197, 198].

Furthermore, a RIS naturally operates in a full-duplex mode without self-interference or introducing thermal and addictive noise since nearly passive RISs cannot amplify or

Table 4.7. RIS vs Relay

Advantage	**RIS**	**Relay**
Spectral efficiency [**?**, 180, 197, 198]	✓ (with large surface area)	✓ ✓ (full duplex)
Hardware complexity/cost [180, 198]	✓ (lower complexity)	×
Noise [**?**, 197, 198]	✓ (not affected by addicted noise, no self-interference)	× (affected)
Energy efficiency [199]	× (except when very high rates are required)	✓
Signal range [180]	× (reduced signal range due to the lack of amplification)	✓
SNR (Fig. 4.3)	×	✓
Power budget [198]	✓	× (power sharing between BS & relay)
Power consumption [180, 198]	✓ (use of printed meta-material requires no amplifiers)	×

regenerate the signals. Therefore, they might achieve higher spectral efficiency than active half-duplex relays.

Then, the unique advantage of RIS is that it reduces the hardware complexity (with analog beamforming, no extra RF chains are needed for demodulation and modulation) at the price of requiring a larger surface [180].

Differently, a system that switches between the Single-Input Single-Output (SISO) and decode-and-forward relaying modes is preferable both in terms of minimizing the transmit power and maximizing the energy efficiency, except when very high rates are required [199]. In this case, when data rate exceeds $8.48\,\mathrm{bit\,s^{-1}\,Hz^{-1}}$, RIS is the best option in therm of energy efficiency. However, another drawback of RIS is the reduced signal range [180] (meaning also reduced SNR) due to the nearly passive nature and, hence, lack of amplification.

Thus, in Fig. 4.3, we show SNR as a function of the proportion of d_{SR} distance to the total $d_{SR} + d_{RD}$ distance of 100 m. We also present the relay SNR with the distance of 100 m between devices for both blockages by large- and small-scale constructions (i.e., NLoS non-blocked and LoS blocked). As one may observe, only RISs of large size can perform comparably to the relays (when they are placed either close to the source –S– or to destination –D–). It is important to note further that distances between relays are usually smaller than 100 m, which means that the SNR between relays will be even higher.

Regarding the power budget, in relay-aided systems, it is usually assumed that the total RF power is allocated between the transmitter and the relay to ensure a total power constraint. Conversely, in the ideal case, the total power reflected by a RIS is the same as the total power of the impinging radio waves [198]. In addition, the use of printed meta-materials requires no amplifiers, which is favorable for energy consumption [180, 198]. Finally, RISs are

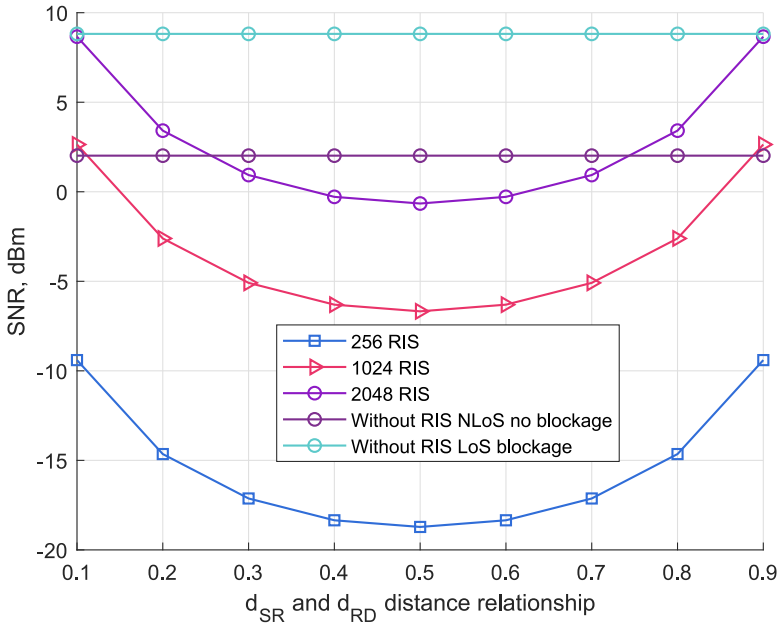


Fig. 4.3. SNR comparison of RIS and relays.

semi-passive devices with low cost, which make them ideal to be mounted on surfaces as well as moving objects [175].

In the following subsection, we aim to provide the localization performance accuracy of RIS- and D2D-assisted localization systems.

4.2.4 Performance Assessment

In this subsection, we compare the performance of D2D- and RIS-cooperative localization systems in terms of positioning accuracy, considering different scenarios (i.e., 3GPP UMi Street Canyon for LoS and NLoS conditions) and deployments of devices as described below. Note that we have chosen one path loss model only for the following reasons. First, we use the same propagation model for the two technologies for a fair comparison. Then, applying more models will lead to the same conclusions in a qualitative matter but varying in terms of quantity. Thus, for the sake of brevity and without losing generality, we focus on the 3GPP UMi Street Canyon and alter anchors' positions.

Simulation Setup

For the **cellular positioning**, we assume the service area $100\,\mathrm{m} \times 100\,\mathrm{m}$, where the four BSs are fixed at the corners of the square (see Fig. 4.4). That is, the coordinates of the BSs are $(0,0)$, $(0,100)$, $(100,100)$, and $(100,0)$. Similarly, in the case of **RIS-aided positioning**, RISs are positioned close to the BSs so that it works effectively (with a BS-RIS distance of $\approx 10\,\mathrm{m}$) [196]. Thus, the chosen RISs coordinates are known and are $(7.1,7.1)$, $(17.10,92.9)$, $(92.9,92.9)$, and $(92.9,7.1)$. We vary the locations of relays depending on the considered

scenario (the three use cases are drawn in Fig. 4.4). Namely, D2D anchors *(i)* have the same coordinates as RIS ones, *(ii)* uniformly distributed within the area of interest, and *(iii)* uniformly distributed within the circle with radius 20 m.In relay-aided systems, it is usually assumed that the total RF power is allocated between the transmitter and the relay to ensure a total power constraint. Moreover, by employing a FD relaying protocol, one pays the cost of introducing high loop-back self-interference I_S at the relay because of the concurrent transmission and reception of signals.

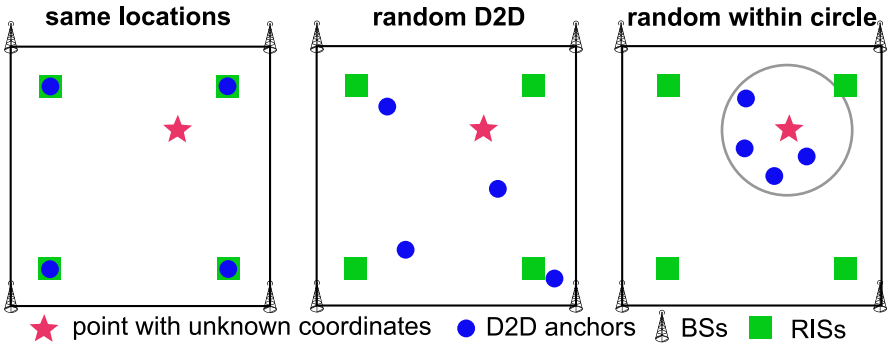


Fig. 4.4. Visualisation of devices deployments: considered use cases.

To quantify the positioning error, we calculate the euclidean distance between the ground truth position and the estimated position as it corresponds to the shortest distance between two points without considering obstacles. We report the mean, median, standard deviation as well as minimum, 75^{th}, and 95^{th} percentile of the positioning errors over M evaluation points and *rep* repetition times. In addition, we also provide the Mean Squared Error (MSE) (4.26) and the Root Mean Squared Error (RMSE) (4.27), as they penalize outlier positioning errors. These metrics have been selected according to the ISO18305 standard [200] and [201]. Recall that the mean is the same as the average value, whereas the median is the mid-value in the ordered data set (for illustration, see Fig. 4.5).

$$\text{MSE} = \frac{1}{M}\sum_{k=1}^{M}(x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2, \tag{4.26}$$

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{k=1}^{M}(x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2}. \tag{4.27}$$

where k stands for the k-th evaluation point, M is the number of evaluation points, (x_k, y_k) is the real position, $(\hat{x}_k, \hat{y}_k)$ is the position estimate.

In the performed simulations, we randomly distribute unknown coordinates (x, y) for $M = 30$ evaluation points and measure the location positioning accuracy 1000 times. i.e., we have run $rep = 1000$ simulations to deal with different random initializations. We use a loose coupling approach for cooperative positioning, which consists in fusing positions obtained

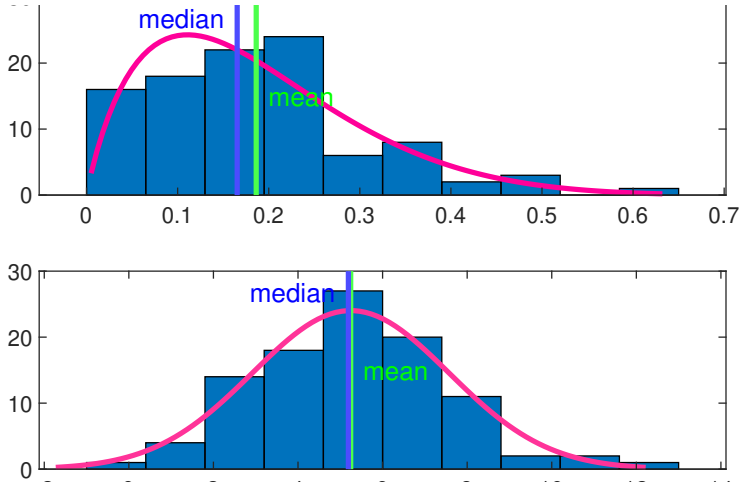


Fig. 4.5. Visualisation of median and mean on skewed and normal distributions.

from different technologies. The D2D-aided cellular position estimate corresponds to the centroid from cellular and D2D positioning. That is, their mean $\hat{x}$ and mean $\hat{y}$.

The organization of this subsection is as follows. We begin with investigating a simple LoS scenario for UMi Street Canyon with a known PLE of $\zeta = 2.1$ and $\sigma = 4$ for equal placement of RISs and relays, random placement of relays, and random placement of relays within 20 m radius, respectively. Then, we investigate the impact of NLoS conditions on D2D and cellular positioning for UMi Street Canyon with $\zeta = 3.19$ and $\sigma = 7.82$, while RIS technology preserves an assumption of LoS operation only. Then, we analyze the impact of the number of RIS elements on positioning accuracy. Finally, we summarize all the results. Table 4.8 provides the parameters for the performed simulations.

Table 4.8. Default Parameters

Parameter	**Value**
Carrier frequency, f_c	28 GHz
Transmit power (RIS/BS), P_T	20 dBm [111]
Operating bandwidth, B_{Hz}	100 MHz
Transmit power (relay), $P_T/2$	10 dBm
Noise power (receiver), N_0	$-174 + 10\log(B) + 10$
Self-Interference, I_S	$10N_0P_T/2$
Height of AP, h_A	3 m
Height of blocker, h_B	1.7 m
Height of UE, h_U	1.5 m
Blocker radius, r_B	0.4 m
BS antenna array	32×32
Number RIS elements, N_{RIS}	1024 el
Environment	UMi Street Canyon [6]
Number of evaluation points, M	30
Number of repetition times, rep	1000

Equal Placement of RISs and Relays

By way of example, Fig. 4.6 illustrates the triangulation method, where pink and blue circles represent real and estimated positions, respectively. In Fig. 4.7(a) and Table 4.9(a), we show the results of the first evaluation campaign when RISs and relay devices have the same locations (i.e., placed equidistantly from the point with unknown coordinates [198]). In a relay system, a total power constraint is assumed. Note that in this particular scenario, no technology dominates over the others for all considered metrics. For example, in Table 4.9(a), the best value of each metric does not belong to the same (almost the same) configuration as in the scenarios below (see discussion in subsections 4.2.4, 4.2.4, and 4.2.4). This might be explained by the fact that since the devices and RISs are deployed at the same positions, the communication happens equidistantly, and the channel quality is similar. However, the noise in the systems is not correlated, which explains why positioning errors are independent.

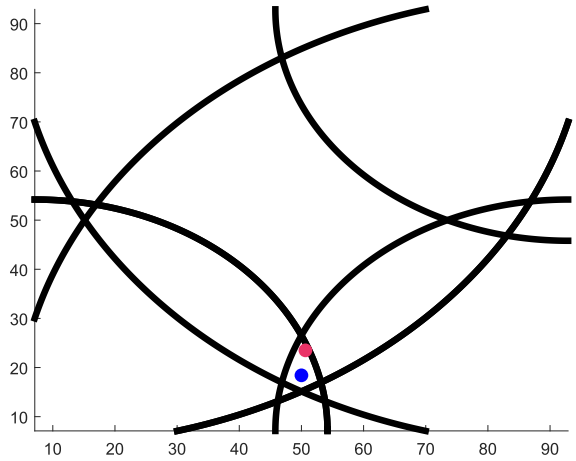
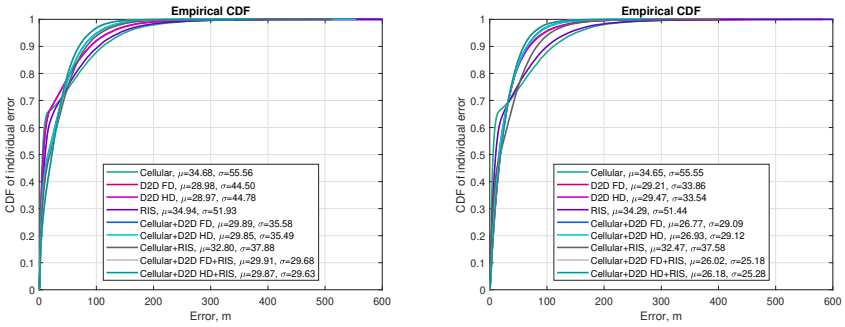


Fig. 4.6. Visualisation of triangulation. Pink and blue circles are real and estimated positions.

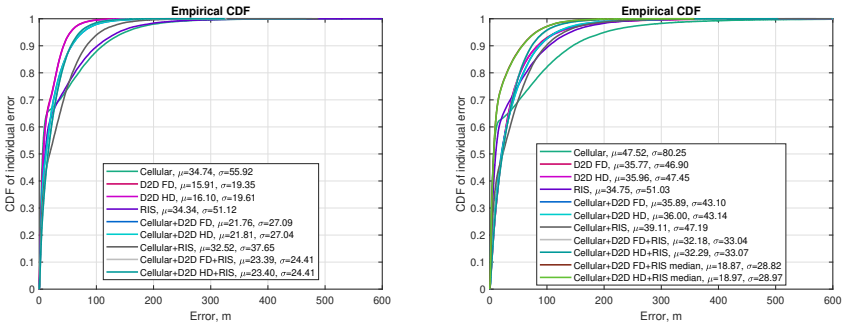
Random Placement of Relays

We randomly deploy four D2D relays within the considered area of interest to participate in D2D-aided localization, as shown in Fig. 4.4 with the caption "random D2D". Other parameters remain the same. Under this deployment option, as shown in Fig. 4.7(b) and Table 4.9(b), D2D localization by itself outperforms the cellular and RIS-based ones. We also note that depending on the noise in the system (a normal distribution with $\sigma_{\mathrm{SF}} = 4$, which is independent for all the technologies), the accuracy may vary to some extent. For example, see RIS and Cellular positioning accuracy in Fig. 4.7(a) and Fig. 4.7(b). In Fig. 4.7(b), RIS outruns the cellular positioning in terms of mean and standard deviation. Also, we again notice that the combination of all three technologies with FD relaying provides an improvement of cellular positioning by 24.9% in terms of mean error.



(a) LoS, equal placement of RISs and relays

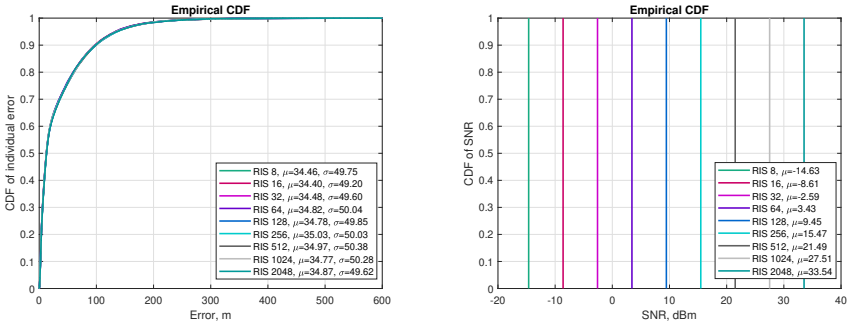
(b) LoS, random placement of relays within 100 m×100 m



(c) LoS, random placement of relays within circle of 20 m radius

(d) NLoS, random placement of relays within 100 m×100 m

Fig. 4.7. CDF of individual error.



(a) RIS, errors.

(b) RIS, SNR.

Fig. 4.8. CDF of individual error, SNR for RIS with different number of antenna elements, LoS, PLE=2.1, SF=4.

Random Placement of Relays Within 20 m Radius

We continue by studying the performance of different localization configurations when D2D relays are randomly distributed around the device with unknown coordinates within an area with radius of 20 m. Analyzing Fig. 4.7(c) and Table 4.9(c), one may notice a substantial

Table 4.9. Individual errors, [m]

(a) LOS, PLE=2.1, SF=4. RISs and relays have same locations									
Configuration	**min**	**max**	**mean**	**median**	**std**	**75thprc**	**95thprc**	**MSE**	**RMSE**
Cellular	0.0176	1024.0	34.68	**5.80**	55.56	52.73	148.69	4290.3	65.50
D2D FD	**0.0072**	597.89	28.98	7.30	44.50	41.24	121.79	2820.6	53.10
D2D HD	0.0124	895.01	**28.97**	7.27	44.78	**40.90**	122.29	2844.2	53.33
RIS	0.02	756.46	34.94	10.36	51.93	49.33	140.35	3917.4	62.58
Cellular+D2D FD	0.020	511.73	29.89	15.57	35.58	46.00	101.35	2159.0	46.46
Cellular+D2D HD	0.0387	553.51	29.85	15.56	35.49	46.18	100.86	2151.0	46.37
Cellular+RIS	0.011	512.31	32.80	18.83	37.88	50.07	107.70	2510.7	50.10
Cell.+D2D FD+RIS	0.018	**341.37**	29.91	22.18	29.68	44.08	**88.93**	1775.8	42.14
Cell.+D2D HD+RIS	0.026	369.23	29.87	22.19	**29.63**	44.01	88.76	**1770.1**	**42.07**
(b) LOS, PLE=2.1, SF=4. D2D relays are uniformly distributed within 100 m×100 m									
Configuration	**min**	**max**	**mean**	**median**	**std**	**75thprc**	**95thprc**	**MSE**	**RMSE**
Cellular	0.012	633.62	34.65	**5.85**	55.55	52.26	150.23	4286.1	65.47
D2D FD	0.018	582.55	29.21	18.45	33.86	37.86	94.65	1999.7	44.72
D2D HD	0.032	444.76	29.47	18.70	33.54	38.44	95.46	1993.4	44.65
RIS	0.039	730.56	34.29	10.39	51.44	47.79	137.38	3821.1	61.82
Cellular+D2D FD	0.054	328.85	26.77	16.08	29.09	38.19	84.65	1563.0	39.53
Cellular+D2D HD	0.035	319.29	26.93	16.41	29.12	38.03	85.99	1573.3	39.66
Cellular+RIS	**0.010**	401.76	32.47	18.04	37.58	49.38	106.38	2466.2	49.66
Cell.+D2D FD+RIS	0.035	**264.09**	**26.02**	18.36	**25.18**	**37.14**	**75.20**	**1310.6**	**36.20**
Cell.+D2D HD+RIS	0.041	270.54	26.18	18.47	25.28	37.24	76.11	1324.6	36.40
(c) LOS, PLE=2.1, SF=4. D2D relays are uniformly distributed within circle with radius of 20 m									
Configuration	**min**	**max**	**mean**	**median**	**std**	**75thprc**	**95thprc**	**MSE**	**RMSE**
Cellular	0.027	959.19	34.74	**5.88**	55.92	52.55	149.77	4333.6	65.83
D2D FD	**0.000025**	**250.32**	**15.91**	7.15	**19.35**	**24.80**	**54.57**	**627.47**	**25.05**
D2D HD	0.00017	340.85	16.10	7.38	19.61	25.07	54.83	643.84	25.37
RIS	0.0159	666.71	34.34	10.43	51.12	47.85	138.66	3792.6	61.58
Cellular+D2D FD	0.0255	477.32	21.77	11.66	27.09	30.30	76.12	1207.3	34.75
Cellular+D2D HD	0.0327	466.58	21.81	11.86	27.04	30.29	75.81	1207.1	34.74
Cellular+RIS	0.039	487.98	32.52	18.36	37.65	49.74	106.23	2475.3	49.75
Cell.+D2D FD+RIS	0.0123	327.66	23.39	15.42	24.41	33.69	71.59	1142.9	33.81
Cell.+D2D HD+RIS	0.0223	325.10	23.40	15.40	24.41	33.80	71.31	1143.4	33.81
(d) NLOS, PLE=3.19, SF=7.82. D2D relays are uniformly distributed within 100 m×100 m **square**									
Configuration	**min**	**max**	**mean**	**median**	**std**	**75thprc**	**95thprc**	**MSE**	**RMSE**
Cellular	0.0174	1056.4	47.52	6.85	80.25	70.66	200.65	8698.8	93.27
D2D FD	0.0236	1095.6	35.77	21.24	46.90	44.34	118.14	3478.9	58.98
D2D HD	0.0225	1472.6	35.96	21.28	47.45	44.62	120.34	3545.0	59.54
RIS	**0.0074**	571.01	34.75	10.35	51.03	49.03	140.99	3811.7	61.74
Cellular+D2D FD	0.1104	550.16	35.87	20.29	43.10	48.96	116.42	3145.2	56.08
Cellular+D2D HD	0.0415	736.62	36.00	20.79	43.14	49.62	116.45	3156.5	56.18
Cellular+RIS	0.0091	598.68	39.11	22.75	47.19	58.42	1129.59	3756.8	61.29
Cell.+D2D FD+RIS	0.0653	413.50	32.18	22.17	33.04	45.01	95.71	2127.0	46.12
Cell.+D2D HD+RIS	0.0204	504.34	32.29	22.18	33.07	45.56	94.69	2136.5	46.22
Cell.+D2D FD+RIS*	0.0135	**357.33**	**18.87**	**6.79**	**28.82**	**20.83**	**80.63**	**1186.8**	**34.45**
Cell.+D2D HD+RIS*	0.0135	**357.33**	**18.97**	6.80	28.97	20.95	81.08	1198.7	34.62
* means that for this configuration we use the median for position-level sensor fusion among the three technologies									

improvement in D2D localization positioning accuracy, which is explained by the fact that D2D devices are located much closer to the node with unknown coordinates, while positions of BSs and RISs are fixed (for illustration, see left part of Fig. 4.4). Moreover, we should emphasize that, differently from the first two scenarios, in this case, pure D2D localization has the highest accuracy level, outperforming even all possible combinations of technologies. The reason is that D2D by itself reports approximately twice as better results compared to cellular and RIS positioning schemes. Hence, in this case, the errors of other technologies can only deteriorate D2D positioning. Alternatively, when all schemes reveal similar performance, their errors could be balanced via the integration of all technologies (first two scenarios). In addition, there is no significant difference between the two D2D schemes (i.e., Half-Duplex (HD) and FD) at variance with the first two deployments due to the shortened distances between MTs.

NLoS: Random Placement of Relays

We proceed by investigating the positioning precision using different technologies in the case of the UMi Street Canyon NLoS scenario with PLE $\zeta = 3.19$ and $\sigma = 7.82$, as shown in Fig. 4.7(d) and Table 4.9(d). We note that RISs are always in LOS conditions since they are designed to provide a virtual LOS path when a direct BS-UE path is blocked [196]. Here, compared to LoS scenario in Fig. 4.7(b), one can observe that RIS-aided positioning outperforms relays in terms of accuracy due to the obstacle on their paths. Then, the configuration "Cellular+D2D FD+RIS" again can balance the errors and provide the highest accuracy among all considered configurations.

Further, in this use case, we introduced the median value for sensor fusion when dealing with the combination of the three technologies. Note that since we work with location-level fusion, there is no sense in applying median value in the case of the two estimated positions. However, using the median for "Cellular+D2D FD+RIS" configuration helps to reduce the localization mean error by 41.36%. Hence, we might conclude that *the optimal sensor fusion metric that minimizes the mean absolute error is the median of a set of estimated locations.* Note that, in this case, the median fusion approach works better because one out of three localization methods fails (cellular positioning), and the two others show a similar level of accuracy. The median operates well in the presence of outliers (e.g., in the case of cellular positioning). In the case of mean sensor fusion, cellular positioning accuracy penalizes the combination of the three methods.

Effects of the Number of RIS Elements

Finally, Fig. 4.8 illustrates the influence of the number of RIS elements on the localization (Fig. 4.8(a)) and communication (Fig. 4.8(b)) performance. One may observe that even though the number of antenna elements impacts the communication channel, i.e., SNR of the link, it does not influence the positioning accuracy. The reason might be that the positions

and the channel propagation model of RISs are known and do not change when varying the number of RIS elements. In addition, as we noticed from previous simulation results, the factor which genuinely impacts the positioning accuracy is the anchors' locations.

Summary

By analyzing the results for all considered scenarios, we highlight that when one of the compared technologies (i.e., D2D or RIS) significantly outperforms either the second one or cellular positioning, the technology by itself provides the best accuracy (see Fig. 4.7(c) and Table 4.9(c)). Alternatively, when the difference between D2D and RIS or D2D/RIS and cellular positioning is not substantial, the combination of all three technologies reveals the best configuration in terms of accuracy.

Previous works have identified noise as a primary source of positioning errors. In our simulations, the white noise injected into simulated D2D, RIS, and cellular data is completely independent (not cross-correlated), making the positioning errors also wholly independent. Therefore, combining those three positioning methods gives us this feeling of better results. However, so far, no experimental studies have demonstrated that noise present in D2D, RIS, and cellular measurements is cross-correlated. Intuitively, those are measurements obtained using different technologies, which might lead to non-cross-correlated noise. Nevertheless, future experimental studies are required to confirm the correctness of our assumption, and, in case it is not validated, the feeling of better results when combining all the technologies should be neglected. Moreover, the location of the anchors affects the positioning performance. For example, the highest accuracy yields the case when D2D relays are located in close proximity to the MT with unknown coordinates (see Fig. 4.7(c) and Table 4.9(c)).

We have selected the use cases in such a way as to show a detailed picture of the performance of the positioning methods compared in this work. From the study emerges that there is only a slight improvement of cellular positioning when applying RIS for LoS environment. In the case of NLoS, RIS substantially contributes to improving the cellular localization accuracy, although still the three methods together work better. Then, in the case of D2D, the location of devices defines the positioning accuracy, in both LoS and NLoS propagation environments. Last, the median appears as a parameter that significantly improves the positioning error when applied to sensor fusion.

4.3 Conclusions

This chapter introduced the concept of cooperative positioning. More precisely, we first provided a survey on the state-of-the-art of D2D-aided cooperative positioning, which can be determined as the information exchange among MTs intending to increase their localization accuracy. In summary, 5G system positioning capabilities have been enhanced compared to previous generations of communication networks, especially thanks to the D2D technology

allowing to perform cooperative positioning in cellular systems. It has already been demonstrated that in such systems, the positioning accuracy of 1 m and below can be reached, thereby satisfying 5G application requirements. More importantly, it has the visible potential to offer centimeter-level accuracy. Nevertheless, there still remain plenty of challenges to be solved by the research community. Among them, the need for efficient protocols, positioning complexity, and green communication issues, among others, have to be addressed.

In the second section of this chapter, we provided the model for D2D- and RIS-aided cellular positioning systems. Then, we drew comparisons between them regarding communication aspects and, more importantly, in terms of positioning precision. Throughout the simulation campaign, we identified the use cases where one technology outperforms the other, as well as presented the localization performance of the D2D-, RIS-assisted cellular localization. In addition, we discussed the role of mean and median metrics in localization systems.

The conclusions of this study are as follows. First, RIS-aided localization serves as a reasonable means when the distribution of D2D anchors is scarce. Alternately, relaying works significantly better than RIS positioning due to the short distances. Moreover, we highlight that when the improvement of RIS and D2D over cellular positioning is mediocre (less than 50%), the combination of RIS, D2D, and cellular technologies contributes to the lowest positioning errors. Further, when one of the technologies shows a significant improvement over cellular positioning, this technology by itself shows the best performance among all configurations.

Moreover, the use of the median for position-level sensor fusion in case of the combination of the locations obtained from the three methods helps to reduce the positioning error by 41% and serves as a practical approach to minimize the mean absolute error. Finally, we note that the number of RIS reflective elements does not affect the positioning accuracy.

The published materials from this chapter can be found in [202].

5

Conclusions

We conclude this thesis with a summary of the main research outcomes and present future research avenues. For more concise summaries, see Conclusions in each chapter of this thesis.

5.1 Summary

In this thesis, a set of mathematical frameworks, including optimal solutions, heuristics, machine learning algorithms, and simulation tools, has been developed to characterize and support multicast traffic delivery in mmWave directional systems. Then, sidelink relaying concept was introduced to deal with the channel condition deterioration of dynamic multicast systems and to ensure both mission-critical and ultra-reliable low-latency communications. It is shown that sidelink relaying helps *(i)* improve the performance of mmWave multicast directional systems in the presence of dynamic users/blockers and *(ii)* satisfy the strict requirements of public safety and factory automation use cases.

Furthermore, cooperative positioning techniques for enhancing cellular positioning accuracy for 5G+ emerging applications that require not only improved communication characteristics but also precise localization have been presented and analyzed. As showed the results, cooperative positioning methods by means of D2D and RIS technologies improve the accuracy and coverage of cellular networks. Further, we proposed a loose-coupling sensor fusion based on a statistical metric to fuse the location information obtained through different methods.

The study of this thesis has led to the following important conclusions:

- The capabilities of modern antenna arrays to utilize multiple beams simultaneously with potentially varying half-power beamwidth and asymmetric power allocation pose the problem of efficient multicast transmission in mmWave NR systems.
- With the optimal multicast scheduling strategies, both the system resource utilization (network side) and perceived user experience (user side) can be improved.
- Sidelink relaying serves as a means to improve the multicast users' performance in a dynamic system by relaying the data between peers in proximity and surpassing base stations.

- NR sidelink is anticipated to become an essential technology to ensure both mission-critical and ultra-reliable low-latency communications.
- D2D-aided cooperative positioning has the visible potential to offer centimeter-level accuracy.
- RIS-aided cooperative localization serves as a reasonable means when the distribution of D2D anchors is scarce. Alternatively, sidelink relaying works significantly better than RIS positioning due to the short distances.

5.2 Future Research

In this section, we discuss the topics that are reserved for future work regarding *(i)* multicasting in mmWave, *(ii)* sidelink relaying standardization, and *(iii)* D2D-based cooperative positioning aspects that have to be further studied.

5.2.1 mmWave Multicasting

An in-depth review of additional mechanisms that can be utilized to further improve multicasting performance in 5G/6G mmWave/sub-THz systems are discussed below.

Fair Coexistence Between Unicast and Multicast Traffic. Although there has been a sizable amount of research on the provision of multicast services in broadband wireless access networks, very few of these studies offer solutions for the simultaneous management of unicast and multicast traffic. Due to the fact that these two types of traffic will undoubtedly coexist in future mobile communication systems, it is imperative to understand their unique properties to ensure fair resource allocation, as discussed in the following.

The specifics of the multicast service operation indirectly introduce priority for the multicast sessions, thereby severely decreasing the unicast session loss probability. As the offered load for the multicast sessions increases, the system fills up nearly entirely with them, leaving the unicast sessions with minimal remaining resources. One must actively prioritize the unicast traffic using bandwidth reservation and connection admission control techniques, among others, to balance out the session drop possibilities. Note that some work on fair multicast and unicast traffic management has been done for LTE systems in [87] and in a review of related studies thereof. However, solutions designed for omnidirectional LTE systems are not suitable for 5G/6G mmWave/sub-THz systems based on directional transmissions. Hence, there is an urgent need to fill this gap.

Reliability Improvements via New Mechanisms. Being inherently prone to outage events due to blockage and micromobility, 5G/6G mmWave/sub-THz bands pose extreme challenges for the provisioning of reliable multicast service. As the main tool to improve session service reliability, 3GPP offers inter- and intra-RAT multi-connectivity operations. However, by utilizing even extremely density deployments on mmWave BS, no sufficient service reliability can be achieved. On the other hand, the use of inter-RAT multi-connectivity

with, e.g., LTE or μWave NR, leads to considerable performance degradation of single-band μWave UEs. However, the observations above are extrapolated from those obtained for unicast services, and there are still no in-depth mathematical frameworks benchmarking performance improvements of these functionalities for multicast services.

There is an urgent need for new advanced mechanisms to improve service reliability in 5G/6G mmWave/sub-THz systems, such as the use of IAB deployments, RISs, and NR sidelink technologies. There are no full comprehensive frameworks allowing for comparing performance of different solutions and algorithms.

5.2.2 Sidelink Relaying

Sidelink technology was standardized for the first time during LTE 3GPP Release 12. However, due to the uncertainty on whether such a technology would have been of interest to the major mobile operators, the use cases that sidelink was supposed to handle were only confined to public safety and V2X. Besides, only a simple set of features ended up in the LTE specification, most of which actually needed to be pre-configured in the UE sim card. Note that pre-configuration is for out-of-coverage operations only, whereas the usual SIB/RRC mechanisms are used for operations in coverage.

However, with the multitude of use cases that 5G NR is expected to support, the sidelink technology again gained momentum among industry and mobile operators and is now considered one of the killer technology to guarantee low delay, extended coverage, and improved energy efficiency to the UEs. Most importantly, the use cases that sidelink is expected to handle are not only confined to public safety and V2X (i.e., as for LTE), but they span from unlicensed applications, IIoT, up to UAV. This is also becoming evident in view of the coming 3GPP Release 18, where vertical technologies and applications are interacting with each other in order to provide the connection to “anything”, “everywhere”, and “anytime” (one of the bases of 5G technologies). We discuss the future sidelink directions that have to be investigated.

Sidelink Relaying. One of the main goals of sidelink relaying is the coverage area extension of both sidelink communication and cellular network. Moreover, energy efficiency and enhanced QoS support are additional essential features. As defined in [203], there are two types of relaying that can be studied: *(i)* UE-to-UE and *(ii)* network-to-UE. The former aims to extend the coverage of the cell through a relay, thereby providing the service for UEs located at the edge or out of the coverage of the cell. The latter means that not only a single-hop relay (supported by Rel-17) can be performed. In this case, multiple relays (multi-hop relaying is currently not supported) can extend the sidelink coverage, but more work needs to be done. The main aspects that have to be studied regarding relaying are relay (re-)selection, relay discovery, UE authorization, QoS provisioning, among others, which is becoming more complex in the case of multi-hop relaying.

Sidelink Positioning. One of the missing functionalities in Rel-17 is sidelink positioning, i.e., to satisfy the strict requirements for absolute positioning. Here, a study on positioning in Rel-17 can be considered as an initial point, and it is expected that sidelink positioning will be one of the main work items in Rel-18.

With the growing complexity of indoor and outdoor environments, the radio propagation between transmitting and receiving devices becomes increasingly complex, especially with dynamic blockers. Although the LoS paths between each pair of target and anchor nodes typically exist in the considered industrial/public safety scenarios, different environmental objects can block the LoS paths, turning them into challenging NLoS scenarios. Thus, novel positioning, tracking, and mapping frameworks employing both the multipath components and the relay paths from intelligent surfaces are seen to form an intriguing open research space to synthesize the location and environmental awareness towards an intelligent positioning and mapping system.

Sidelink and Artificial Intelligence. Artificial Intelligence (AI) and ML can improve sidelink further communication. For instance, ML algorithms can perform resource allocation with the quality of the radio channel, road traffic conditions, among other input algorithm parameters. This, in turn, will reduce the overall latency and the throughput for future sidelink communications.

In conclusion, sidelink as a technology is continually evolving, and it is now evident that it will be one of the pillars not only in the further development of the 5G system but will also be at the center of the close-to-come 6G technology that will be the first standardization in 3GPP (hopefully) during 2023.

5.2.3 Cooperative Positioning

This subsection discusses further aspects related to D2D-based cooperative positioning that still need future investigations by the research and industry communities.

Green Communications. There is a rising interest in developing more energy-efficient protocols and algorithms for a D2D-based collaborative positioning, which can reduce the economic and environmental impact and promote Green Communications. Since, in 5G communication systems, the use of energy and spectrum resources plays a significant role, green communications represent a very timely topic. Also, adaptive power control should be analyzed in order to gain a more detailed understanding, e.g., on the effect of transmissions. However, there are some implementation difficulties to cope with, e.g., green communications pose a notable challenge to sustainable network development in the case of industrial applications.

Mobility. In many cases, due to the portability of the nodes, mobility becomes an issue to be considered in the design of these communication protocols. Also, mobility makes localization techniques increasingly less accurate, and these errors usually increase with node speed. Moreover, the impact that anchors' and agents' mobility, needing frequent handovers

in a Multi-RAT network, may have on energy consumption has still received limited attention from the research community. On the other side, some studies prove that positioning accuracy could be improved by leveraging the mobility capabilities of the nodes depending on the level of mobility of the nodes within the network.

Utilization of the Beam Training Period for Positioning. Since highly directional transmissions (e.g., by using mmWave and terahertz bands) are considered as one of the main components of 5G and B5G systems, the feasibility of utilizing the beam training period for positioning is of extreme interest.

Synchronization. Network synchronization is a critical aspect of cellular ranging-based location methods, such as OTDOA/UTDOA in LTE, which should not exceed the order of nanoseconds for accurate positioning. Thus, to support ranging-based methods, cellular standards should take into account tight network synchronization requirements. Indoors, e.g., the use of advanced network time protocols or accurate round-trip time should be considered. Moreover, the chosen waveform will impact the synchronization requirements, potentially gaining localization accuracy.

LTE sidelink communication can handle the synchronization problem between MTs in distributed cooperative localization systems through the primary sidelink synchronization signal. However, it can be managed only for those MTs that are under the eNodeBs coverage, whereas synchronization is a challenging problem when MTs are out-of-network coverage.

Higher carrier frequencies increase the resolution of multipaths. Higher carrier frequencies, such as mmWave spectrum, increase the LoS reception probability since any NLoS condition is likely to be blocked due to the severe penetration and propagation properties. Consequently, the channel becomes more sparse in the sense of few dominant multipath components and very few reflected paths. A sparse channel means that it is easier to identify individual specular multipath components that can be used for high-accurate positioning by reducing the risk of positioning errors due to the NLoS bias.

References

1. N. Chukhno, O. Chukhno, D. Moltchanov, A. Molinaro, *et al.*, "Optimal Multicasting in Millimeter Wave 5G NR with Multi-beam Directional Antennas," *IEEE Trans. on Mobile Computing (Early Access)*, 2021.
2. N. Chukhno, O. Chukhno, D. Moltchanov, A. Gaydamaka, A. Samuylov, A. Molinaro, Y. Koucheryavy, A. Iera, and G. Araniti, "The Use of Machine Learning Techniques for Optimal Multicasting in 5G NR Systems," *IEEE Transactions on Broadcasting*, 2022.
3. 3GPP, "NR; Physical channels and modulation (Release 15)," 3GPP TR 38.211, Dec 2017.
4. 3GPP, "Study on Positioning Use Cases (Release 16)," 3GPP TR 22.872 V16.1.0 , September 2018.
5. 3GPP, "Service Requirements for the 5G System (Release 18)," tech. rep., 3GPP TS 22.261 V18.2.0, March 2021.
6. 3GPP, "Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 14)," 3GPP TR 38.901 V14.1.1, July 2017.
7. N. Docomo, "5G Channel Model for Bands up to 100 GHz," tech. rep., Tech. Report, Oct, 2016.
8. M. Peter, K. Haneda, S. Nguyen, A. Karttunen, and J. Järveläinen, "Measurement Results and Final mmMAGIC Channel Models," *Deliverable D2*, vol. 2, 2017.
9. G. Sanfilippo, O. Galinina, S. Andreev, S. Pizzi, and G. Araniti, "A Concise Review of 5G New Radio Capabilities for Directional Access at mmWave Frequencies," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, pp. 340–354, Springer, 2018.
10. 3GPP, "NG-RAN; Xn Application Protocol (XnAP)," 3GPP TR 38.423 (Draft), June 2018.
11. S. Pizzi, C. Suraci, A. Iera, A. Molinaro, and G. Araniti, "A sidelink-aided approach for secure multicast service delivery: From human-oriented multimedia traffic to machine type communications," *IEEE Transactions on Broadcasting*, vol. 67, no. 1, pp. 313–323, 2020.

12. N. Chukhno, O. Chukhno, S. Pizzi, A. Molinaro, A. Iera, and G. Araniti, "Efficient management of multicast traffic in directional mmwave networks," *IEEE Transactions on Broadcasting*, vol. 67, no. 3, pp. 593–605, 2021.
13. Y. M. Tsang and A. S. Poon, "Detecting Human Blockage and Device Movement in mmWave Communication System," in *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011*, pp. 1–6, IEEE, 2011.
14. G. R. MacCartney, T. S. Rappaport, and S. Rangan, "Rapid Fading Due to Human Blockage in Pedestrian Crowds at 5G Millimeter-Wave Frequencies," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–7, IEEE, 2017.
15. G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter-Wave Human Blockage at 73 GHz With a Simple Double Knife-Edge Diffraction Model and Extension for Directional Antennas," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pp. 1–6, IEEE, 2016.
16. A. Biason and M. Zorzi, "Multicast via Point to Multipoint Transmissions in Directional 5G mmWave Communications," *IEEE Communications Magazine*, vol. 57, no. 2, pp. 88–94, 2019.
17. F. Rinaldi, S. Pizzi, A. Orsino, A. Iera, A. Molinaro, and G. Araniti, "A Novel Approach for MBSFN Area Formation Aided by D2D Communications for eMBB Service Delivery in 5G NR Systems," *IEEE Trans. on Vehicular Technology*, vol. 69, no. 2, pp. 2058–2070, 2019.
18. P. Mach and Z. Becvar, "Device-to-Device Relaying: Optimization, Performance Perspectives, and Open Challenges Towards 6G Networks," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1336–1393, 2022.
19. G. H. Sim, A. Loch, A. Asadi, V. Mancuso, and J. Widmer, "5G Millimeter-Wave and D2D Symbiosis: 60 GHz for Proximity-based Services," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 140–145, 2017.
20. Y. Niu, L. Yu, Y. Li, Z. Zhong, B. Ai, *et al.*, "Device-to-Device Communications Enabled Multicast Scheduling with the Multi-Level Codebook in mmWave Small Cells," *Mobile Networks and Applications*, vol. 24, no. 5, pp. 1603–1617, 2019.
21. S. Zhang, D. Liu, J. Lv, and Z. Zhang, "D2D-enabled Multicast Optimal Scheduling in mmWave Cellular Networks," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 442–447, IEEE, 2020.
22. Y. Niu, Y. Liu, Y. Li, X. Chen, Z. Zhong, and Z. Han, "Device-to-Device Communications Enabled Energy Efficient Multicast Scheduling in mmWave Small Cells," *IEEE Trans. on Communications*, vol. 66, no. 3, pp. 1093–1109, 2017.
23. Y. Niu, L. Yu, Y. Li, Z. Zhong, and B. Ai, "Device-to-Device Communications Enabled Multicast Scheduling for mmWave Small Cells using Multi-Level Codebooks," *IEEE Trans. on Vehicular Technology*, vol. 68, no. 3, pp. 2724–2738, 2018.

24. G. Hong, M. Mousavi, L. Wang, A. Klein, M. Hollick, *et al.*, "Joint Relaying and Spatial Sharing Multicast Scheduling for mmWave Networks," *arXiv preprint arXiv:1907.13085*, 2019.

25. A. Orsino, D. Moltchanov, M. Gapeyenko, A. Samuylov, S. Andreev, L. Militano, G. Araniti, and Y. Koucheryavy, "Direct Connection on the Move: Characterization of User Mobility in Cellular-Assisted D2D Systems," *IEEE Vehicular Technology Magazine*, vol. 11, no. 3, pp. 38–48, 2016.

26. A. Orsino, A. Ometov, G. Fodor, D. Moltchanov, L. Militano, S. Andreev, O. N. Yilmaz, T. Tirronen, J. Torsner, G. Araniti, *et al.*, "Effects of Heterogeneous Mobility on D2D-and Drone-Assisted Mission-Critical MTC in 5G," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 79–87, 2017.

27. Y. Geng, E. Liu, R. Wang, and Y. Liu, "Hierarchical Reinforcement Learning for Relay Selection and Power Optimization in Two-Hop Cooperative Relay Network," *IEEE Transactions on Communications*, vol. 70, no. 1, pp. 171–184, 2021.

28. Y. Gu, Q. Wang, H. Chen, Y. Li, and B. Vucetic, "Optimizing Information Freshness in Two-Hop Status Update Systems Under a Resource Constraint," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1380–1392, 2021.

29. O. Kanhere, S. Ju, Y. Xing, and T. S. Rappaport, "Map-Assisted Millimeter Wave Localization for Accurate Position Location," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2019.

30. S. Wu, S. Zhang, and D. Huang, "A TOA-based Localization Algorithm With Simultaneous NLOS Mitigation and Synchronization Error Elimination," *IEEE Sensors Letters*, vol. 3, no. 3, pp. 1–4, 2019.

31. W. Wang, Y. Zhang, and L. Tian, "TOA-based NLOS Error Mitigation Algorithm for 3D Indoor Localization," *China Communications*, vol. 17, no. 1, pp. 63–72, 2020.

32. Cisco VNI Forecast, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022," 2019.

33. 3GPP, "Release 15 Description; Summary of Rel-15 Work Items (Rel. 15)," 3GPP TR 21.915 V15.0.0, Oct 2019.

34. 3GPP, "Release 16 Description; Summary of Rel-16 Work Items (Rel. 16)," 3GPP TR 21.916 V0.6.0, Sept 2020.

35. 3GPP, "Study on Architectural Enhancements for 5G Multicast-Broadcast Services (Rel. 17)," TR 23.757 V1.2.0, November 2020.

36. O. Vikhrova, S. Pizzi, A. Iera, A. Molinaro, K. Samuylov, and G. Araniti, "Performance Analysis of Paging Strategies and Data Delivery Approaches for Supporting Group-Oriented IoT Traffic in 5G Networks," in *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–5, IEEE, 2019.

37. A. Samuylov, D. Moltchanov, R. Kovalchukov, R. Pirmagomedov, Y. Gaidamaka, S. Andreev, Y. Koucheryavy, and K. Samouylov, "Characterizing resource allocation trade-offs

in 5g nr serving multicast and unicast traffic," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3421–3434, 2020.

38. S. Ahmadi, *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards*. Academic Press, 2019.
39. M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A Tutorial on Beam Management for 3GPP NR at mmWave Frequencies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 173–196, 2018.
40. E. Garro, M. Fuentes, J. Carcel, H. Chen, D. Mi, F. Tesema, J. Gimenez, and D. Gomez-Barquero, "5G Mixed Mode: NR Multicast-Broadcast Services," *IEEE Transactions on broadcasting*, vol. 66, no. 2 Part II, 2020.
41. G. R. MacCartney, J. Zhang, S. Nie, and T. S. Rappaport, "Path Loss Models for 5G Millimeter Wave Propagation Channels in Urban Microcells," in *2013 IEEE Global Communications Conference (GLOBECOM)*, pp. 3948–3953, IEEE, 2013.
42. M. Gapeyenko, A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, E. Aryafar, S.-p. Yeh, N. Himayat, S. Andreev, and Y. Koucheryavy, "Analysis of Human-Body Blockage in Urban Millimeter-Wave Cellular Communications," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–7, IEEE, 2016.
43. A. Biason and M. Zorzi, "Multicast transmissions in directional mmwave communications," in *European Wireless 2017; 23th European Wireless Conference*, pp. 1–7, VDE, 2017.
44. H. Park and C.-H. Kang, "A Group-aware Multicast Scheme in 60GHz WLANs," *TIIS*, vol. 5, no. 5, pp. 1028–1048, 2011.
45. H. Park, S. Park, T. Song, and S. Pack, "An incremental multicast grouping scheme for mmWave networks with directional antennas," *IEEE Communications Letters*, vol. 17, no. 3, pp. 616–619, 2013.
46. K. Sundaresan, K. Ramachandran, and S. Rangarajan, "Optimal beam scheduling for multicasting in wireless networks," in *Proceedings of the 15th annual international conference on Mobile computing and networking*, pp. 205–216, 2009.
47. E. Aryafar, M. A. Khojastepour, K. Sundaresan, S. Rangarajan, and E. Knightly, "ADAM: An Adaptive Beamforming System for Multicasting in Wireless LANs," *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1595–1608, 2013.
48. 3GPP, "Radio Frequency (RF) requirements for LTE Pico Node B (Release 12)," 3GPP TR 36.911 V12.0.0, 3GPP, Sep. 2012.
49. R. Kovalchukov, D. Moltchanov, A. Pyattaev, and A. Ometov, "Evaluating multi-connectivity in 5G NR systems with mixture of unicast and multicast traffic," in *International Conference on Wired/Wireless Internet Communication*, pp. 118–128, Springer, 2019.
50. A. B. Constantine, *Antenna Theory: Analysis and Design*. Wiley-Interscience, 2005.
51. V. Petrov, M. Komarov, D. Moltchanov, J. M. Jornet, and Y. Koucheryavy, "Interference and SINR in Millimeter Wave and Terahertz Communication Systems With Blocking and

Directional Antennas," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1791–1808, 2017.

52. V. Begishev, D. Moltchanov, E. Sopin, A. Samuylov, S. Andreev, Y. Koucheryavy, and K. Samouylov, "Quantifying the impact of guard capacity on session continuity in 3GPP new radio systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12345–12359, 2019.
53. C. A. Balanis, *Antenna Theory: Analysis and Design*. John wiley & sons, 2015.
54. 3GPP, "Study on Indoor Positioning Enhancements for UTRA and LTE (Release 13)," 3GPP TR 37.857 V13.1.0, January 2016.
55. L. Rayleigh, "XII. On the Resultant of a Large Number of Vibrations of the Same Pitch and of Arbitrary Phase," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 10, no. 60, pp. 73–78, 1880.
56. S. O. Rice, "Statistical Properties of a Sine Wave Plus Random Noise," *The Bell System Technical Journal*, vol. 27, no. 1, pp. 109–157, 1948.
57. M. Nakagami, "The m-Distribution—A General Formula of Intensity Distribution of Rapid Fading," in *Statistical methods in radio wave propagation*, pp. 3–36, Elsevier, 1960.
58. R. S. Hoyt, "Probability Functions for the Modulus and Angle of the Normal Complex Variate," *The Bell System Technical Journal*, vol. 26, no. 2, pp. 318–359, 1947.
59. M. Taneda, "A New Approach to Fading: Weibull Model," in *Proc. Int. Symp. Personal, Indoor, Mobile Radio Commun.*, pp. 711–715, 1999.
60. M. K. Simon and M.-S. Alouini, *Digital Communication Over Fading Channels*, vol. 95. John Wiley & Sons, 2005.
61. R. Series, "Propagation Data and Prediction Methods Required for the Design of Terrestrial Line-of-Sight Systems," *Recommendation ITU-R*, pp. 530–12, 2015.
62. S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, Inc., 1990.
63. T. G. Crainic, F. D. Fomeni, and W. Rei, *The Multi-Period Variable Cost and Size Bin Packing Problem with Assignment Cost: Efficient Constructive Heuristics*. CIRRELT, 2019.
64. J. Fan, Q. Yin, G. Y. Li, *et al.*, "MCS Selection for Throughput Improvement in Downlink LTE Systems," in *2011 Proceedings of 20th international conference on computer communications and networks (ICCCN)*, pp. 1–5, IEEE, 2011.
65. ITU-T Rec. Y.3172, "Architectural Framework for Machine Learning in Future Networks Including IMT-2020," *ITU-T*, 2020.
66. ITU-T Rec. Y.3174, "Framework for Data Handling to Enable Machine Learning in Future Networks Including IMT-2020," *ITU-T*, 2020.
67. ITU-T Rec. Y.3176, "Machine Learning Marketplace Integration in Future Networks Including IMT-2020," *ITU-T*, 2020.
68. A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. " O'Reilly Media, Inc.", 2017.

69. M. Merenda, C. Porcaro, and D. Iero, "Edge Machine Learning for AI-enabled IoT devices: A Review," *Sensors*, vol. 20, no. 9, p. 2533, 2020.

70. M. Fischetti and A. Lodi, "Local Branching," *Mathematical programming*, vol. 98, no. 1, pp. 23–47, 2003.

71. E. Danna, E. Rothberg, and C. Le Pape, "Exploring Relaxation Induced Neighborhoods to Improve MIP Solutions," *Mathematical Programming*, vol. 102, no. 1, pp. 71–90, 2005.

72. Y. Fu and A. Banerjee, "Heuristic/Meta-Heuristic Methods for Restricted Bin Packing Problem," *Journal of Heuristics*, vol. 26, no. 5, pp. 637–662, 2020.

73. M.-T. Vakil-Baghmisheh and A. Navarbaf, "A Modified Very Fast Simulated Annealing Algorithm," in *2008 International Symposium on Telecommunications*, pp. 61–66, IEEE, 2008.

74. E. H. Aarts and J. H. Korst, "Boltzmann Machines as a Model for Parallel Annealing," *Algorithmica*, vol. 6, no. 1, pp. 437–465, 1991.

75. M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks*. Elsevier, 2004.

76. M. Eusuff, K. Lansey, and F. Pasha, "Shuffled Frog-Leaping Algorithm: A Memetic Meta-Heuristic for Discrete Optimization," *Engineering optimization*, vol. 38, no. 2, pp. 129–154, 2006.

77. C. Gallo and V. Capozzi, "A Simulated Annealing Algorithm for Scheduling Problems," *Journal of Applied Mathematics and Physics*, vol. 7, no. 11, pp. 2579–2594, 2019.

78. O. Hasançebi, S. Çarbaş, and M. P. Saka, "Improving the Performance of Simulated Annealing in Structural Optimization," *Structural and Multidisciplinary Optimization*, vol. 41, no. 2, pp. 189–203, 2010.

79. A. Schnack, M. J. Wright, and J. L. Holdershaw, "Immersive Virtual Reality Technology in a Three-Dimensional Virtual Simulated Store: Investigating Telepresence and Usability," *Food Research International*, vol. 117, pp. 40–49, 2019.

80. S. Zhang, X. Cai, W. Zhou, and Y. Wang, "Green 5G Enabling Technologies: an Overview," *IET Communications*, vol. 13, no. 2, pp. 135–143, 2019.

81. N. Chukhno, O. Chukhno, S. Pizzi, A. Molinaro, A. Iera, and G. Araniti, "Unsupervised Learning for D2D-Assisted Multicast Scheduling in mmWave Networks," in *2021 IEEE BMSB*, pp. 1–6, IEEE, 2021.

82. X. Wang, Y. Xu, J. Chen, C. Li, X. Liu, D. Liu, and Y. Xu, "Mean Field Reinforcement Learning based Anti-Jamming Communications for Ultra-Dense Internet of Ihings in 6G," in *2020 International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 195–200, IEEE, 2020.

83. H. Wei, G. Zheng, V. Gayah, and Z. Li, "Recent Advances in Reinforcement Learning for Traffic Signal Control: A Survey of Models and Evaluation," *ACM SIGKDD Explorations Newsletter*, vol. 22, no. 2, pp. 12–18, 2021.

84. D. Lee, N. He, P. Kamalaruban, and V. Cevher, "Optimization for Reinforcement Learning: From a Single Agent to Cooperative Agents," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 123–135, 2020.

85. H. Liu and W. Wu, "Federated Reinforcement Learning for Decentralized Voltage Control in Distribution Networks," *IEEE Transactions on Smart Grid*, 2022.

86. L. Feng, Z. Yang, Y. Yang, X. Que, and K. Zhang, "Smart Mode Selection Using Online Reinforcement Learning for VR Broadband Broadcasting in D2D Assisted 5G HetNets," *IEEE Trans. on Broadcasting*, vol. 66, no. 2, pp. 600–611, 2020.

87. S. Pizzi, M. Condoluci, G. Araniti, A. Molinaro, A. Iera, and G.-M. Muntean, "A Unified Approach for Efficient Delivery of Unicast and Multicast Wireless Video Services," *IEEE Trans. on Wireless Comm.*, vol. 15, no. 12, pp. 8063–8076, 2016.

88. O. Chukhno, N. Chukhno, O. Galinina, Y. Gaidamaka, S. Andreev, and K. Samouylov, "Analysis of 3D Deafness Effects in Highly Directional mmWave Communications," in *2019 IEEE GLOBECOM*, pp. 1–6, IEEE, 2019.

89. O. Chukhno, N. Chukhno, O. Galinina, S. Andreev, Y. Gaidamaka, K. Samouylov, and G. Araniti, "A Holistic Assessment of Directional Deafness in mmWave-based Distributed 3D Networks," *IEEE Trans. on Wireless Communications (Early Access)*, 2022.

90. D. Helbing and P. Molnar, "Social Force Model for Pedestrian Dynamics," *Physical review E*, vol. 51, no. 5, p. 4282, 1995.

91. F. Farina, D. Fontanelli, A. Garulli, A. Giannitrapani, and D. Prattichizzo, "Walking ahead: The headed social force model," *PloS one*, vol. 12, no. 1, p. e0169734, 2017.

92. F. Nielsen, "Hierarchical Clustering," in *Introduction to HPC with MPI for Data Science*, pp. 195–211, Springer, 2016.

93. D. Müllner, "Modern Hierarchical, Agglomerative Clustering Algorithms," *arXiv preprint arXiv:1109.2378*, 2011.

94. D. Lopez-Perez, I. Guvenc, G. De la Roche, M. Kountouris, T. Q. Quek, and J. Zhang, "Enhanced Intercell Interference Coordination Challenges in Heterogeneous Networks," *IEEE Wireless communications*, vol. 18, no. 3, pp. 22–30, 2011.

95. S. Akoum, O. El Ayach, and R. W. Heath, "Coverage and capacity in mmWave cellular systems," in *2012 conference record of the forty sixth Asilomar conference on signals, systems and computers (ASILOMAR)*, pp. 688–692, IEEE, 2012.

96. S. A. Ashraf, R. Blasco, H. Do, G. Fodor, C. Zhang, and W. Sun, "Supporting Vehicle-to-Everything Services by 5G New Radio Release-16 Systems," *IEEE Communications Standards Magazine*, vol. 4, no. 1, pp. 26–32, 2020.

97. X. Wang and S. Akoum, "Forward Compatible New Radio Sidelink Slot Format Signalling," Dec. 22 2020. US Patent 10,873,944.

98. S. Saafi, J. Hosek, and A. Kolackova, "Enabling Next-Generation Public Safety Operations With Mission-Critical Networks and Wearable Applications," *Sensors*, vol. 21, no. 17, p. 5790, 2021.

99. A. Yarali, *Public safety networks from LTE to 5G*. John Wiley & Sons, 2020.

100. A. Othman and N. A. Nayan, "Public Safety Mobile Broadband System: From Shared Network to Logically Dedicated Approach Leveraging 5G Network Slicing," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2109–2120, 2020.

101. M. Höyhtyä, O. Apilo, and M. Lasanen, "Review of Latest Advances in 3GPP Standardization: D2D Communication in 5G Systems and its Energy Consumption Models," *Future Internet*, vol. 10, no. 1, p. 3, 2018.

102. N. Bonjorn, F. Foukalas, and P. Pop, "Enhanced 5G V2X services using sidelink device-to-device communications," in *2018 17th annual mediterranean ad hoc networking workshop (Med-Hoc-Net)*, pp. 1–7, IEEE, 2018.

103. "Mission-critical Services in 3GPP." Available online: `https://www.3gpp.org/news-events/3gpp-news/1875-mc_services` (accessed on January 24, 2023).

104. G. Fodor, H. Do, S. A. Ashraf, R. Blasco, W. Sun, M. Belleschi, and L. Hu, "Supporting Enhanced Vehicle-to-Everything Services by LTE Release 15 Systems," *IEEE Communications Standards Magazine*, vol. 3, no. 1, pp. 26–33, 2019.

105. J. Peisa, P. Persson, S. Parkvall, E. Dahlman, A. Grovlen, C. Hoymann, and D. Gerstenberger, "5G New Radio Evolution," *Ericsson technology review*, February 2020.

106. 3GPP, "Enhanced Relays for Energy Efficiency and Extensive Coverage (Release 17)," 3GPP TR 22.866 V17.1.0, December 2019.

107. Q. Zhao, S. Paris, T. Veijalainen, and S. Ali, "Hierarchical Multi-Objective Deep Reinforcement Learning for Packet Duplication in Multi-Connectivity for URLLC," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pp. 142–147, IEEE, 2021.

108. 3GPP, "Service requirements for the 5G system (Release 18)," 3GPP TS 22.261 V18.3.0, June 2021.

109. 3GPP, "Technical Specification 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; User Equipment (UE) radio transmission and reception; Part 3: Range 1 and Range 2 Interworking operation with other radios (Release 16)," *TS 38.101-3*, vol. V16.1.0, Sept. 2019.

110. "Technical Specification Group Radio Access Network; Study on Channel Model for Frequency Spectrum above 6 GHz (Release 14)," tech. rep., 3GPP TR 38.900 V14.2.0, December 2016.

111. D. Solomitckii, A. Orsino, S. Andreev, Y. Koucheryavy, and M. Valkama, "Characterization of mmWave Channel Properties at 28 and 60 GHz in Factory Automation Deployments," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2018.

112. 3GPP, "Study on LTE Device to Device Proximity Services; Radio Aspects(Release 12)," *TS 36.843*, vol. V12.0.1, Mar. 2014.

113. N. Chukhno, A. Orsino, J. Torsner, A. Iera, and G. Araniti, "5G NR Sidelink Multi-Hop Transmission in Public Safety and Factory Automation Scenarios," *IEEE Network*, 2023.

114. R. Hattachi and J. Erfanian, "5G White Paper," *Next Generation Mobile Networks Alliance*, 2015.

115. 3GPP, "Feasibility Study on New Services and Markets Technology Enablers for Critical Communications," 3GPP TR 22.862 V14.1.0, September 2016.

116. J. Talvitie, T. Levanen, M. Koivisto, K. Pajukoski, M. Renfors, and M. Valkama, "Positioning of High-Speed Trains Using 5G New Radio Synchronization Signals," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2018.

117. J. Talvitie, T. Levanen, M. Koivisto, T. Ihalainen, K. Pajukoski, and M. Valkama, "Positioning and Location-Aware Communications for Modern Railways with 5G New Radio," *IEEE Communications Magazine*, vol. 57, no. 9, pp. 24–30, 2019.

118. P. Zhang, J. Lu, Y. Wang, and Q. Wang, "Cooperative Localization in 5G Networks: A Survey," *Ict Express*, vol. 3, no. 1, pp. 27–32, 2017.

119. R. M. Buehrer, H. Wymeersch, and R. M. Vaghefi, "Collaborative Sensor Network Localization: Algorithms and Practical Issues," *Proceedings of the IEEE*, vol. 106, no. 6, pp. 1089–1114, 2018.

120. S. Khandker, J. Torres-Sospedra, and T. Ristaniemi, "Improving RF Fingerprinting Methods by Means of D2D Communication Protocol," *Electronics*, vol. 8, no. 1, p. 97, 2019.

121. 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 Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2017.

122. 3GPP, "Service Requirements for Cyber-Physical Control Applications in Vertical Domains (Release 18)," tech. rep., 3GPP TS 22.104 V18.0.0, March 2021.

123. S. Bartoletti, A. Conti, D. Dardari, and A. Giorgetti, "5G Localization and Context-Awareness," *University of Bologna, University of Ferrara*, 2018.

124. 3GPP, "Study on Indoor Positioning Enhancements for UTRA and LTE (Release 13)," 3GPP TR 37.857 V13.1.0, December 2015.

125. K. McDermott, R. M. Vaghefi, and R. M. Buehrer, "Cooperative UTDOA positioning in LTE Cellular Systems," in *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, IEEE, 2015.

126. R. S. Campos, "Evolution of Positioning Techniques in Cellular Networks From 2G to 4G," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.

127. F. Mogyorósi, P. Revisnyei, A. Pašić, Z. Papp, I. Törös, P. Varga, and A. Pašić, "Positioning in 5G and 6G Networks—A Survey," *Sensors*, vol. 22, no. 13, p. 4757, 2022.

128. H. Wymeersch, G. Seco-Granados, G. Destino, D. Dardari, and F. Tufvesson, "5G mmWave Positioning for Vehicular Networks," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 80–86, 2017.

129. H. Wymeersch, N. Garcia, H. Kim, G. Seco-Granados, S. Kim, F. Wen, and M. Fröhle, "5G mmWave Downlink Vehicular Positioning," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 206–212, IEEE, 2018.

130. C. Ma, J. Yang, J. Chen, and Y. Tang, "Indoor and Outdoor Positioning System based on Navigation Signal Simulator and Pseudolites," *Advances in Space Research*, vol. 62, no. 9, pp. 2509–2517, 2018.

131. P. Pascacio, S. Casteleyn, J. Torres-Sospedra, E. S. Lohan, and J. Nurmi, "Collaborative Indoor Positioning Systems: A Systematic Review," *Sensors*, vol. 21, no. 3, p. 1002, 2021.

132. J. Palacios, P. Casari, and J. Widmer, "JADE: Zero-Knowledge Device Localization and Environment Mapping for Millimeter Wave Systems," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pp. 1–9, IEEE, 2017.

133. O. Kanhere and T. S. Rappaport, "Position Locationing for Millimeter Wave Systems," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 206–212, IEEE, 2018.

134. N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive MIMO," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2475–2487, 2017.

135. H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative Localization in Wireless Networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

136. N. Garcia, A. M. Haimovich, J. A. Dabin, M. Coulon, and M. Lops, "Direct Localization of Emitters Using Widely Spaced Sensors in Multipath Environments," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 695–700, IEEE, 2014.

137. N. Pirzada, M. Y. Nayan, F. Subhan, M. F. Hassan, and M. A. Khan, "Comparative Analysis of Active and Passive Indoor Localization Systems," *AASRI Procedia*, vol. 5, pp. 92–97, 2013.

138. Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan, and Y. Xiang, "Passive Indoor Localization based on SCI and Naive Bayes Classification," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1566–1577, 2017.

139. N. Al-Sabbagh and A. Al-Omary, "A Centralized Multi-Floor Indoor Navigation System for a Large Mall," in *2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*, pp. 1–6, IEEE, 2019.

140. M. Sakr, A. Masiero, and N. El-Sheimy, "LocSpeck: A Collaborative and Distributed Positioning System for Asymmetric Nodes Based on UWB ad-hoc Network and Wi-Fi Fingerprinting," *Sensors*, vol. 20, no. 1, p. 78, 2019.

141. Y. H. Ho and H. C. Chan, "Decentralized Adaptive Indoor Positioning Protocol Using Bluetooth Low Energy," *Computer Communications*, vol. 159, pp. 231–244, 2020.

142. A. Das and G. Dubbelman, "An Experimental Study on Relative and Absolute Pose Graph Fusion for Vehicle Localization," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 630–635, IEEE, 2018.

143. P. Zhang and Q. Wang, "On Using the Relative Configuration to Explore Cooperative Localization," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 968–980, 2014.

144. R. Klus, L. Klus, D. Solomitckii, M. Valkama, and J. Talvitie, "Deep Learning Based Localization and HO Optimization in 5G NR Networks," in *2020 International Conference on Localization and GNSS (ICL-GNSS)*, pp. 1–6, IEEE, 2020.

145. A. Ometov, A. Orsino, L. Militano, G. Araniti, D. Moltchanov, and S. Andreev, "A Novel Security-Centric Framework for D2D Connectivity based on Spatial and Social Proximity," *Computer Networks*, vol. 107, pp. 327–338, 2016.

146. H. Sun, Z. Zhang, R. Q. Hu, and Y. Qian, "Wearable Communications in 5G: Challenges and Enabling Technologies," *ieee vehicular technology magazine*, vol. 13, no. 3, pp. 100–109, 2018.

147. A. Ometov, V. Shubina, L. Klus, J. Skibińska, S. Saafi, P. Pascacio, L. Flueratoru, D. Q. Gaibor, N. Chukhno, O. Chukhno, *et al.*, "A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges," *Computer Networks*, vol. 193, p. 108074, 2021.

148. A. Asadi and V. Mancuso, "WiFi Direct and LTE D2D in action," in *2013 IFIP Wireless Days (WD)*, pp. 1–8, IEEE, 2013.

149. S. Saafi, J. Hosek, and A. Kolackova, "Cellular-enabled Wearables in Public Safety Networks: State of the Art and Performance Evaluation," in *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 201–207, IEEE, 2020.

150. M. M. Alam, D. B. Arbia, and E. B. Hamida, "Research Trends in Multi-Standard Device-to-Device Communication in Wearable Wireless Networks," in *International Conference on Cognitive Radio Oriented Wireless Networks*, pp. 735–746, Springer, 2015.

151. S.-Y. Lien, D.-J. Deng, C.-C. Lin, H.-L. Tsai, T. Chen, C. Guo, and S.-M. Cheng, "3GPP NR Sidelink Transmissions Toward 5G V2X," *IEEE Access*, vol. 8, pp. 35368–35382, 2020.

152. H. Bagheri, Z. Liu, H. Lee, D. Pesch, K. Moessner, P. Xiao, *et al.*, "5G NR-V2X: Towards Connected and Cooperative Autonomous Driving," *arXiv preprint arXiv:2009.03638*, 2020.

153. M. Koivisto, M. Costa, J. Werner, K. Heiska, J. Talvitie, K. Leppänen, V. Koivunen, and M. Valkama, "Joint Device Positioning and Clock Synchronization in 5G Ultra-Dense Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2866–2881, 2017.

154. A. Dammann, R. Raulefs, and S. Zhang, "On Prospects of Positioning in 5G," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1207–1213, IEEE, 2015.

155. X. Cui, T. A. Gulliver, H. Song, and J. Li, "Real-Time Positioning based on Millimeter Wave Device to Device Communications," *IEEE Access*, vol. 4, pp. 5520–5530, 2016.

156. X. Cui, T. A. Gulliver, J. Li, and H. Zhang, "Vehicle Positioning Using 5G millimeter-wave Systems," *IEEE Access*, vol. 4, pp. 6964–6973, 2016.

157. A. Papapostolou, W. Xiao, and H. Chaouchi, "Cooperative Fingerprint-based Indoor Localization Using Self-Organizing Maps," in *2011 7th International Wireless Communications and Mobile Computing Conference*, pp. 1814–1819, IEEE, 2011.

158. T. Iwase and R. Shibasaki, "Infra-free Indoor Positioning Using Only Smartphone Sensors," in *International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–8, IEEE, 2013.

159. P. Raveneau, S. D'Alu, and H. Rivano, "Localisation based on Wi-Fi Fingerprints: A Crowdsensing Approach with a Device-to-Device Aim," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 321–325, IEEE, 2017.

160. M. Karlsson and F. Karlsson, "Cooperative Indoor Positioning by Exchange of Bluetooth Signals and State Sstimates Between Users," in *2016 European Control Conference (ECC)*, pp. 1440–1444, IEEE, 2016.

161. R. M. Vaghefi and R. M. Buehrer, "Improving Positioning in LTE through Collaboration," in *2014 11th Workshop on Positioning, Navigation and Communication (WPNC)*, pp. 1–6, IEEE, 2014.

162. H. Wymeersch, G. R. de Campos, P. Falcone, L. Svensson, and E. G. Ström, "Challenges for Cooperative ITS: Improving Road Safety Through the Integration of Wireless Communications, Control, and Positioning," in *2015 International Conference on Computing, Networking and Communications (ICNC)*, pp. 573–578, IEEE, 2015.

163. R. M. Vaghefi and R. M. Buehrer, "Cooperative RF Pattern Matching Positioning for LTE Cellular Systems," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 264–269, IEEE, 2014.

164. L. Yin, Q. Ni, and Z. Deng, "A GNSS/5G Integrated Positioning Methodology in D2D Communication Networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 2, pp. 351–362, 2018.

165. Z. Zhou, H. Yu, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, "Dependable Content Distribution in D2D-based Cooperative Vehicular Networks: A Big Data-Integrated Coalition Game Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 953–964, 2018.

166. R. Raulefs, S. Zhang, and C. Mensing, "Bound-based Spectrum Allocation for Cooperative Positioning," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 1, pp. 69–83, 2013.

167. A. Bourdoux, A. N. Barreto, B. van Liempd, C. de Lima, D. Dardari, D. Belot, E.-S. Lohan, G. Seco-Granados, H. Sarieddeen, H. Wymeersch, *et al.*, "6G White Paper on Localization and Sensing," *arXiv preprint arXiv:2006.01779*, 2020.

168. C. De Lima, D. Belot, R. Berkvens, A. Bourdoux, D. Dardari, M. Guillaud, M. Isomursu, E.-S. Lohan, Y. Miao, A. N. Barreto, *et al.*, "Convergent Communication, Sensing and Localization in 6G Systems: An Overview of Technologies, Opportunities and Challenges," *IEEE Access*, vol. 9, pp. 26902–26925, 2021.

169. M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppanen, and M. Valkama, "High-Efficiency Device Positioning and Location-Aware Communications in Dense 5G Networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 188–195, 2017.

170. D. K. P. Tan, J. He, Y. Li, A. Bayesteh, Y. Chen, P. Zhu, and W. Tong, "Integrated Sensing and Communication in 6G: Motivations, Use cases, Requirements, Challenges and Future Directions," in *2021 1st IEEE International Online Symposium on Joint Communications & Sensing (JC&S)*, IEEE, 2021.

171. Z. Xiao and Y. Zeng, "An Overview on Integrated Localization and Communication Towards 6G," *Science China Information Sciences*, vol. 65, no. 3, pp. 1–46, 2022.

172. H. Wymeersch, D. Shrestha, C. M. De Lima, V. Yajnanarayana, B. Richerzhagen, M. F. Keskin, K. Schindhelm, A. Ramirez, A. Wolfgang, M. F. De Guzman, *et al.*, "Integration of Communication and Sensing in 6G: A Joint Industrial and Academic Perspective," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, IEEE, 2021.

173. M. Gerasimenko, D. Moltchanov, M. Gapeyenko, S. Andreev, and Y. Koucheryavy, "Capacity of Multiconnectivity mmWave Systems with Dynamic Blockage and Directional Antennas," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3534–3549, 2019.

174. H. Sarieddeen, N. Saeed, T. Y. Al-Naffouri, and M.-S. Alouini, "Next Generation Terahertz Communications: A Rendezvous of Sensing, Imaging, and Localization," *IEEE Communications Magazine*, vol. 58, no. 5, pp. 69–75, 2020.

175. K. Keykhosravi, M. F. Keskin, S. Dwivedi, G. Seco-Granados, and H. Wymeersch, "Semi-Passive 3D Positioning of Multiple RIS-Enabled Users," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 11073–11077, 2021.

176. H. Wymeersch, J. He, B. Denis, A. Clemente, and M. Juntti, "Radio Localization and Mapping with Reconfigurable Intelligent Surfaces: Challenges, Opportunities, and Research Directions," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 52–61, 2020.

177. J. He, H. Wymeersch, T. Sanguanpuak, O. Silvén, and M. Juntti, "Adaptive Beamforming Design for mmWave RIS-aided Joint Localization and Communication," in *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, IEEE, 2020.

178. Y. Cui and H. Yin, "Channel Estimation for RIS-aided mmWave Communications via 3D Positioning," in *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, pp. 399–404, IEEE, 2021.

179. M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, and S. Tretyakov, "Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and the Road Ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

180. E. Björnson, Ö. Özdogan, and E. G. Larsson, "Reconfigurable Intelligent Surfaces: Three Myths and Two critical Questions," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, 2020.

181. M. Heino, D. Korpi, T. Huusari, E. Antonio-Rodriguez, S. Venkatasubramanian, T. Riihonen, L. Anttila, C. Icheln, K. Haneda, R. Wichman, *et al.*, "Recent Advances in Antenna Design and Interference Cancellation Algorithms for In-band Full Duplex Relays," *IEEE Communications magazine*, vol. 53, no. 5, pp. 91–101, 2015.

182. S. Series, "Comparison of Time-Difference-of-Arrival and Angle-of-Arrival Methods of Signal Geolocation," *ITU: Geneva, Switzerland*, 2011.

183. Z. Zheng, L. Liu, and W. Hu, "Accuracy of Ranging based on DMT Visible Light Communication for Indoor Positioning," *IEEE Photonics Technology Letters*, vol. 29, no. 8, pp. 679–682, 2017.

184. K. Järvinen, H. Leppäkoski, E.-S. Lohan, P. Richter, T. Schneider, O. Tkachenko, and Z. Yang, "PILOT: Practical Privacy-Preserving Indoor Localization Using Outsourcing," in *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 448–463, IEEE, 2019.

185. V. Shubina, A. Ometov, S. Andreev, D. Niculescu, and E. S. Lohan, "Privacy versus Location Accuracy in Opportunistic Wearable Networks," in *2020 International Conference on Localization and GNSS (ICL-GNSS)*, IEEE, 2020.

186. J. Zhang, Z. Zheng, Z. Fei, and Z. Han, "Energy-Efficient Multi-User Localization in the RIS-Assisted IoT Networks," *IEEE Internet of Things Journal*, 2022.

187. N. Alam and A. G. Dempster, "Cooperative Positioning for Vehicular Networks: Facts and future," *IEEE transactions on intelligent transportation systems*, vol. 14, no. 4, pp. 1708–1717, 2013.

188. H. Xiong, M. Peng, S. Gong, and Z. Du, "A Novel Hybrid RSS and TOA Positioning Algorithm for Multi-Objective Cooperative Wireless Sensor Networks," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9343–9351, 2018.

189. K. Pahlavan and A. H. Levesque, *Wireless Information Networks*. John Wiley & Sons, 2005.

190. M. K. Simon and M.-S. Alouini, "Digital Communications Over Fading Channels (MK Simon and MS Alouini; 2005)[Book Review]," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3369–3370, 2008.

191. S. Mazuelas, F. A. Lago, D. González, A. Bahillo, J. Blas, P. Fernandez, R. M. Lorenzo, and E. J. Abril, "Dynamic Estimation of Optimum Path Loss Model in a RSS Positioning System," in *2008 IEEE/ION Position, Location and Navigation Symposium*, pp. 679–684, IEEE, 2008.

192. J. Du, C. Yuan, M. Yue, and T. Ma, "A Novel Localization Algorithm Based on RSSI and Multilateration for Indoor Environments," *Electronics*, vol. 11, no. 2, p. 289, 2022.

193. M. Gapeyenko, V. Petrov, D. Moltchanov, M. R. Akdeniz, S. Andreev, N. Himayat, and Y. Koucheryavy, "On the Degree of Multi-Connectivity in 5G Millimeter-wave Cel-

lular Urban Deployments," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1973–1978, 2018.

194. S. Zeng, H. Zhang, B. Di, Z. Han, and L. Song, "Reconfigurable Intelligent Surface (RIS) Assisted Wireless Coverage Extension: RIS Orientation and Location Optimization," *IEEE Communications Letters*, vol. 25, no. 1, pp. 269–273, 2020.
195. S. W. Ellingson, "Path loss in Reconfigurable Intelligent Surface-enabled Channels," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 829–835, IEEE, 2021.
196. I. Yildirim, A. Uyrus, and E. Basar, "Modeling and Analysis of Reconfigurable Intelligent Surfaces for Indoor and Outdoor Applications in Future Wireless Networks," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1290–1301, 2020.
197. C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, *et al.*, "Reconfigurable Intelligent Surfaces for 6G Systems: Principles, Applications, and Research Directions," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 14–20, 2021.
198. M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, *et al.*, "Reconfigurable Intelligent Surfaces vs. Relaying: Differences, Similarities, and Performance Comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.
199. E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent Reflecting Surface versus Decode-and-Forward: How Large Surfaces Are Needed to Beat Relaying?," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2019.
200. ISO Central Secretary, "Information Technology—Real Time Locating Systems—Test and Evaluation of Localization and Tracking Systems," Standard ISO/IEC JTC 1/SC 31, International Organization for Standardization, 2016.
201. F. Potortì, A. Crivello, and F. Palumbo, "The EvAAL Evaluation Framework and the IPIN Competitions," in *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*, pp. 209–224, Elsevier, 2019.
202. N. Chukhno, S. Trilles, J. Torres-Sospedra, A. Iera, and G. Araniti, "D2D-based Cooperative Positioning Paradigm for Future Wireless Systems: A Survey," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5101–5112, 2021.
203. 3GPP, "Study on NR Sidelink Relay," *RP-193253*, 2019.

Nadezhda Chukhno is an Early Stage Researcher at A-WEAR and Doctoral Researcher at Mediterranea University of Reggio Calabria, Italy and Jaume I University, Spain. She graduated from RUDN University, Russia, and received her B.Sc. in Business Informatics (2017) and M.Sc. in Fundamental Informatics and Information technologies (2019). Her current research activity mainly focuses on wireless communications, 5G+ networks, multicasting, D2D, and wearable technologies.

Recently, the popularity of Millimeter Wave (mmWave) wireless networks has increased due to their capability to cope with the escalation of mobile data demands caused by the unprecedented proliferation of smart devices in the fifth-generation (5G). Extremely high frequency or mmWave band is a fundamental pillar in the provision of the expected gigabit data rates. Hence, according to both academic and industrial communities, mmWave technology, e.g., 5G New Radio (NR) and WiGig (60 GHz), is considered as one of the main components of 5G and beyond networks. Particularly, the 3rd Generation Partnership Project (3GPP) provides for the use of licensed mmWave sub-bands for the 5G mmWave cellular networks, whereas IEEE actively explores the unlicensed band at 60 GHz for the next-generation wireless local area networks. In this regard, mmWave has been envisaged as a new technology layout for real-time heavy-traffic and wearable applications.

This very work is devoted to solving the problem of mmWave band communication system while enhancing its vantages through utilizing the direct communication radio interface for NR multicasting, cooperative positioning, and mission-critical applications. The *main contributions* presented in this work include: *(i)* a set of mathematical frameworks and simulation tools to characterize multicast traffic delivery in mmWave directional systems; *(ii)* sidelink relaying concept exploitation to deal with the channel condition deterioration of dynamic multicast systems and to ensure mission-critical and ultra-reliable low-latency communications; *(iii)* cooperative positioning techniques analysis for enhancing cellular positioning accuracy for 5G+ emerging applications that require not only improved communication characteristics but also precise localization.

Our study indicates the need for additional mechanisms/research that can be utilized: *(i)* to further improve multicasting performance in 5G/6G systems; *(ii)* to investigate sidelink aspects, including, but not limited to, standardization perspective and the next relay selection strategies; and *(iii)* to design cooperative positioning ystems based on Device-to-Device (D2D) technology.

5G, New Radio, millimeter Wave, sidelink, device-to-device (D2D), multi-hop, public safety, factory automation, collaborative localization, cooperative localization, wearables.

UNIVERSITAT JAUME·I

UNIVERSITÀ DEGLI STUDI DI REGGIO CALABRIA Mediterranea