UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF INFORMATION ENGINEERING

PH.D. COURSE IN INFORMATION ENGINEERING
SCIENCE AND INFORMATION TECHNOLOGY CURRICULUM

XXXII SERIES

The Potential of Millimeter Waves for Future 5G Cellular and Vehicular Networks

Ph.D. Candidate
MARCO GIORDANI

Supervisor
Prof. Michele Zorzi

Coordinator
Prof. Andrea Neviani

Academic Year 2018/2019

To the memory of my father,
my strength, my inspiration, my role model...

Abstract

The fifth generation of wireless technology (5G) is positioned to address the demands and business contexts of 2020 and beyond. It is expected to enable a fully mobile and connected society, related to the significant growth in connectivity and volume of traffic that will be required in the near future. In this context, the millimeter wave (mmWave) spectrum is rapidly emerging as a key enabler of the 5G performance demands, thanks to the large available bandwidth at such high frequencies. Communication at mmWaves, however, suffers from severe path and penetration loss, requires the maintenance of directional transmissions and calls for the definition of new control operations for both cellular and vehicular networks. Among all the challenges that will be faced, in this thesis we (i) focus on the design of mobility management strategies for devices in idle and connected mode, (ii) investigate how to deploy mmWave networking architectures, (iii) validate the potential of the mmWave technology as a means to foster the automotive revolution towards connected and autonomous transportation systems, (iv) study the most promising options to broadcast vehicular sensory observations in an efficient way, and (v) envision how 5G technologies can evolve into 6G to address the needs of the future digital society. Among other results, we demonstrate the importance of combining multiple radio technologies into a single solution that is more robust and efficient than any individual approach, discuss the trade-offs of mobility management in 3GPP NR, and evaluate practical strategies for assigning value of information in 5G networks.

Sommario

La tecnologia cellulare di quinta generazione (5G) nasce per far fronte alla crescente richiesta di una connettività veloce, a bassa latenza e robusta, in linea con le esigenze di mercato della società del prossimo futuro. In questo contesto, lo spettro delle onde millimetriche (mmWave) sta rapidamente emergendo come un catalizzatore essenziale per lo sviluppo delle tecnologie 5G, in particolare grazie all'enorme quantità di banda disponibile ad alte frequenze. La comunicazione ad onde millimetriche, tuttavia, è ostacolata da una serie di problematiche, legate principalmente alla rapida attenuazione della potenza in spazio libero, all'esigenza di mantenere comunicazioni direzionali, e alla necessità di definire nuove procedure di controllo per reti cellulari e veicolari. Tra tutte le sfide che si dovranno affrontare, in questa tesi (i) progetteremo architetture per gestire la mobilità degli utenti e le trasmissioni direzionali, (ii) investigheremo come installare efficientemente reti cellulari a onde millimetriche, (iii) valideremo l'effettivo potenziale delle onde millimetriche per favorire lo sviluppo di sistemi di trasporto totalmente autonomi e connessi, (iv) analizzeremo meccanismi per trasmettere dati tra veicoli in modo efficiente, (v) discuteremo di come i sistemi 5G possano evolversi in una nuova generazione (6G) con l'obiettivo di garantire comunicazioni con prestazioni sempre migliori. Tra i vari risultati, dimostreremo l'importanza di utilizzare le tecnologie radio in parallelo, discuteremo i pro e contro di diverse proposte per il controllo della mobilità dei dispositivi cellulari e veicolari, e valuteremo le prestazioni di strategie per assegnare valore all'informazione.

Contents

List of Figures

List of Tables

Acronyms

3GPP 3rd Generation Partnership Project
5G 5th generation
5GC 5G Core
6G 6th generation
ADC Analog to Digital Converter
AHP Analytic Hierarchy Process
AM Acknowledged Mode
AMF Access and Mobility Management Function
AoI Age of Information
BF Bayesian Filtering
BSM Basic Safety Message
C-ITS Connected Intelligent Transportation System
CA Carrier Aggregation
CAM Cooperative Awareness Message
CAV Connected and Autonomous Vehicle
CDF Cumulative Distribution Function
CQI Channel Quality Information
CRS Cell Reference Signal
CSCC Channel Sensing Congestion Control
CSI Channel State Information
CSI-RS Channel State Information - Reference Signal
CSMA/CA Carrier Sense Multiple Access with Collision Avoidance
CTRA Constant Turn Rate and Acceleration
CU Central Unit
DAG Directed Acyclic Graph
DASH Dynamic Adaptive Streaming over HTTP
DCI Downlink Control Information
DCM Dual Carrier Modulation
DL Downlink
DL-SCH Downlink Shared Cannel
DMRS DeModulation Reference Signal
DR Dead Reckoning
DSRC Dedicated Short Range Communication
DU Distributed Unit
EC European Commission
eMBB enhanced Mobile Broadband
eNB eNodeB
EPC Evolved Packet Core
ETB Error Threshold Broadcasting
FCC Federal Communications Commission
FDD Frequency Division Duplexing
FDM Frequency Division Multiplexing

FS Fast Switching
gNB Next Generation Node Base
GPS Global Positioning System
GTP GPRS Tunneling Protocol
HARQ Hybrid Automatic Repeat reQuest
HH Hard Handover
HMM Hidden Markov Model
HQF Highest-Quality-First
HTTP HyperText Transfer Protocol
IA Initial Access
IAB Integrated Access and Backhaul
IoT Internet of Things
ITU International Telecommunication Union
KF Kalman Filter
KPI Key Performance Index
LDM Local Dynamic Map
LDPC Low-Density Parity-Check
LED Light Emitting Diode
LIDAR Light Detection and Ranging
LOS Line of Sight
LTE Long Term Evolution
M2M Machine to Machine
MAC Medium Access Control
MC Multi-Connectivity
MCS Modulation and Coding Scheme
MEC Mobile Edge Cloud
MIB Master Information Block
MIMO Multiple Input Multiple Output
ML Machine Learning
MLR Maximum-Local-Rate
MME Mobility Management Entity
mmWave Millimeter Wave
MR Maximum Rate
MS Maximum SNR
MT Mobile Termination
NACC Neighbor Aware Congestion Control
NFV Network Function Virtualization
NLOS Non Line of Sight
NLOSv Vehicle Non Line of Sight
NN Neural Network
ns-3 Network Simulator 3
NSA non-standalone
OFDM Orthogonal Frequency Division Multiplexing
PA Position-Aware
PB Periodic Broadcasting
PBCH Physical Broadcast Channel
PDCP Packet Data Convergence Protocol
PF Proportional Fair
PHY Physical
PPP Poission Point Process
PSFCH Physical Sidelink Feedback Channel
PSS Primary Synchronization Signal
QoE Quality of Experience
QoS Quality of Service
RA Requirement-Aware
RA Random Access
RACH Random Access Channel

RAN Radio Access Network
RAT Radio Access Technology
RBF Radial Basis Function
RF Radio Frequency
RLC Radio Link Control
RLF Radio Link Failure
RR Round Robin
RRC Radio Resource Control
RRM Radio Resource Management
RSRP Reference Signal Received Power
RSS Received Signal Strength
RT Report Table
RTT Round Trip Time
SA standalone
SCH Secondary Cell Handover
SDAP Service Data Adaptation Protocol
SDM Space Division Multiplexing
SDMA Spatial Division Multiple Access
SI Study Item
SIB System Information Block
SINR Signal to Interference plus Noise Ratio
SNR Signal to Noise Ratio
SRS Sounding Reference Signal
SS Synchronization Signal
SSS Secondary Synchronization Signal
ST Spanning Tree
SUMO Simulation of Urban MObility
SVM Support Vector Machine
TCP Transmission Control Protocol
TDD Time Division Duplexing
TDM Time Division Multiplexing
TDMA Time Division Multiple Access
TTT Time-to-Trigger
UAV Unmanned Aerial Vehicle
UDP User Datagram Protocol
UE User Equipment
UKF Unscented Kalman Filter
UL Uplink
UMTS Universal Mobile Telecommunication Systems
UPA Uniform Planar Array
URLLC Ultra-Reliable Low Latency Communication
V2N Vehicle-to-Network
V2V Vehicle-to-Vehicle
V2X Vehicle-to-Everything
VLC Visible Light Communications
VN Vehicular Node
VoI Value of Information
WBF Wired Bias Function
WF Wired-first

Chapter 1

Introduction

From analogue cellular since Long Term Evolution (LTE), each generation of mobile technology has been designed to meet the needs of end users and network operators. In this regard, the 5th generation (5G) of mobile technology will not simply be an evolution of earlier cellular standards. On the contrary, it is positioned to address the demands and business contexts of 2020 and beyond. It is expected to enable a fully mobile and connected society, related to the tremendous growth in connectivity and density/volume of traffic that will be required in the near future [1, 2]. According to the Cisco Visual Networking Index (VNI) [3][1], in fact, it is expected that global mobile data traffic will grow at a Compound Annual Growth Rate (CAGR) of 47 percent between 2019 to 2022 (more than the rate at which both the population and Internet users are growing) reaching 77.5 exabytes per month by 2022. Mobile devices and connections will grow to 12.3 billion within the next four years – and over 422 million of those will be 5G capable.

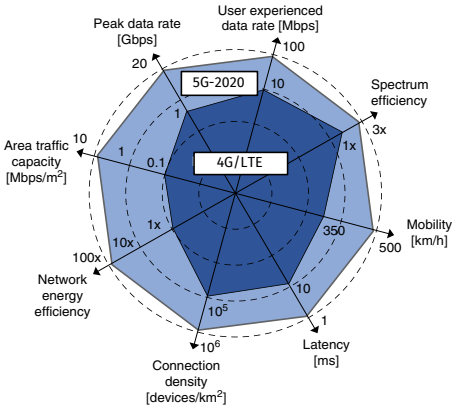
Previous generations of wireless networks have always been tailored towards one particular need and a particular business ecosystem, such as mobile broadband in the case of LTE, and will not reasonably be able to support this continuing growth in demand from subscribers for better mobile broadband experiences [4]. In contrast, 5G will be associated with the need for multi-service and multi-tenancy support, and is commonly understood to comprise a heterogeneous variety of tightly integrated radio solutions and breakthrough technologies to address different service needs.

1.1 5G Service Requirements

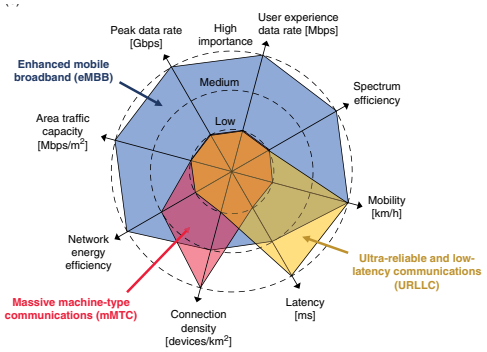
According to the International Telecommunication Union (ITU), 5G wireless networks will need to satisfy the following key requirements [4, 5], as illustrated in Fig. 1.1

- *Peak data rate*, referred to the maximum achievable data rate under ideal conditions per user or device in bits per second. The minimum 5G requirements for peak data rate are 20 Gbps in the Downlink (DL) and 10 Gbps in the Uplink (UL). In particular, per-user experienced data rate should reach up to 100 Mbps in the DL and 50 Mbps in the UL.

[1] The Cisco Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update is part of the comprehensive Cisco VNI Forecast, an ongoing initiative to track and forecast the impact of visual networking applications on global networks, presenting some of Cisco's major global mobile data traffic projections and growth trends.



(a) 5G vs. 4G/LTE KPIs enhancements.



(b) 5G KPIs importance.

Fig. 1.1: Key capabilities of 5G wireless systems [4].

- *Area traffic capacity*, defined as the total traffic throughput served per geographic area in Mbps/m^2. ITU has defined this objective with a target of 10 Mbps/m^2 for the DL.
- *Latency*, defined as the time from when the source sends a packet to when the destination receives it. The one-way end-to-end latency requirement is set to 4 ms for enhanced Mobile Broadband (eMBB) services and 1 ms for Ultra-Reliable Low Latency Communication (URLLC). Control plane latency, instead, should be bounded to less than 20 ms.
- *Connection density*, corresponding to the total number of connected and/or accessible devices per unit area. ITU has specified a target of 1 million devices per km^2 for machine-type communication services.
- *Energy efficiency*, on the network side referring to the quantity of information bits transmitted to or received from users. The specification given by ITU in this respect is that 5G air interfaces must have the ability to support a high sleep ratio and long sleep duration.
- *Reliability*, defined as the success probability of delivering a data packet before a given deadline. The target is to transmit packets of 32 bytes in less than 1 ms in the cell edge of the dense urban test environment with 99.999% probability.
- *Mobility*, defined as the maximum speed at which a defined Quality of Service (QoS) can be achieved. For the rural test environment, the normalized traffic channel link data rate at 500 km/h, reflecting the average user spectral efficiency, must be larger than 0.45 bps/Hz in the UL. At the same time, mobility interruption time, i.e., the time during which the device cannot exchange data packets because of handover procedures, should be as low as 0 ms, essentially meaning that a make-before-break paradigm has to be applied, i.e., the connection to the new cell has to be set up before the old one is dropped.

1.2 5G Use Cases and Applications

5G will need to support, apart from the evolution of mobile broadband, new use cases ranging from delay-sensitive video applications to ultra-reliable safety services. Researchers, industry vectors and standardization bodies, like 3GPP and ITU, have categorized use cases in a few broad application domains (as depicted in Fig. 1.2) that, for their generality and complementarity, are good representatives of future 5G services [4, 6].

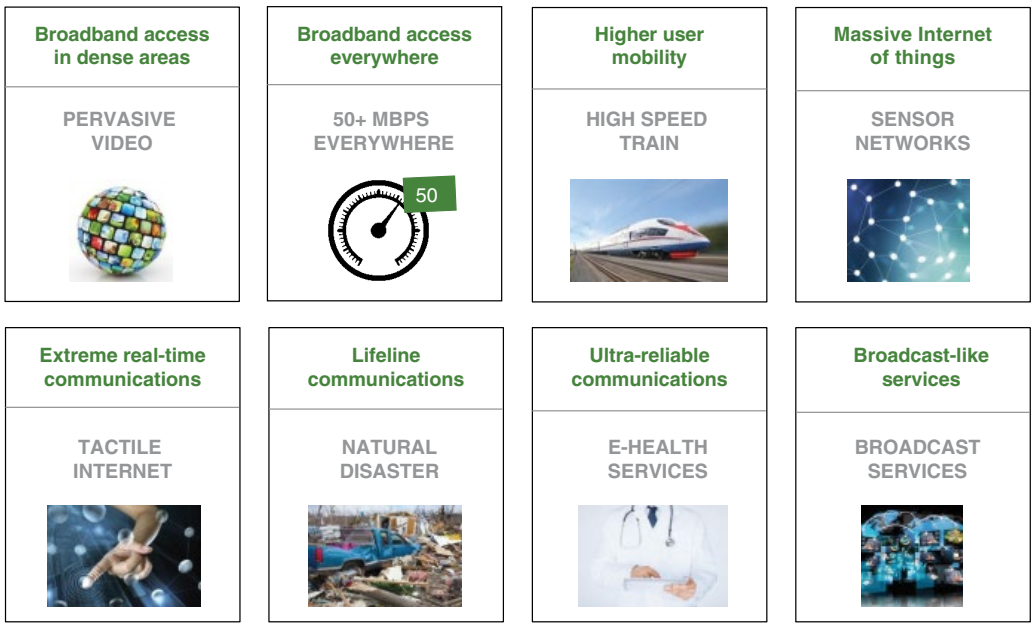


Fig. 1.2: Representative 5G use cases [4].

- *Broadband access.* It refers to the use cases requiring high data rates across a wide coverage area, including enhanced access to multimedia content, services and data with improved performance and increasingly seamless user experience. Applications include (i) pervasive video streaming (video traffic will be 82% of all IP traffic by 2022, up from 75% in 2017), (ii) operator cloud services, (iii) information sharing and traffic support in ultra-dense open-air gathering events (e.g., where several hundred thousands of users per km^2 request connectivity).
- *High-speed user mobility.* Beyond 2020, there will be a growing demand for mobile services in very high speed scenarios. Vehicles, in particular, will demand enhanced connectivity for on-board entertainment, accessing the Internet, enhanced navigation through real-time information dissemination, autonomous driving, safety and vehicle diagnostics. 5G will foster the automotive revolution towards fully autonomous transportation systems.
- *Internet of Things and smart cities.* 5G will accelerate and, in some cases, enable the adoption of solutions for so-called smart cities, improving the quality of life through better energy-efficient architecture, environment and waste management, improved city transportation.
- *e-health services.* 5G will revolutionize health services, for instance through the possibility of wirelessly enabled smart pharmaceuticals or remote surgery with haptic feedback. Depending on the patient's device, treatment reactions based on monitored data may be required, and these should be immediate and (semi-)automatic. Such applications will involve significant growth in remote operation and control and will require extremely low latency and ultra-high reliability, due to the sensitive nature of the exchanged information.
- *Lifeline communication.* Public safety and emergency services that are provided today are continuously improving. In addition to new capabilities for authority-to-citizen communication, these use cases will evolve to include emergency prediction and disaster relief, as well as remote control of vehicles and machines in dangerous or inaccessible areas, e.g., in the fields of mining. Specifically, 5G should be able to provide robust communications in case of natural disasters (e.g., earthquakes, floods, hurricanes), energy-efficient network management, high

level of connection availability in addition to the ability to support traffic surges.

- *Broadcast services.* While personalization of communication will lead to a reducing demand for currently deployed broadcast services, e.g., linear TV, future fully mobile and connected society will nonetheless need efficient distribution of information from one source to many destinations for the support of interactive services. These applications (which are also well suited to accommodate vertical industries' needs) require ubiquitous connectivity as well as continuous service availability.

Most importantly, 5G will support the unprecedentedly stringent throughput and latency demands of future cellular networks, and will play a key role for the automotive sector towards the development of future Connected Intelligent Transportation Systems (C-ITSs) and Connected and Autonomous Vehicles (CAVs), as we will discuss in the following sections.

1.3 5G Cellular Networks

As of July 2019, while technical standard development for 5G radio services is still an ongoing process, 293 operators in 98 countries have demonstrated, are testing, or have started deploying working 5G commercial cellular networks (the equivalent numbers of 5G investments in November 2018 were 192 operators in 81 countries) [7]. The wide diversity of technology drivers and use cases of future society will not make 5G cellular deployments just a simple evolution of 4G networks with new spectrum bands, higher spectral efficiencies and higher peak throughput, but also target new services and business models. Several standardization bodies are involved in the definition of the forthcoming 5G cellular system. The most important one is the 3rd Generation Partnership Project (3GPP) which recently defined a new Radio Access Technology (RAT), i.e., 3GPP NR[2], that introduces novel designs and technologies that will comply with the 5G requirements. NR has been standardized by 3GPP with a first set of specifications[3] (Release 15) in December 2017 and a complete one published in June 2018. Release 16 for NR is expected to be completed in December 2019 and will be composed of a set of specifications that match the ITU 5G requirements previously described in Sec. 1.1.

1.3.1 5G Cellular Networks Standardization: 3GPP NR

Their main characteristic of NR is flexibility: the specifications, indeed, provide a general technology framework that addresses different and, in some cases, conflicting 5G requirements [10] and is forward compatible, to accommodate future applications and use cases. Fig. 1.3 shows the main novelties of NR with respect to LTE. For the Radio Access Network (RAN), as illustrated in the left part of Fig. 1.3, NR features (i) a more flexible frame structure, providing a choice of physical-layer parameters that enable support for a wide variety of use cases; (ii) the support for a much larger spectrum, with frequencies also in the Millimeter Wave (mmWave) band (up to 52.6 GHz); and (iii) the design of Physical (PHY) and Medium Access Control (MAC) layer procedures for beam management [9, 11]. In the core network, instead, the 5G Core (5GC) introduces network slicing and a higher level of flexibility and virtualization with respect to the traditional LTE Evolved Packet Core (EPC). Moreover, different deployment options and inter-networking with LTE are supported, as

[2]While NR was originally meant as the acronym for "New Radio" [8], according to the latest 3GPP specifications [9] it has lost its original meaning and it now refers to the 5G Radio Access Network.

[3]The specifications for NR are in the Technical Specification (TS) of 3GPP *38 series*, together with Technical Reports (TRs) that contain related studies. Other relevant RAN specifications can be found in the 36 (LTE) and 37 (LTE-NR inter-networking) series.

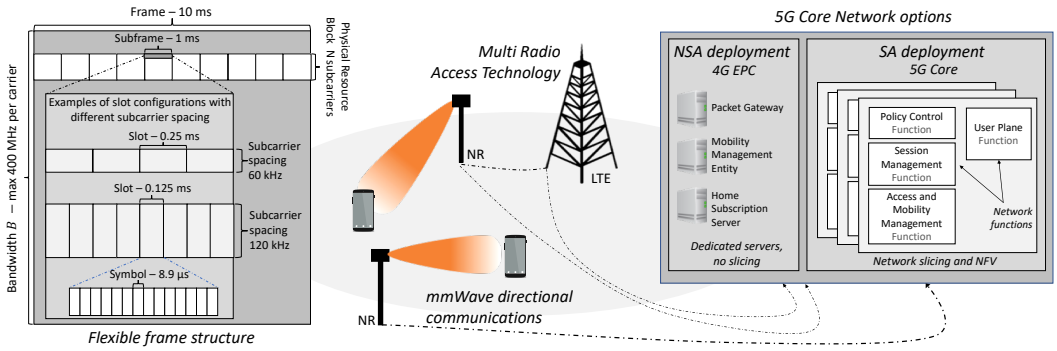


Fig. 1.3: Key novelties and options of the 5G architecture: flexible frame structure, multi RAT, mmWaves and core network options.

shown in the right part of Fig. 1.3. In the following paragraphs, we will introduce the main novelties related to the NR frame structure and the 5G deployment architectures [12, 13].

5G NR Frame Structure The main characteristics of the frame structure can be found in [11]. The waveform is Orthogonal Frequency Division Multiplexing (OFDM) with a cyclic prefix, and the available resources are organized in a time and frequency grid, with OFDM symbols and subcarriers representing the minimum units in time and frequency. As in LTE, and as shown in Fig. 1.3, 14 symbols are grouped in one slot, slots in subframes (1 ms) and subframes in frames (10 ms). However, NR's main novelty is flexibility: it supports multiple numerologies, i.e., sets of parameters for the OFDM waveform, also multiplexed in time and frequency, as long as they are aligned on a subframe basis. This enables the different 5G use cases: for example, a higher subcarrier spacing supports high-data-rate traffic, while a lower subcarrier spacing enables low-frequency narrowband communications. The maximum bandwidth for each carrier is 400 MHz, with the option to aggregate up to 16 carriers. In time, each subframe is composed by 2^n slots, with n ranging from 0 to 4. This parameter controls also the subcarrier spacing Δ_f, equal to 15×2^n kHz. For frequencies above 6 GHz, the minimum value of n is 2, therefore the minimum subcarrier spacing is 60 kHz. The left part of Fig. 1.3 shows an example of NR frame structure with different subcarrier spacings ($\Delta_f = 60$ and 120 kHz) and number of slots per subframe (4 and 8, respectively).

The NR frame structure also supports ultra-low-latency communications. First, with *mini-slots*, data transmission does not need to be allocated synchronously with respect to the beginning of a standard slot, thus it does not need to wait for the next slot. Moreover, a subframe can be self-contained, i.e., an exchange with a first transmission and the corresponding acknowledgment can be completed in 1 ms. Therefore, NR supports sub-ms latency for acknowledged transmissions.

5G NR Network Deployment The flexibility provided by the 3GPP specifications extends to the possible deployment architectures and interconnectivity between 4G and new 5G networks, as shown in Fig 1.3. In particular, in order to smooth the transition between the different generations and reuse the widely deployed LTE and EPC infrastructure, the NR specifications foresee a non-standalone (NSA) deployment, in which the network operator does not use the new 5GC, but only deploys NR Next Generation Node Bases (gNBs)[4] which are connected to EPC. Alternatively, in a standalone (SA) deployment, both the RAN and the core network follow the 5G specifications, which, as Fig. 1.3 illustrates, support network slicing and follow the Network Function Virtualization

[4]Notice that gNB is the NR term for a base station.

(NFV) paradigm [14].

With NR, the gNB can also be split into separate physical units, i.e., the Distributed Unit (DU), which contains the lower layers of the protocol stack and is deployed in the field, and the Central Unit (CU) incorporating complete stack functionalities, which can be co-located with the DU or hosted in a data center facility. As discussed in [15], this allows network operators to deploy the 5G RAN according to the use cases they want to serve, e.g., an ultra-dense small cell deployment with low utilization but high peak rate can rely on the CU/DU split to maximize the multiplexing and enable a centralized control of the RAN, while a rural low-density deployment for the support of Internet of Things (IoT) applications can feature complete gNB nodes. Finally, the 5G core network has been redesigned with respect to the 4G core following a service-based approach [14]: the 5G core is composed of multiple network functions, that provide mobility, authentication and routing support, that can be dynamically instantiated in data centers according to the load and traffic demands of the network. For example, while in LTE/EPC networks the control plane for the mobility of the user was handled by a single server (e.g., the Mobility Management Entity (MME)), with the 5GC multiple network functions concur to offer the same set of services, but can be deployed in different data center locations and quickly turned off and on to decrease resource utilization.

Moreover, the 5GC supports network slicing [16], i.e., the resources of the network can be split to serve different portions of traffic, that have different QoS requirements (e.g., IoT and mobile broadband traffic). The service-based 5GC architecture is an important enabler of network slicing in 5G, given that network functions can be provisioned dynamically to serve new network slices without the need to use separate servers, as would happen with the EPC.

1.4 5G Vehicular Networks

Over the past few decades, advances in the automotive industry have opened up the potential for CAVs as a means to offer safer and more efficient driving. From a safety perspective, CAVs can eliminate or mitigate the severity of traffic accidents, which also account for around 25% of traffic congestion, since more than 90% of accidents are due to human error [17]. In addition, connected cars can guarantee improved traffic management through smart platooning, adaptive and cooperative cruise control, lane change assistance, and traffic light coordination. In particular, CAVs can lead to more than 50% highway capacity increase at 80% market penetration, resulting in an estimated 10% reduction in infrastructure investment and 4 to 10% improved fuel economy [18]. The transition to autonomous, computer-controlled vehicles is also expected to contribute to a 60% fall in emissions, with positive implications on the environment [19]. The hands-free driving environment of CAVs can finally reduce drivers' stress and tedium, as well as increase their productivity (CAVs could save over 2.7 billion unproductive hours in the US annually in work commutes, according to some estimates [20]). Overall, in the US alone, the total associated monetized annual market resulting from CAVs is estimated in a huge 1.3 trillion USD, or 8% of the entire US GDP, thereby stimulating research efforts towards fully autonomous vehicles [21].

CAVs, when fully commercialized, will have ever more stringent regulations in terms of road safety and traffic efficiency [23]. In this regard, new use cases (and relative requirements) specific to future vehicular services have been defined in [22], as summarized in Fig. 1.4.

- *Platooning.* It refers to the set of services that enable the vehicles to cooperatively travel in close proximity to one another at highway speeds. The data rate ranges from a few Kbps up to 65 Mbps depending on whether sensor sharing is required, while the latency ranges from 10 ms to 500 ms depending on the distance. Vehicle platooning poses also very strict requirements

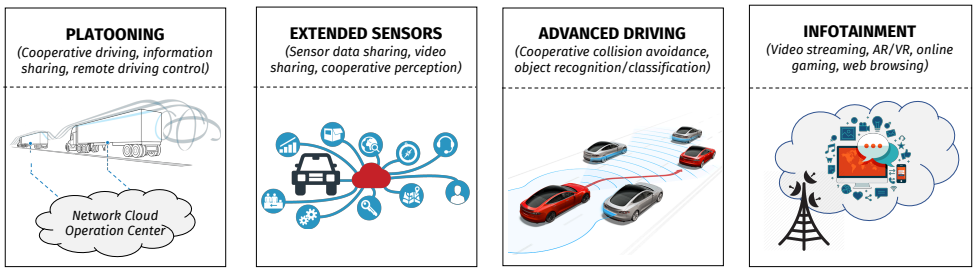


Fig. 1.4: Representation of the use cases specified by the 3GPP in [22] for future CAVS systems.

in terms of connection reliability.

- *Extended Sensors.* It enables the exchange of raw or processed data gathered through local sensors, thereby enhancing the perception range of the vehicles beyond the capabilities of their on-board instrumentation. The data rate demands are proportional to the resolution of the acquired sensory data and range from around 10 Mbps for a 300-beam 32-bit LIDAR up to approximately 1 Gbps for high-quality uncompressed camera images [24]. Due to the sensitive nature of the exchanged information, the maximum tolerable latency varies from approximately 3 ms up to 100 ms for lower degrees of automation.
- *Advanced Driving.* It enables semi- or fully-automated driving through persistent dissemination of perception data. The required throughput is relatively low (i.e., less than 50 Mbps), while latency must be very small (i.e., less than 100 ms for high degree of automation) to ensure prompt reactions to unpredictable events.
- *Infotainment.* It generically refers to a set of services that deliver a combination of information and entertainment (e.g., video streaming, media download applications, Internet browsing). Data rate requirements are in the order of hundreds of Mbps, and the dynamic maintenance of multicast communication may be required.

In this perspective, the potential of future CAV deployments can be fully unleashed through wireless communications to and from roadside infrastructures and among vehicles, a concept that is usually referred to as Vehicle-to-Everything (V2X) connectivity. Today, the two key technologies that enable V2X communications are IEEE 802.11p [25] and Cellular-V2X (C-V2X) [26] – an LTE-based RAT – that, however, fall short of providing the desired QoS performance of future V2X use cases. Therefore, going forward into the 5G era, different standardization activities are currently being promoted as a means to overcome current technology limitations, as described in the following subsections [27].

1.4.1 5G Vehicular Networks Standardization: 3GPP NR-V2X

3GPP V2X related specification so far has been focused on enhancements to LTE, which however may not be able to satisfy the unprecedentedly stringent demands (e.g., in terms of latency and throughput) of envisioned vehicular services [29]. However, 3GPP has an agreed roadmap to support V2X as part of its ongoing 5G effort, i.e., within the so called NR-V2X work item for Release 16 [28,30–32]. These developments will include (i) more sophisticated channel models, (ii) sidelink and network architecture improvements, (iii) enhancement of legacy communication standards for V2X transmissions, (iv) RAT selection support, and (v) dynamic QoS management. A comparison between the C-V2X and NR-V2X key features is presented in Table 1.1.

Table 1.1: Comparison of C-V2X and NR-V2X [28].

Feature	C-V2X	NR-V2X
Communication types	Brodacast	Broadcast, Unicast, Groupcast
Modulation and Coding Scheme (MCS)	QPSK, 16-QAM	QPSK, 16-QAM, 64-QAM
Waveform	SC-FDMA	OFDM
Re-transmissions	Blind	HARQ
PHY Channels	PSCCH, PSSCH	PSCCH, PSSCH, PSFCH
Control and data multiplexing	FDM	TDM
Subcarrier spacing	15 kHz	Flexible (up to 120 kHz)
Scheduling interval	One sub-frame	Slot, mini-slot, multi-slot
Carrier frequency	Sub-6 GHz	Sub-6 GHz, mmWaves

Sidelink Modes Like C-V2X, NR-V2X will support two sidelink modes. The NR-V2X *sidelink mode 1* defines mechanisms that allow direct vehicular communications within gNB coverage, which allocates resources to the vehicles and controls users' operations. The NR-V2X *sidelink mode 2*, instead, supports direct vehicular communications in the out-of-coverage scenario and vehicles autonomously determine sidelik transmission resources within pre-configured resource polls. Moreover, while C-V2X only provisions support for broadcast transmissions, NR-V2X enables broadacst, unicast and groupcast (i.e., communication with only a specific sub-set of vehicles in the sender's vicinity) transmission services, e.g., for a platoon leader that communicates with its platoon members using the groupcast mode as well as with vehicles that are not part of the platoon using the broadcast mode.

PHY Enhancements A large number of NR-V2X use cases are based on reliable and high-throughput delivery of messages. To support such services, 3GPP investigates, as for cellular transmissions, general aspects of NR operations in the mmWave band above 52.6 GHz (e.g., target spectrum range, use cases, regulatory requirements) [33]. NSA functionalities (i.e., the synergistic orchestration among different radios, including coexistence of C-V2X and NR-V2X within a single device) are also being discussed to deliver more flexible and resilient transmissions.

At the PHY layer, NR-V2X systems support a flexible numerology, as introduced in 3GPP Release 15 for cellular scenarios (see Sec. 1.3). *Mini-slots* are also possible, where vehicles that have latency-critical messages to send can occupy any number of OFDM symbols within the slot. Furthermore, slot-aggregation, i.e., combining two or more slots to form a *multi-slot*, will also be supported in NR-V2X to cater to use-cases that require exchange of large-sized packets [28]. C-V2X also provides support for re-transmissions, to increase the communication reliability, even through these re-transmissions are blind, i.e., the source, if configured, re-transmits without knowing if the initial transmission has been received by surrounding destination nodes. Such blind re-transmissions are resource inefficient if the initial transmission is successful. Blind re-transmissions are also ineffective if more than two transmissions are required for a given reliability requirement. NR-V2X will thus introduce a new Physical Sidelink Feedback Channel (PSFCH), which enables feedback-based re-transmissions and channel state information acquisition [34].

Other PHY layer enhancements include the use of Low-Density Parity-Check (LDPC) coding, higher order Modulation and Coding Scheme (MCS) including 64-QAM, and a flexible number of DeModulation Reference Signal (DMRS) symbols per slot [28].

Resource Scheduling Operations 3GPP has began evaluating the following schemes:

Table 1.2: Comparison of IEEE 802.11p and IEEE 802.11bd [28].

Feature	IEEE 802.11p	IEEE 802.11bd
Channel coding	BCC	LDCP
Re-transmissions	None	Congestion dependent
Doppler management	None	Midambles
Subcarrier spacing	156.25 kHz	Flexible (down to 78.125 kHz)
Supported relative speeds	252 km/h	500 km/h

- For sidelink mode 1, a gNB schedules resources to be used by the UE for sidelink transmission(s) (this is similar to C-V2X PC5 mode 3).
- For sidelink mode 2, the UE determines, i.e., a gNB does not schedule, sidelink transmission resources within sidelink resources configured by a gNB or pre-configured by the network. The work item considers four sub-modes [32][5]:

  – *Mode 2 (a)*: Each UE autonomously selects its resources (this mode is similar to C-V2X PC5 mode 4) by sensing and analyzing the channel occupation time during the sensing window, and using this information to identify the most appropriate set of resources.

  – *Mode 2 (b)*: UEs assist other UEs in performing resource selection. The UE providing assistance can be the receiver UE, which can potentially notify the transmitting UE of its preferred resources using the PSFCH.

  – *Mode 2 (c)*: UEs use pre-configured sidelink grants to transmit their messages. This sub-mode will be facilitated through the design of two-dimensional time-frequency patterns.

  – *Mode 2 (d)*: a UE performs resource allocation for a group of UEs in its vicinity, thereby allowing a significant reduction in the number of collisions between group member UEs. This sub-mode is especially useful in applications that require groupcast or unicast transmissions, e.g., for platooning [36]. The UE performing resource allocation for other UEs within the group is referred to as the *scheduling UE* (S-UE) [37]. The mechanism to select a UE as the S-UE is still under study, and some possible options include selection based on geo-location or pre-configuration [38].

1.4.2 5G Vehicular Networks Standardization: IEEE 802.11bd

IEEE developed the first V2X standard, i.e., IEEE 802.11p, in 2010, to assist basic vehicular safety and better traffic management. Since then, however, advanced PHY and MAC techniques, introduced in 802.11n/ac/ax, stimulated researchers to create the IEEE 802.11bd Task Group to promote IEEE 802.11p enhancements targeting future V2X application requirements [39]. Although technical details have not yet been throughly discussed, IEEE 802.11bd is positioned to support twice the throughput and the communication range of 802.11p, with relative velocities up to 500 km/h, through (i) new transmission mechanisms, (ii) dual carrier modulation, (iii) re-designed PHY and MAC features, and (iv) support of high-frequency operations. A comparison between the IEEE 802.11p and IEEE 802.11bd key features is presented in Table 1.2.

Transmission Mechanism Improvements In 802.11p systems, preambles are used for initial channel estimation. However, for fast-varying channels, the initial estimate may quickly become

[5]In the latest 3GPP meetings, it has been agreed to no longer support modes 2(b) and 2(c) as separate sub-modes [35].

obsolete. 802.11bd, in turn, proposes to use *midambles* as a means to better handle high Doppler shifts experienced in high-mobility environments. Unlike preambles, which are allocated at the beginning of the frame, midambles will be introduced in between the OFDM symbols with appropriate frequency, and will be used in channel tracking so that an accurate channel estimate is obtained for all symbols [28]. Furthermore, it is possible to repeat certain signaling fields of the preamble to achieve higher range and robustness [40]. Finally, new re-transmission methods have been discussed by the IEEE 802.11bd Task Group to increase packet reliability. In particular, 802.11bd proposes an adaptive re-transmission scheme where decisions to re-transmit a frame and the number of re-transmissions are based on the congestion level [41].

PHY Enhancements IEEE 802.11p PHY layer is OFDM-based with 64 subcarriers, typically with a subcarrier spacing of 156.25 kHz. To increase the OFDM efficiency, IEEE 802.11bd Task Group members are exploring the use of narrower OFDM numerologies such that the number of subcarriers is increased while still occupying a 10 MHz channel [42]. The design of alternate OFDM numerologies must, however, take the maximum relative velocities into consideration, to avoid that fast-fading channel variations result in inter-carrier interference [28]. 802.11bd will also support Dual Carrier Modulation (DCM) techniques, that allow transmitting the same symbols twice over sufficiently far-apart subcarriers such that frequency diversity is achieved [43]. Because each symbol transmission is repeated over two different subcarriers, the modulation order must be doubled (e.g., from QPSK to 16-QAM) to maintain the throughput while improving block-error-rate performance.

IEEE 802.11p devices operate in the 5.9 GHz band and count on a total spectrum of 75 MHz. IEEE 802.11db systems, instead, support operations in the above 6-GHz bands, including therefore the mmWave frequencies [44]. At the current stage, communications at 60 GHz are included in the scope as an optional feature.

Other PHY layer enhancements include the use of LDPC coding and multiple antennas to increase the reliability using spatial diversity or increase the throughput using spatial multiplexing [45].

1.5 5G Technologies and Challenges

The wide diversity of technology drivers and use cases is a unique characteristic of the 5G paradigm, whose potential will be fully unleashed only through cornerstone technological advancements and novel network designs [12]. In particular, 5G encompasses new architecture developments to boost wireless capacity through (i) new frequency bands (e.g., the mmWave spectrum up to 100 GHz), (ii) advanced spectrum usage and management, (iii) new techniques such as massive Multiple Input Multiple Output (MIMO), (iv) seamless integration of licensed and unlicensed bands, as well as multi-radio access technology orchestration, and (v) a new core network design offering network slicing and virtualization. These concepts will be further studied and discussed in Chapter 2.

In particular, communications at very high frequency (up to 100 GHz), including standalone operations in the mmWave bands, will enable 5G performance demands by delivering orders of magnitude higher bit-rates than legacy cellular networks [10, 46, 47]. Even though the standardization is moving full pace ahead towards first 5G deployments operating at mmWaves, there are still various questions to be answered for the proper design of both cellular and vehicular networks, and many topics are still open for long-term research.

PHY Layer From a PHY-layer perspective, even though the literature provides valuable insights into the propagation characteristics of mmWave signals, there remain some open problems which call for more accurate channel modeling, including:

- Characterization of second order statistics: The lack of temporally correlated channel measurements in the mmWave band significantly limits the level of detail that can be achieved in simulations, and prevents the applicability of existing channel models to dynamic scenarios.
- Characterization of vehicular channel. Current channel models are derived from cellular measurement campaigns which might not be fully representative of a vehicular system due to the more challenging propagation characteristics of highly mobile vehicular nodes.
- Characterization of scenario-dependent fading statistics. Fading is vital to describe the fluctuations of the received power over time. Most existing measurements lack characterization of the correlation among signals in a multipath environment, e.g., the role played by ground reflection is often underestimated, especially in a mobile setting.
- Characterization of directional scenarios. The effects of directional transmissions (through MIMO technologies) have not been numerically characterized by currently available channel measurements, which make use of isotropic antennas with the assumption of unit gain or of horn antennas with fixed pointing direction. It has also been demonstrated that the delay spread decreases with narrow beams, but measurements are lacking [48, 49].

MAC Layer Signals propagating in the mmWave band suffer from increased pathloss and severe channel intermittency, and are blocked by many common materials such as brick or mortar [50, 51], and even the changing position of the body relative to the mobile device can lead to rapid drops in signal strength (see Chapter 2 for more details). This requires the establishment and maintenance of highly-directional transmission links, to benefit from the resulting beamforming gain and sustain an acceptable communication quality. From a MAC-layer perspective, fine alignment of the transmitter and receiver beams has important implications for the design of a variety of control tasks, including (i) Initial Access (IA) [52] for idle users, which allows a mobile UE to establish a physical link connection with a gNB, and (ii) beam tracking, for connected users, which enables beam adaptation schemes, handover, path selection, and radio link failure recovery procedures [53]. In current legacy systems, these procedures are performed using omnidirectional signals while, in the mmWave bands, it may be essential to exploit the antenna gains even for control operations. In the vehicular context, these challenges are further exacerbated considering a highly dynamic scenario, as the beam alignment may be lost before a data exchange is completed [54]. Also, the increased Doppler effect experienced at high speed could make the assumption of channel reciprocity not valid and could impair the feedback over a broadcast channel (e.g., during synchronization).

Optimizing resource allocation in a multi-user high-rate mmWave directional scenario poses another interesting challenge. On one hand, typical Reference Signal Received Power (RSRP)-based association can lead to inefficient use of the network resource, thereby resulting in overload of the transmission links. At the same time, mmWave systems become noise-limited [55], thereby making legacy resource allocation methods, which are suited for an interference-limited homogeneous regime, ineffective. This calls for the design of fair and efficient attachment policies, specifically tailored to the characteristics of the above-6 GHz bands.

TCP/IP Layer In the mmWave context, the propagation characteristics and the directional nature of above-6 GHz links bring several challenges for network protocol design. For instance, due to the presence of communication blockages, the shortest path connecting two network nodes (in terms of geographical or topological distance) is not necessarily the best, and may actually yield lower throughput and higher packet loss than a longer path. It is thus important to make a judicious selection of relaying nodes, for example trying to keep the number of hops to a minimum when using multi-hop communications to overcome an impaired direct path.

Another relevant issue is the performance analysis of transport protocols, especially congestion control using the Transmission Control Protocol (TCP). First, standard slow start mechanisms can take several Round Trip Times (RTTs) to achieve the full throughput offered by the mmWave physical layer, increasing the latency of the communication. Second, sudden drops in the data rate, which are likely to occur in LOS/NLOS transitions, can result in very large queuing at the nodes, dramatically increasing the packet drop probability. Third, after a retransmission timeout, even aggressive TCP protocols (e.g., `Cubic`) can take inordinately long to recover the full data rate [56].

Network Architecture At mmWaves, the combination of the high propagation loss and the blockage phenomenon calls for a high-density deployment [57]. Such ultra-dense deployment can be costly for network operators due to (i) the need to interconnect a large number of small cells through wired backhaul links, (ii) the significant capital and operational expenditures for the deployment of the base stations (which are not scalable with the number of nodes), (iii) increased traffic control and signaling load derived from frequent handover and radio link failure events. In this perspective, how to forward massive wireless traffic from/to the core network in a low cost and energy efficient way is still an open issue that deserves further discussion.

Base station densification in 5G calls also for the design of an energy efficient wireless network [58]. At mmWaves, in particular, the maintenance of multi-Gbps wireless links will have a major impact on end users' battery life, thereby posing energy, complexity and cost issues.

1.6 Thesis Structure

Given the above introduction, in this thesis we study the challenges pertaining to both future cellular and vehicular networks operating at mmWaves, compare all the existing solutions, highlight pros and cons, and propose new valid approaches. More specifically, the thesis is organized as follows.

- In Chapter 2 (based on **[B3]**) we overview some of the key enabling technologies that could lead to fundamental changes in the design of 5G networks, including the use of (i) the mmWave spectrum to enable the foreseen application performance demands, (ii) the MIMO technology to improve the reliability and spectral efficiency of future 5G networks, and (iii) multi connectivity (in non-standalone deployments) to improve communication robustness while increasing network capacity.
- In Chapter 3 (based on **[J1, J5, C4]**) we study, analyze and compare some possible implementations of initial access techniques for 5G cellular networks, where we argue that directionality should be used also in the initial synchronization phase. In particular, we focus on the solutions that have been recently proposed by the 3GPP in its Release 15, as part of NR specifications. We show that there exist trade-offs among better detection accuracy, improved reactiveness and reduced overhead.
- In Chapter 4 (based on **[J2, J3, J5, J6, C2, B1]**), after a brief overview of the proposed frame structure and reference signals in 3GPP NR, we focus on the settings for communication at frequencies above 6 GHz. In particular, we describe several beam management and handover procedures according to different network architectures (standalone and non-standalone) and signal transmission directions (downlink or uplink). Specifically, we provide insights and guidelines for determining the strategies in different mmWave network deployments, according to the need of the network operator and the specific environment in which the nodes are deployed. We finally demonstrate that multi-connectivity is able to improve the handover performance of an end-to-end network with mmWave access links with respect to several

metrics, including latency, throughput (in terms of both average and stability), radio control signaling and packet loss.

In Chapter 5 (based on **[J7, C7, C8]**) we investigate Integrated Access and Backhaul (IAB) solutions as an approach to relay access traffic to the core network wirelessly, thereby removing the need for all base stations to be equipped with fiber backhaul. We review the characteristics of IAB that are currently being standardized in 3GPP NR Release 16 and evaluate the performance of IAB networks for different applications and traffic types such as Internet browsing and video streaming. We show that IAB represents a viable solution to efficiently relay cell-edge traffic, although the benefits decrease for more congested networks. We also highlight the limitations of the IAB paradigm and provide guidelines on how to overcome them.

- In Chapter 6 (based on **[J4, J9, C1, C3, C5, C6, C9, C12,C14, C16, B2]**) we focus on vehicular network operations and provide an end-to-end performance comparison between the different radio technologies that are currently discussed for V2X deployments (including LTE, IEEE 802.11p and mmWaves). The impact of several automotive specific parameters (i.e., the gNB density, the vehicular scenario and the application data rate) are investigated in terms of experienced throughput, communication latency and fairness. We conclude that, although legacy communication systems deliver a good compromise between fairness and low latency, the combination of massive bandwidth and spatial degrees of freedom has the potential for mmWave systems to meet some of the boldest requirements of next-generation transportation systems, including high peak per user data rate and very low latency, both in urban and high-mobility highway scenarios. We conclude that the end-to-end communication performance can be improved by using multiple radios in parallel (i.e., *hybrid networking*), to complement the limitations of each type of network and deliver more flexible and resilient transmissions.
- In Chapter 7 (based on **[J10, C10, C11, C13, C15, P1]**) we propose to use Value of Information as an essential tool to discriminate the importance of the different information sources in vehicular networks. In particular, we study the trade-off between ensuring accurate position information and preventing the congestion of the communication channel, and design an innovative threshold-based broadcasting algorithm that forces vehicles to distribute state information if the estimated positioning error is above a certain error threshold. We show through simulations that the proposed approach outperforms a conventional broadcasting strategy, which relies on a periodic transmission of state information and channel sensing, since it reduces the positioning error with no additional resources. Moreover, we propose a method that quantifies the expected Value of Information based on time, space and quality inter-dependencies, and evaluate the impact of the operating distance, the type of observation, the type of sensor, the propagation scenario and the age of information on the value assessment operations.
- In Chapter 8 (based on **[J8]**) we present an overview of the applications and the technologies that may characterize future 6G networks. To this aim, we consider several potential 6G use cases (including massive scale connectivity, truly indoor coverage, eHealth, robotics and unmanned applications) and attempt to identify the most promising technologies, architectures and deployment models that can provide the basis for 6G systems. We expect 6G networks to adopt new spectrum bands, combining advancements throughout the whole network stack, from circuit and antenna design to network architectures, protocols and artificial intelligence.
- In Chapter 9 we conclude this thesis with suggestions for future research.

Chapter 2

Millimeter Waves and 5G Technologies

As introduced in Chapter 1, 5G will not simply be an evolution of earlier cellular standards. On the contrary, 5G will foster the 4th industrial revolution by supporting, besides enhanced mobile broadband and massive machine-type communications, new use cases with unprecedented stringent demands in terms of ultra-high data rates, ultra-low latency and support for a massive number of connections. To meet these requirements, 5G encompasses novel architecture developments and design options including [59] (i) new frequency bands (e.g., the *Millimeter Wave (mmWave)* spectrum up to 100 GHz), (ii) new techniques such as massive *Multiple Input Multiple Output (MIMO)*, and (iii) seamless integration of different frequency bands through *Multi-Connectivity (MC)*. In this chapter we will overview the characteristics of these key innovations.

2.1 Millimeter Waves: the Next Spectrum Frontier

The mmWave spectrum – roughly above 10 GHz[1] – has rapidly emerged as an enabler of the 5G performance demands in micro and picocellular networks [60]. These frequencies, combined with high-order modulation, offer much more bandwidth than 4G/LTE systems operating in the congested bands below 6 GHz, and initial capacity estimates have suggested that networks operating at mmWaves can offer orders of magnitude higher bit-rates than legacy cellular networks. The physical size of antennas at mmWave frequencies is so small that it becomes practical to build complex antenna arrays to realize directional communications, thereby obtaining high antenna gains through beamforming. Directional communications, at the same time, tend to isolate the users and deliver reduced interference. Security and privacy are also inherently improved because of the short-range transmissions which are typically established [61]. Recognizing the potential of mmWave for wireless applications in the 5G context, the Federal Communications Commission (FCC) has proposed to authorize the 27.5–28.35 GHz band and the 38.6–40 GHz band for mmWave services subject to county-sized geographic area licenses [62]. Similarly, the European Commission (EC) has licensed the 24.25–27.5 GHz frequency band for initial 5G trails [63]. 3GPP and IEEE are also supporting, for

[1] Although strictly speaking mmWave bands include frequencies between 30 and 300 GHz, industry has loosely defined it to include any frequency above 10 GHz.

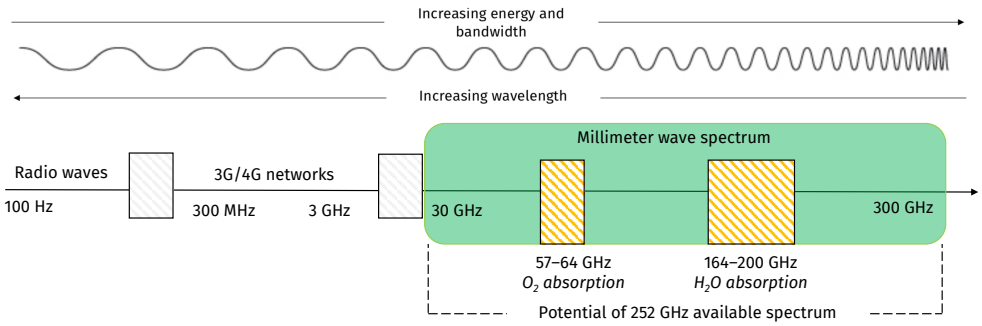


Fig. 2.1: The millimeter wave spectrum.

the first time, frequencies above 6 GHz, including therefore mmWave bands [64], for both cellular and vehicular network operations, as described in Secs. 1.3 and 1.4.

Despite these promising features, communication at mmWave introduces new challenges for the whole protocol stack, which may have a significant impact on the overall end-to-end system performance, as explained in the next subsection.

2.1.1 Challenges of the Millimeter Wave Spectrum

Propagation Loss. Propagation loss is given by Friis' free-space formula, i.e.,

$$P_{\text{rx}} = P_{\text{tx}} G_{\text{tx}} G_{\text{rx}} \left(\frac{\lambda}{4\pi d} \right)^2, \tag{2.1}$$

where P_{tx} and P_{rx} are the transmit and receive powers, respectively, G_{tx} and G_{rx} are the transmit and receive antenna gains, respectively, λ is the wavelength and d is the distance between the transmitter and the receiver. A direct result of (2.1) is that the receive power scales with λ^2 and hence, for mmWaves, due to the high frequency and, consequently, to the small wavelength, the communication suffers from severe path loss, thereby preventing long-range omnidirectional transmissions.

Penetration Loss. The poor diffraction capability and high penetration loss make blockage from common material (e.g., brick and mortar) an important effect in mmWave propagation [50]. For example, it was observed that there is a received power difference of more than 40 dB at 28 GHz and 73 GHz when a mobile receiver goes around a building corner [47]. Even changes in the position of the body relative to the mobile device can lead to rapid drops in signal strength. In particular, Raghavan *et al.* [65] have experimentally captured the impact of the hand and the human body on the mmWave signal propagation, showing that a median loss of 15 dB is incurred by the hand even in the most pessimistic scenario of a hard hand grip. The authors also proved that the time-scales at which the mmWave signals are disrupted by blockage are on the order of a few hundreds of ms or more. High attenuations for certain materials might even result in the mmWave signals transmitted from outdoor base stations being confined to streets and other outdoor structures. The indoor coverage must thus be provided by other means such as using indoor Wi-Fi or femtocells.

Environmental Loss. Signal propagation at mmWaves is affected by weather absorption [51]. Raindrops, for instance, are roughly the same size as the mmWave wavelengths and, therefore, cause scattering of the radio signal. Under heavy rain, the attenuation is in the 8-18 dB/km range. Fortunately, the most intense rain tends to fall in selected countries of the world and happens in short bursts, while light rain yields just a little attenuation. Foliage loss is another limiting propagation

impairment at mmWave frequencies. At 80 GHz and 10 meters foliage penetration, for example, the loss can be as severe as 23.5 dB, which is about 15 dB higher compared to sub-6 GHz channels.

Atmospheric Loss. MmWave signals are absorbed by gas molecules to some extent, even though the absorption level depends on the frequency [61]. While the attenuation under normal atmospheric conditions is in the range between 7–15.5 dB/km, oxygen (O_2) and water vapor (H_2O) absorption peaks at 60 GHz and 180 GHz and causes an overall attenuation of 15 dB/km and 20–30 dB/km, respectively, thereby resulting in severe attenuation even at short distances.

Delay Spread and Doppler Spread. Delay and Doppler spreads are important characteristics of the mmWave propagation channels. The delay spread is a measure of the propagation delay of each multipath component and determines the severity of intersymbol interference. It was observed in [66] that large delay spreads are rare for link distances greater than 50 m. The Doppler spread refers to the difference in the perceived frequency of a traveling wave from its true frequency, and is closely related to the coherence time, which determines the temporal selectivity of the channels. Its effect increases with the carrier frequency and the speed of the mobile terminals, and may lead to network disconnections [67].

To deal with these impairments, mmWave networks must provide a set of mechanisms by which the endpoints establish highly directional transmission links to benefit from the resulting beamforming gain and sustain an acceptable communication quality. This is typically achieved using high-dimensional phased arrays, as described in the next section.

2.2 Massive MIMO: a Key Component of 5G Systems

While the combination of extreme cell densification, increased system bandwidth through mmWaves, and more flexible spectrum usage (e.g., by resource sharing) represents a feasible and sustainable solution to meet 5G performance requirements, MIMO techniques have also emerged in modern wireless networks to improve reliability and spectral efficiency. The main concept is to use multiple transmit and receive antennas to exploit multipath propagation. Among the possible antenna array designs, the most suitable approach is the use of Uniform Planar Arrays (UPAs) where the antenna elements are evenly spaced on a two-dimensional plane and a 3D beam can be synthesized by adapting both azimuth and elevation planes, as illustrated in Fig. 2.2.

Depending on the channel properties, MIMO systems can be configured for [68, 69]:

- *Spatial diversity*, i.e., sufficiently separated antennas are used to transmit redundant versions of the same message over multiple paths. The quasi-independent fading characteristics of the channel are thereby exploited to make links more robust and decrease the outage probability.
- *Beamforming*, i.e., multiple antenna elements are adaptively phased to form a concentrated beam pattern towards a specific direction. Beamforming provides significant array gains, thereby guaranteeing increased Signal to Noise Ratio (SNR) (since propagation path loss is mitigated) and reduced co-channel interference (resulting from the spatial selectivity of the directional antenna).
- *Spatial multiplexing*, i.e., an outgoing signal is split into multiple independent streams, transmitted simultaneously and in parallel on the same channel through different antennas.

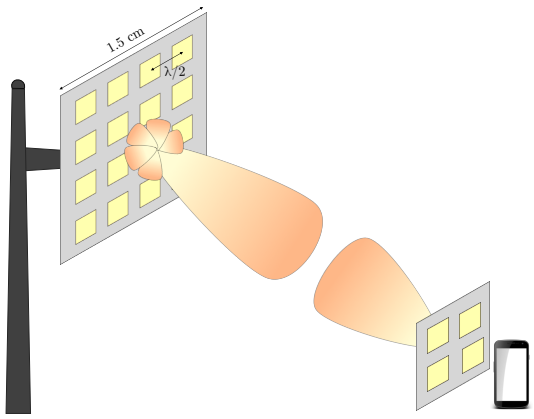


Fig. 2.2: Illustration of an UPA MIMO array. At the base station side, the array (which has dimension of roughly 1.5 cm $\times$1.5 cm) is comprised of 4×4 elements, at the user terminal side the array has 2×2 elements. The antenna element radiation pattern is modeled as a patch antenna element with horizontal and vertical spacing equal to $\lambda/2$.

Moreover, multi-user MIMO (MU-MIMO) can be implemented through *Spatial Division Multiple Access (SDMA)*, in which the multipath properties of the channel are used to multiplex users in the spatial dimension while operating in the same time-frequency resource.

Typical current MIMO installations use relatively few (i.e., less than 10) antennas, and the corresponding improvement in spectral efficiency has been relatively modest [70]. When combined with mmWave propagation, instead, the full potential of the MIMO paradigm can be truly unleashed. In fact, the physical size of antennas at mmWave frequencies makes it practical to build large antenna arrays (e.g., with 100 or more elements), thereby scaling up the network performance by possibly orders of magnitude compared to state of the art MIMO implementations. The concept of using a number of antennas in network nodes which is much higher than the number of users is usually referred to as *massive MIMO* [71]. For mmWave transmissions, massive MIMO is mainly used for beamforming while, at sub-6 GHz, it provides *channel hardening*, i.e., the combined usage of a massive number of antennas decreases the channel variability by averaging the small-scale fading [72].

However, massive MIMO comes with its own set of challenges, mainly related to:

- *Hardware impairments:* massive MIMO systems exploit channel reciprocity to estimate the channel responses on the uplink and use such information for both uplink and downlink transmissions. Since the transceiver hardware is generally not reciprocal, calibration is needed to exploit the channel reciprocity in practice.

- *Energy-consumption vs. flexibility trade-off:* while it is desirable to design *digital* beamforming architectures (which enable the transceiver to generate beams in multiple directions at the same time), they may suffer from increased energy consumption with respect to an *analog* strategy (which, in turn, has little flexibility since the transceiver can only beamform in one direction at a time).

- *Channel State Information (CSI) acquisition*: dynamic environments impose a finite coherence interval during which CSI must be acquired and utilized. As a consequence, there is a finite number of orthogonal pilot sequences that can be assigned to the network terminals. Reuse of such pilots may result in *pilot contamination* and coherent interference, which cause performance degradation.

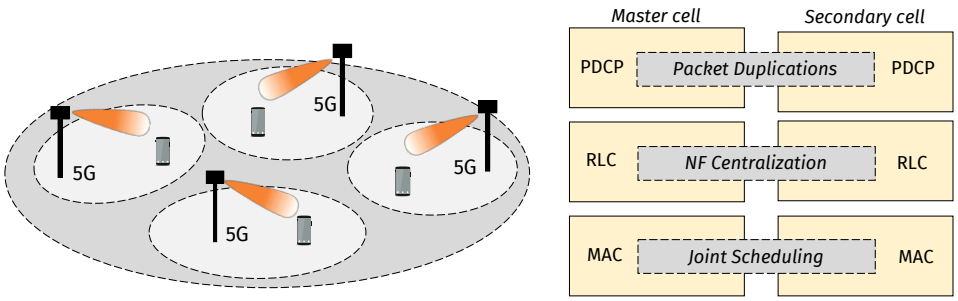
2.3 Multi-Connectivity: Towards Non-Standalone Deployments

Despite the use of MIMO technologies, the increased carrier frequency of mmWave systems may render the propagation conditions more demanding than at the lower frequencies traditionally used for wireless services, especially in terms of communication resilience. In this context, one likely key feature of cellular networks that can improve robustness is *Multi-Connectivity (MC)* [73, 74], which enables each user to maintain multiple possible signal paths to different cells so that drops in one link can be overcome by switching data paths [75]. An MC architecture can be both among multiple 5G cells (SA deployment) and between 5G cells and traditional 4G/LTE cells in the legacy spectrum (NSA deployment).

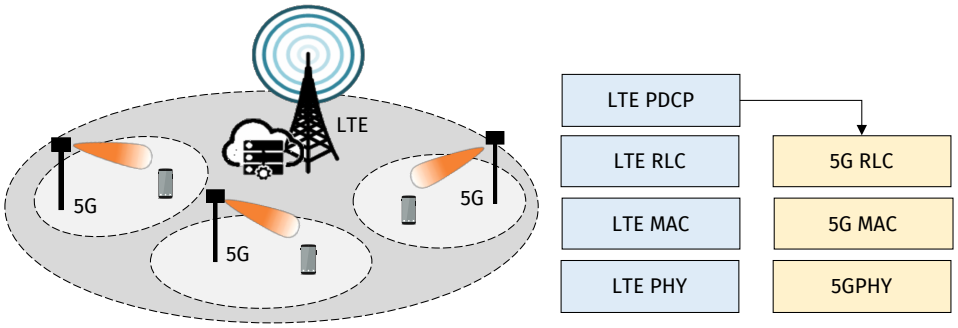
Multi-Connectivity in SA Deployments. In this configuration (see Fig. 2.3a), multiple 5G cells, possibly operating in the mmWave spectrum, can be integrated through MC to perform network function integration, bearer split, and user plane aggregation at lower layers of the protocol stack, including [4]:

- *Bearer split at the Packet Data Convergence Protocol (PDCP) level*, with a duplication of data packets towards the secondary cell. The duplicated data can then be used to enhance diversity and hence reliability whenever required by a specific service type;
- *Bearer split at the Radio Link Control (RLC) level*, such that the secondary cell only implements the lower parts of the RLC functionality, while the higher parts of the RLC are centralized at the master cell, allowing for a faster reaction of traffic steering mechanisms towards the radio legs;
- *Aggregation at the MAC level*, like in LTE Release 10 Carrier Aggregation (CA). This provides the possibility of joint scheduling across the radio legs, and hence the fastest possible reaction to changing radio conditions. This level of aggregation, however, typically requires ideal backhaul between the base stations, or these should be co-located, and may be complicated in the context of different numerologies involved.

Multi-Connectivity in NSA Deployments. In this configuration (see Fig. 2.3b), a 5G base station, possibly operating in the mmWave spectrum, uses a 4G/LTE cell as support for the control plane management, and mobile terminals exploit MC to maintain multiple possible connections to provide *macro-diversity*. Mobiles with such 4G/5G integration feature benefit from both the high bit-rates that can be provided by the mmWave links and the more robust, but lower-rate, legacy channels, thereby opening up new ways to solve capacity issues as well as provide good mobile network performance and robustness. Some literature works, such as [76], consider only the bands under 6 GHz for the control channel of 5G networks, to provide robustness against blockage and a wider coverage range, but this solution does not provide the high capacities that can be obtained when exploiting mmWave frequencies. The potential of combining legacy and mmWave technologies in outdoor scenarios has also been investigated in [77], highlighting the significant benefits that a mmWave network achieves with flexible, dynamic support from LTE technologies. Article [78] also proposes an MC framework as a solution for mobility-related link failures and throughput degradation of cell-edge users, enabling increased reliability with different levels of mobility. The 3GPP is also envisioning to adopt key principles of the MC concept from LTE Release 12 [79] for NSA deployments, where the user plane of LTE and 5G are aggregated at PDCP level.



(a) SA deployment.



(b) NSA deployment.

Fig. 2.3: Possible MC configurations. An MC architecture can be both among multiple 5G cells and between 5G cells and traditional 4G/LTE cells in the legacy spectrum.

The 3GPP, in its Release 15, supports both SA and NSA deployments. The former includes the NR core and RAN, the latter relies on the LTE RAN as a radio overlay [80].

Finally, notice that an MC scheme should not be confused with a mmWave version of the Joint Transmission (JT) Coordinated MultiPoint (CoMP) nor of the Coordinated Scheduling/Beamforming (CS/CB) CoMP [81,82]. In the first case, multiple base stations are simultaneously and cooperatively selected as transmission cells to achieve better reception of users at the cell edge. In the second case, users receive data only from their current serving cells, and base stations share their associated users' channel state information and their relative scheduling information, with the overall goal of enabling inter-cell interference mitigation in a distributed way. Importantly, unlike in standard CoMP, the mobile terminals do not need to maintain relative phase information for the links from different cells – a task that would be extremely difficult in a mmWave setting due to the high Doppler.

2.4 Conclusions

In this chapter we described the main technological novelties that are being discussed in the research community to match the 5G performance requirements. We focused on the potentials and limitations of the mmWave bands as a solution to achieve sufficient link budget, massive MIMO to establish directional communications and provide antenna gain by beamforming, and the multi-connectivity paradigm to guarantee service continuity in case of network disconnections.

Part I

Millimeter Waves for Future 5G Cellular Networks

Chapter 3

Initial Access in 5G Millimeter Wave Cellular Networks

3.1 Introduction

As introduced in Sec. 2.1, the use of the mmWave spectrum for communication will be one of the innovations of 5G cellular mobile networks. It will provide unprecedented data rates, but is highly susceptible to rapid channel variations and suffers from severe isotropic pathloss. Highly directional antennas at the transmitter and the receiver will be used to compensate for these shortcomings and achieve sufficient link budget in wide area networks. However, directionality demands precise alignment of the transmitter and the receiver beams, an operation which has important implications for control plane procedures, like *Initial Access (IA)*, and may increase the delay of the data transmission.

3.1.1 Initial Access in 4G/LTE Networks

Initial access [83] is the procedure by which a mobile terminal establishes an initial physical link connection with a cell, a necessary step to access the network, and is composed of three steps, as described in the following paragraphs.

Step 1: Cell Search Cell search aims at detecting surrounding cells and measuring the strength of received signal from each of these cells. One of them will become the User Equipment (UE)'s entry point to join the network. This is done by synchronizing to each available frequency in the network and checking whether this is a frequency from the right operator which it wants to connect to.

This synchronization procedure makes use of two specially designed physical signals which are broadcast omnidirectionally in the downlink by the eNodeB (eNB) in each cell: the Primary Synchronization Signal (PSS) and the Secondary Synchronization Signal (SSS). Each UE in the cell is aware a priori of when and where the synchronization control channel is and, thereby, it can extract and detect those signals. The detection of the PSS and SSS not only enables time and frequency synchronization, but also provides the UE with the physical layer identity of the cell and the cyclic prefix length, and informs it whether the cell uses Time Division Duplexing (TDD) of Frequency

This chapter is based on the contributions presented in **[J1, J5, C4]**. Part of the results included in this chapter is also based on joint work with Michele Polese.

Division Duplexing (FDD). Since, usually, multiple neighbour eNBs send their own PSS and SSS messages, the PHY entity of each UE generates a list of detected cells, each with its corresponding cell ID and the averaged recorded RSRP.[1] Finally, inspecting the report, the terminal can choose the cell with the strongest RSRP, instructing back its PHY entity to synchronize to this particular cell.

Step 2: System Information Acquisition At this moment, the terminal has to acquire the cell system information. This information is repeatedly broadcast by the network and needs to be acquired by terminals in order for them to access and, in general, operate properly within the network and a specific cell. The system information includes, among other things, downlink and uplink cell bandwidths, uplink and downlink configuration in case of TDD, detailed parameters related to Random Access (RA) transmission, number of transmit antennas and uplink power control.

In LTE, system information is delivered by two different mechanisms relying on two different transport channels: (i) a limited amount of system information, corresponding to the so-called Master Information Block (MIB), is transmitted using the Physical Broadcast Channel (PBCH); (ii) the main part of the system information, corresponding to different so-called System Information Blocks (SIBs), is transmitted using the Downlink Shared Cannel (DL-SCH).

Step 3: Random Access At this stage, the UE does not have any resource or channel available to inform the network about its desire to connect to it; the LTE Random Access Channel (RACH) therefore plays a key role to perform this connection. The RA procedure comes in two forms, allowing access to be either contention-based (implying an inherent risk of collision) or contention-free.

- In the contention-based procedure, a RA preamble signature is randomly chosen by the UE, with the result that it is possible for more than one UE to transmit simultaneously the same signature, leading to the need for a subsequent *contention resolution* process.
- In the contention-free procedure, a dedicated signature is allocated to a certain UE. In such a way, the whole procedure is faster than the contention-based one, provided that enough dedicated signatures can be assigned to each UE accessing the network.

3.1.2 Initial Access in 5G Networks: Limitations

In current LTE systems, IA is performed on omnidirectional channels, whereas beamforming can only be performed after a physical link is established [84]. However, there are several factors that make 4G/LTE procedures unsuitable for use in a 5G mmWave context [52].

Discovery range mismatch In LTE systems, acquiring time-frequency synchronization during cell search is facilitated, as signals are transmitted omnidirectionally in the downlink. In mmWave bands, it may instead be essential to exploit the beamforming gains even during the cell search phase, since omnidirectional signaling may generate a mismatch between the relatively short range at which a cell can be detected (C-plane range), and the much longer range at which a user could directionally send and receive data (U-plane range) [85,86].

Multi-connectivity To ensure sufficient coverage, mmWave networks will be much denser. Each user is expected to simultaneously detect multiple potential serving stations, including at least a macro eNB operating in the LTE spectrum. Consequently, the IA procedures have to be redesigned in order to capture this fundamental new feature.

[1] RSRP is the linear average of the downlink reference signals across the channel bandwidth and provides information about signal strength, without giving any indication of signal quality.

Deafness and blockage In mmWave cellular networks, IA messages may not be received due to deafness or blockage phenomena. Deafness refers to a situation where the transmit-receive beams do not point to each other, whereas blockage causes a failed message delivery due to a channel drop, which may be related to obstacles, hand rotations, and other mm-Wave-sensitive events, as described in Sec. 2.1.1. Increasing the transmission power, or waiting for a random back-off time (as done in LTE), are not suitable approaches in mmWave networks, hence new techniques have to be introduced to discriminate among the different reasons for access failure.

Dynamics-aware access Due to denser topologies, association schemes based on RSRP would be highly inefficient in mmWave cellular networks, an issue already encountered in heterogeneous networks [87]. However, the challenge with higher frequencies is the need to also account for dynamics such as directionality and intermittency.

3.1.3 Motivations and Chapter Structure

Given the limitations outlined in Sec. 3.1.2, there is an urge to extend current LTE procedure for IA with innovative mmWave-aware algorithms and methods. A natural (and practical) solution is to use beamforming even in the first stages of the IA procedure, keeping in mind that a fully directional data plane requires a directional IA procedure in the new frequency band. However, when considering an analog multiantenna architecture, directionality means that only one direction can be considered at a time, thereby losing the broadcast property of the wireless medium, with important implications for protocol design and delay performance that must be carefully taken into consideration.

Along these lines, in this chapter we provide a global comprehensive evaluation of mmWave measurement frameworks that have been proposed for IA, and we focus in particular on the strategies that have been recently standardized by the 3GPP in its Release 15, as part of NR specification. We assess how to optimally design fast, accurate and robust control-plane IA schemes through measurement reports. We focus on a DL framework, and on SA and NSA architectures that may or may not leverage the MC paradigm. Performance is assessed in terms of (i) *detection accuracy*, i.e., how representative the measurement is; (ii) *reactiveness*, i.e., how quickly a mobile user gets access to the network; and (iii) *overhead*, i.e., how many resources are needed for the measurement operations. Finally, we illustrate some of the complex trade-offs to be considered when designing IA solutions for 3GPP NR. The results prove that the optimal design for implementing efficient and fast IA must account for specific features such as the gNBs density, the antenna geometry, the beamforming configuration and the level of harmonization of different technologies.

The rest of the chapter is organized as follows. Sec. 3.2 reports related work on IA procedures at mmWave frequencies. In Sec.3.3 we review the 3GPP NR procedures for IA, and present the IA frameworks we will evaluate. In Sec. 3.4 we describe the system model, the metrics that will be considered and the 3GPP parameters that will be configured, and present major findings and results. Finally, in Sec. 3.5 we identify guidelines for the design of IA strategies at mmWaves.

3.2 Related Work

Papers on IA in 5G mmWave cellular systems are very recent.[2] Most literature refers to challenges that have been analyzed in the past at lower frequencies in ad hoc wireless network scenarios or, more recently, referred to the 60 GHz IEEE 802.11ad WLAN and WPAN scenarios (e.g., [103–105]).

[2]We refer to works [52, 88, 102] for a detailed taxonomy of recent IA strategies.

Table 3.1: Literature on initial access strategies for mmWave networks.

Topic	Relevant References
Initial Access [52,88,89]	[85,90,91] Exhaustive search. [92–94] More advanced searching schemes. [95–98] Context-aware initial access. [99–101] Performance comparison.

However, most of the proposed solutions are unsuitable for next-generation cellular network requirements and present many limitations (e.g., they are appropriate for short-range, static and indoor scenarios, which do not match well the requirements of 5G systems). Therefore, new specifically designed solutions for cellular networks need to be found.

Jeong *et al.* [90] and Barati *et al.* [85] proposed an exhaustive method that performs directional communication over mmWave frequencies by periodically transmitting synchronization signals to scan the angular space. The result of this approach is that the growth of the number of antenna elements at either the transmitter or the receiver provides a large performance gain compared to the case of an omnidirectional antenna. However, this solution leads to a long duration of the IA with respect to LTE, and poorly reactive tracking. Similarly, in [91], measurement reporting design options were compared, considering different scanning and signaling procedures, to evaluate access delay and system overhead. The channel structure and multiple access issues were also considered. The analysis demonstrated significant benefits of low-resolution fully digital architectures in comparison to single stream analog beamforming. Additionally, more sophisticated discovery techniques (e.g., [92,93]) were studied to alleviate the exhaustive search delay through the implementation of a multi-phase hierarchical procedure based on the access signals being initially sent in few directions over wide beams, which are iteratively refined until the communication is sufficiently directional. In [94] a low-complexity beam selection method by low-cost analog beamforming was derived by exploiting a certain sparsity of mmWave channels. It was shown that beam selection can be carried out without explicit channel estimation, using the notion of compressive sensing.

Context information can also be exploited to improve the cell discovery procedure and minimize the delay [95,96], while capturing the effects of position inaccuracy in the presence of obstacles. In the scheme proposed in [97], booster cells (operating at mmWaves) were deployed under the coverage of an anchor cell (operating at LTE frequencies). The anchor base station gets control over IA informing the booster cell about user locations, in order to enable mmWave gNB to directly steer towards the user position. Finally, in [98], the authors studied how the performance of analog beamforming degrades in the presence of angular errors in the available context information during the initial access or tracking procedures, according to the status of the UE (connected or non-connected, respectively).

The performance of the association techniques also depends on the beamforming architecture implemented in the transceivers. Preliminary works aiming at finding the optimal beamforming strategy refer to WLAN scenarios. For example, the algorithm proposed in [106] takes into account the spatial distribution of nodes to allocate the beamwidth of each antenna pattern in an adaptive fashion and satisfy the required link budget criterion. Since the proposed algorithm minimizes the collisions, it also minimizes the average time required to transmit a data packet from the source to the destination through a specific direction. In 5G scenarios, papers [85,90,92] gave some insights on trade-offs among different beamforming architectures in terms of users' communication quality. In this context, articles [99,100] evaluated the mmWave cellular network performance while accounting for the beam training, association overhead and beamforming architecture. More recently the authors

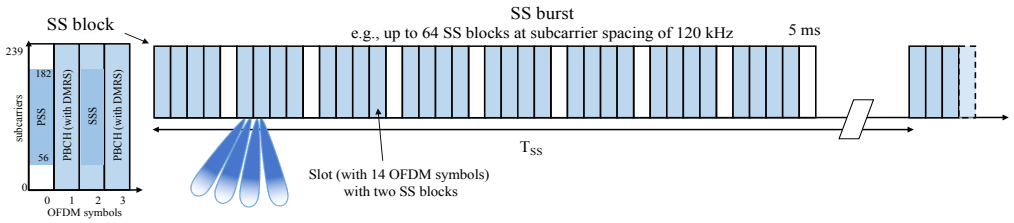


Fig. 3.1: SS and burst structure for IA in NR systems. SS blocks are sent every T_{SS} and are grouped in bursts of up to L blocks.

in [53, 89, 101] compared the performance of several IA schemes, based on current 5G NR slot design considerations, in terms of coverage and search delays, and for different antenna array settings. The results showed that, although employing wide beams, initial beam training with full pilot reuse is nearly as good as perfect beam alignment. Finally, paper [107] provided an overview of the main features of NR with respect to initial access and multi-beam operations, and article [108] discussed about how to collect channel state information in NR.

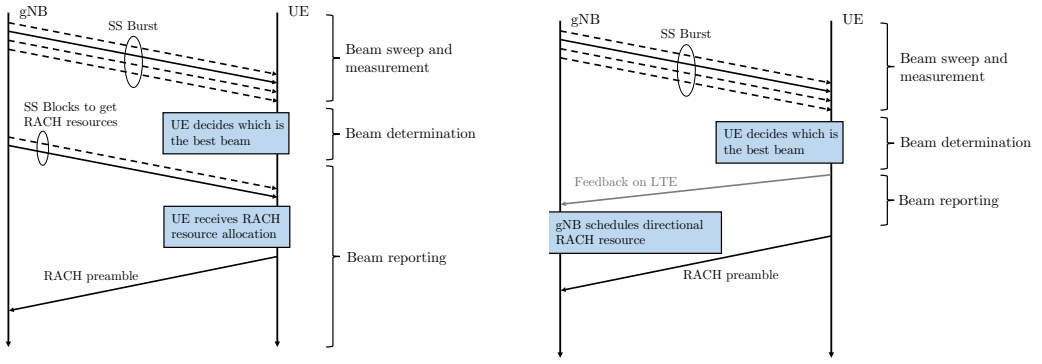
3.3 Initial Access in 5G Networks: the 3GPP NR Approach

Given that NR will support communication at mmWave frequencies, it is necessary to account for beamforming and directionality in the design of its PHY and MAC layers. The NR specifications will thus include a set of parameters for the frame structure dedicated to high carrier frequencies, as well as synchronization signals that enable IA procedures [64]. In this regard, in Sec. 3.3.1 we introduce the 3GPP IA measurement signals proposed for NR, while in Sec. 3.3.2 we present the IA frameworks we consider.

3.3.1 3GPP NR Measurement Signals for Initial Access

In LTE, as reviewed in 3.1.1, the synchronization procedure for IA relies on two specifically designed signals, broadcast omnidirectionally in the downlink, namely the PSS and the SSS. Each UE in the cell knows a priori when and where the synchronization control channel is and can extract and detect those signals. For NR, the 3GPP has defined a directional version of synchronization signals introducing Synchronization Signal (SS) blocks and bursts [11], as represented in Fig. 3.1.

An SS block is composed by 240 subcarriers in frequency and 4 OFDM symbols [11] in time, and carries the PSS, the SSS and the PBCH, used to estimate the RSRP and select the optimal beam to communicate. The SS blocks are transmitted in the first 5 ms of an SS burst, which is then repeated with periodicity $T_{SS} \in \{5, 10, 20, 40, 80, 160\}$ ms [109]. The maximum number of SS blocks per burst L varies according to the carrier frequency, with 64 blocks per burst above 6 GHz. At mmWaves, each gNB directionally transmits the SS blocks by sequentially sweeping different angular directions to cover a whole cell sector. Based on the measured quality of the received signal, the SS blocks can be exploited to identify their initial directions of transmission (in this case, to reduce the impact of SS transmissions and guarantee prompt network access operations, SS can be sent through wider beams) and for beam tracking purposes, as we will see in Chapter 4.



(a) SA deployment.

(b) NSA deployment.

Fig. 3.2: Signals and messages exchanged during the beam management procedure according to the 3GPP NR specifications, for SA and NSA deployments. Notice that the duration of the three phases is not in scale, since it depends on the actual configuration of the network parameters.

3.3.2 3GPP NR Initial Access Frameworks

3GPP NR specifications include a set of basic procedures for IA [9]. We consider two different deployment architectures. With the *standalone* option, the UE connects only to an NR gNB at mmWave frequencies. With *multi-connectivity standalone* option, instead, each UE maintains multiple possible signal paths to different cells at different frequencies (e.g., NR at mmWaves and LTE at conventional frequencies), thus providing both high capacity and robust connections [110]. A downlink framework is analyzed, where the gNBs transmit synchronization signals (i.e., SS blocks) which are collected by the surrounding UEs.

For an SA deployment, the IA scheme proposed in 3GPP NR is composed of the following steps, as illustrated in Fig. 3.2a:

- **Step 1 (Beam sweeping)**: UEs and gNBs cover a spatial area with a set of beams transmitted and received according to pre-specified intervals and directions. The measurements are carried out with an exhaustive search, i.e., users and base stations have a predefined codebook of directions (each identified by a beamforming vector) that cover the whole angular space and are used sequentially to transmit/receive synchronization signals [90].
- **Step 2 (Beam measurement)**: UEs evaluate the quality of the received signal, embedded in the SS blocks. Different metrics could be used [111]. In this thesis, we consider the SNR, which is the linear average of the received power on different resources with synchronization signals divided by the noise power.
- **Step 3 (Beam determination)**: UEs select the beam through which they experienced the maximum SNR, if above a predefined threshold. The corresponding sector will be chosen for the subsequent transmissions and receptions and benefit from the resulting antenna gain. We recall that the optimal beam pair for each link can be determined only after a complete scan, since the gNBs have to detect all UEs within their whole angular range. Steps 1, 2 and 3 correspond to cell search in 4G/LTE IA (see Sec. 3.1.1).
- **Step 4 (Beam reporting)**: UEs send beam quality and beam decision information to the RAN. This corresponds to random access in 4G/LTE IA (see Sec. 3.1.1). According to the 3GPP, after the best beam is determined, the mobile terminal has to wait for the gNB to schedule the RACH opportunity towards the best direction that the UE just determined, for performing

random access and implicitly informing the selected serving infrastructure of the optimal direction (or set of directions) through which it has to steer its beam, in order to be properly aligned. It has been agreed that for each SS block the gNB will specify one or more RACH opportunities with a certain time and frequency offset and direction, so that the UE knows when to transmit the RACH preamble [9, 112]. This may require an additional complete directional scan of the gNB, thus further increasing the time it takes to access the network, especially when there is a large number of directions to scan.

For an NSA deployment it may be possible to decrease the beam reporting information latency in Step 4 compared to SA when large antenna arrays are used, as illustrated in Fig. 3.2b. In the NSA case, the UE leverages the support of the LTE overlay to receive information on the RACH opportunity of interest, and thus a full beamformed RACH may be scheduled without the need to monitor multiple directional SS blocks and associated RACH opportunities.

3.4 Performance Analysis

In Sec. 3.4.1 we define the metrics that will be used to compare the performance of the different IA frameworks, in Sec. 3.4.2 we list the relevant parameters that affect the performance of the investigated solutions, and finally in Sec. 3.4.3 we present our main simulation results.

3.4.1 Performance Metrics

The performance of the different IA architectures and configurations will be assessed using three different metrics. The *detection accuracy* is measured in terms of probability of misdetection P_{MD}, defined as the probability that the UE is not detected by the base station (i.e., the perceived SNR is below a threshold Γ_{th}). The *reactiveness* is represented by the average time to find the best beam pair, i.e., the average time for a UE to connect to a gNB. The *overhead* is the amount of time and frequency resources allocated to the framework with respect to the total amount of available resources.

The simulations for the detection accuracy performance evaluation are based on realistic system design configurations where multiple gNBs are deployed according to a Poission Point Process (PPP). The channel model is based on recent real-world measurements at 28 GHz in New York City, to provide a realistic assessment of mmWave micro and picocellular networks in a dense urban deployment. A complete description of the channel parameters can be found in [113].

3.4.2 Performance Parameters

Frame Structure As depicted in Fig. 3.3, we consider the frame structure of 3GPP NR, with different subcarrier spacings Δ_f. Given that in [11] the only subcarrier spacings considered for IA at frequencies above 6 GHz are $\Delta_f = 120$ and 240 kHz, i.e., 15×2^n kHz, with $n \in [3, 4]$, we will only consider these cases. The slot duration in ms and the symbol duration in μs are, respectively

$$T_{\text{slot}} = \frac{1}{2^n} \qquad \text{and} \qquad T_{\text{symb}} = \frac{71.35}{2^n}. \tag{3.1}$$

Therefore, for $n = 3$ and 4 the slot duration is 125 μs or 62.5 μs, respectively [114]. Moreover, according to the 3GPP specifications [11], the maximum number of subcarriers allocated to the SS blocks is 240, thus the bandwidth reserved for the SS blocks would be respectively 28.8 and 57.6 MHz. We consider a maximum channel bandwidth $B = 400$ MHz per carrier [64].

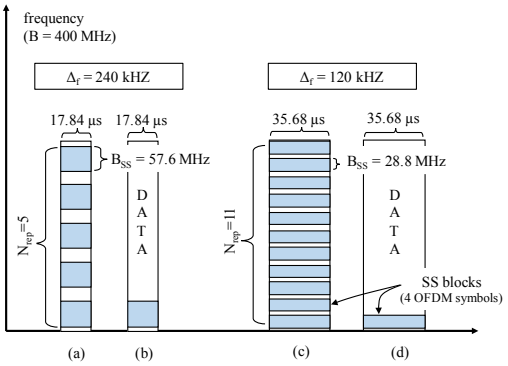


Fig. 3.3: SS block structure. For configurations (a) and (b), each blue rectangle is an SS block (with 4 OFDM symbols) of duration 17.84 μs (i.e., $\Delta_f = 240$ kHz) and bandwidth $B_{\text{SS}} = 57.6$ MHz. For configurations (c) and (d) (for which $\Delta_f = 120$ kHz), instead, the blocks last 35.68 μs and have bandwidth $B_{\text{SS}} = 28.8$ MHz. Cases (a) and (c) implement a *frequency repetition* scheme (with $N_{rep} = 5$ and 11, respectively) while, for cases (b) and (d), a *data* solution (i.e., $N_{rep} = 1$) is preferred.

Frequency Diversity It is possible to configure the system to exploit frequency diversity, D. Given that 240 subcarriers are allocated in frequency to an SS, the remaining bandwidth in the symbols which contain an SS block is $B - 240\Delta_f$. Therefore, it is possible to adopt two different strategies: (i) *data* (as represented in Figs. 3.3(b) and (d)), i.e., the remaining bandwidth $B - 240\Delta_f$ is used for data transmission towards users , or (ii) *repetition* (as displayed in Figs. 3.3(a) and (c)), i.e., the information in the first 240 subcarriers is repeated in the remaining subcarriers to increase the robustness against noise and enhance the detection capabilities. The number of repetitions is therefore $N_{rep} = 1$ if frequency diversity is not used (i.e., $D = 0$, and a single chunk of the available bandwidth is used for the SS block), and $N_{rep} = 11$ or $N_{rep} = 5$ when repetition is used (i.e., $D = 1$) with $\Delta_f = 120$ kHz or $\Delta_f = 240$ kHz, respectively. There is a guard interval in frequency among the different repetitions of the SS blocks, to provide a good trade-off between frequency diversity and coherent combining [85]. Notice that 3GPP does not provide specifications for the repetition scheme.

SS Block Configuration We consider different configurations of the SS blocks and bursts. The maximum number N_{SS} of SS blocks in a burst for our frame structure and carrier frequencies is $L = 64$. We assume that, if $N_{\text{SS}} < L$, the SS blocks will be transmitted in the first N_{SS} opportunities. The actual maximum duration of an SS burst is $D_{\text{max,SS}} = 2.5$ ms for $\Delta_f = 240$ kHz and $D_{\text{max,SS}} = 5$ ms for $\Delta_f = 120$ kHz. We will also investigate all the possible values for the SS burst periodicity T_{ss}, as defined in [109, 115], i.e., $T_{\text{SS}} \in \{5, 10, 20, 40, 80, 160\}$ ms.

Array Geometry As shown in Table 3.2, another fundamental parameter is the array geometry, i.e., the number of antenna elements M at the gNB and UE and the number of directions that need to be covered, both in azimuth N_θ and in elevation N_ϕ. In general, the antenna elements can be deployed as uniform linear or planar arrays, i.e., ULA and UPA respectively, and can be arranged as either rectangular or square arrays. Among the possible antenna designs, the most suitable approach is the use of UPAs, since they can enable 3D beamforming by adapting the beam in both azimuth and elevation planes [49]. In the simulations, the spacing of the elements is set to $\lambda/2$, where λ is the wavelength, since this pattern was shown to offer excellent system capacity in small-cell urban deployments, as well as easy packageability (e.g., at 28 GHz, a 4×4 array has a size of roughly 1.5 cm $\times$ 1.5 cm) [85]. At the gNB we consider a single sector in a three sector site, i.e., the azimuth θ varies from -60 to 60 degrees, for a total of $\Delta_\theta = 120$ degrees. The elevation ϕ varies between

Table 3.2: Relationship between M, θ and N_θ, for the azimuth case. Each gNB sector sweeps through $\Delta_{\theta,\text{gNB}} = 120°$, while the UE scans over $\Delta_{\theta,\text{UE}} = 360°$. In our evaluation, we consider a single antenna array at the UE modeled as a uniform rectangular array with isotropic antenna elements, following the approach of the literature [116]. Real handheld devices will be equipped with multiple patch antennas able to cover the whole angular space.

M	θ [deg]	$N_{\theta,\text{gNB}}$	$N_{\theta,\text{UE}}$
4	60	2	6
16	26	5	14
64	13	10	28

-30 and 30 degrees, for a total of $\Delta_\phi = 60$ degrees, and also includes a fixed mechanical tilt of the array pointing towards the ground. There exists a strong correlation among beamwidth, number of antenna elements and BF gain. The more antenna elements in the system, the narrower the beams, the higher the gain that can be achieved by beamforming, and the more precise and directional the transmission. Thus, given the array geometry, we compute the beamwidth Δ_{beam} at 3 dB of the main lobe of the beamforming vector, and then $N_\theta = \Delta_\theta / \Delta_{\text{beam}}$ and $N_\phi = \Delta_\phi / \Delta_{\text{beam}}$.

Network Deployment Finally, the last parameters are the number of users $\mathcal{N} \in \{5, 10, 20\}$ per sector of the gNBs and the density of base stations λ_b, expressed in gNB/km^2.

3.4.3 Simulation Results

3.4.3.1 Detection Accuracy Results

Array size and gNB density – Fig. 3.4 shows the Cumulative Distribution Function (CDF) of the SNR between the mobile terminal and the gNB it is associated to, for different antenna configurations and considering two density values. Notice that the curves are not smooth because of the progressive transitions of the SNR among the different path loss regimes, i.e., Line of Sight (LOS), Non Line of Sight (NLOS) and outage. We see that better detection accuracy performance can be achieved when densifying the network and when using larger arrays. In the first case, the endpoints are progressively closer, thus ensuring better signal quality and, in general, stronger received power. In the

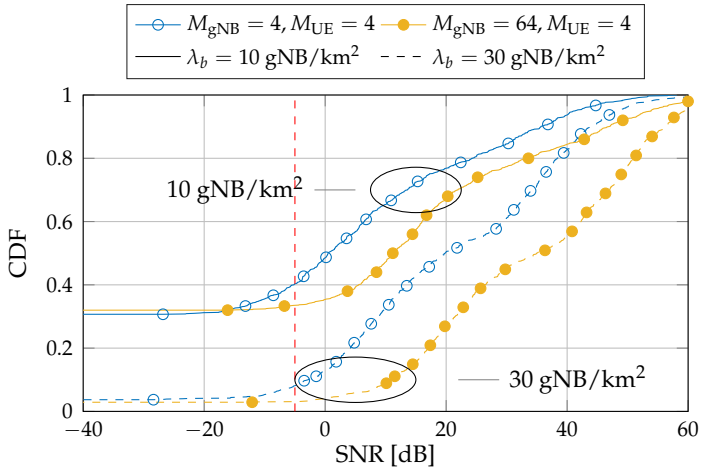


Fig. 3.4: CDF of the SNR, for different antenna configurations. $\Delta_f = 120$ kHz, $N_{rep} = 0$. The red dashed line represents the SNR threshold $\Gamma_{\text{th}} = -5$ dB that has been considered in this study.

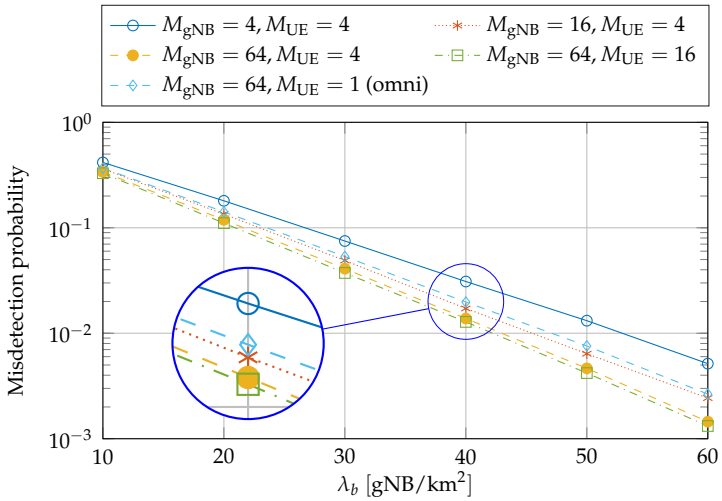


Fig. 3.5: P_{MD} vs. λ_b, for different antenna configurations.

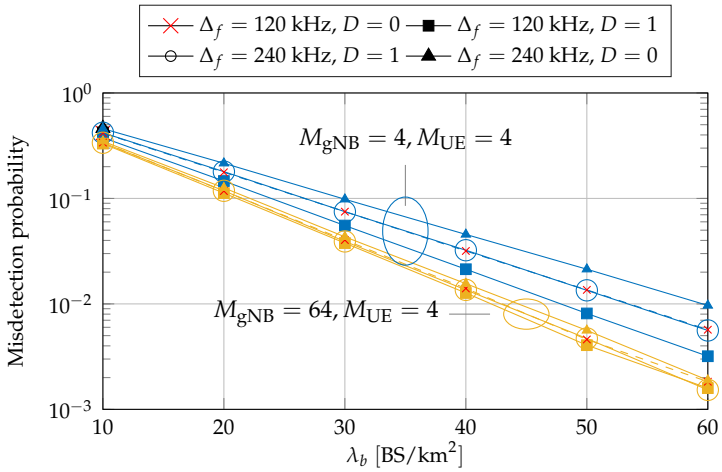


Fig. 3.6: P_{MD} as a function of λ_b, for different subcarrier spacings Δ_f and repetition strategies and for different antenna configurations. $M_{gNB} = 4, M_{UE} = 4, \Gamma_{th} = -5$ dB.

second case, narrower beams can be steered thus guaranteeing higher gains produced by beamforming. We also notice that, for good SNR regimes, the $M_{gNB} = 4, M_{UE} = 4$ and $M_{gNB} = 64, M_{UE} = 4$ configurations present good enough SNR values: in these regions, the channel conditions are sufficiently good to ensure satisfactory signal quality (and, consequently, acceptable misdetection) even when considering small antenna factors. Finally, the red line represents the SNR threshold $\Gamma_{th} = -5$ dB that we will consider in this thesis.

Similar considerations can be deduced from Fig. 3.5, which illustrates how the misdetection probability monotonically decreases when the gNB density λ_b progressively increases or when the transceiver is equipped with a larger number of antenna elements, since more focused beams can be generated in this case. Moreover, we notice that the beamforming strategy in which the UE transmits or receives omnidirectionally, although guaranteeing fast access operations, does not ensure accurate IA performance and leads to degraded detection capabilities. More specifically, the gap with a fully directional architecture (e.g., $M_{gNB} = 64, M_{UE} = 16$) is quite remarkable for very dense scenarios, and increases as the gNB density increases. For example, the configuration with 16 antennas (i.e., $M_{UE} = 16$) and that with a single omnidirectional antenna at the UE reach the same P_{MD}, but at dif-

ferent values of gNB density λ_b, respectively 30 and 35 gNB/km^2: the omnidirectional configuration requires a higher density (i.e., 5 gNB/km^2 more) to compensate for the smaller beamforming gain.

Subcarrier spacing and frequency diversity – Fig. 3.6 reports the misdetection probability related to λ_b, for different subcarrier spacings Δ_f and repetition strategies D. First, we see that, if no repetitions are used (i.e., $D = 0$), lower detection accuracy performance is associated with the $\Delta_f = 240$ kHz configuration, due to the resulting larger impact of the thermal noise and the consequent SNR degradation. Furthermore, the detection efficiency can be enhanced by repeating the SS block information embedded in the first 240 subcarriers in the remaining subcarriers (i.e., $D = 1$), to increase the robustness of the communication and mitigate the effect of the noise in the detection process. In fact, if a frequency diversity approach is preferred, the UE has $N_{rep} > 1$ attempts to properly collect the synchronization signals exchanged during the beam sweeping phase, compared to the single opportunity the nodes would have had if they had not implemented any repetition strategy. We also observe that the $\Delta_f = 120$ kHz with no frequency diversity configuration and the $\Delta_f = 240$ kHz scheme with $N_{rep} = 5$ produce the same detection accuracy results, thus showing how the effect of increasing the subcarrier spacing and the number of repetitions of the SS block information in multiple frequency subbands is similar in terms of misdetection capabilities. Finally, we observe that the impact of the frequency diversity D and the subcarrier spacing Δ_f is less significant when increasing the array factor, as can be seen from the reduced gap between the curves plotted in Fig. 3.6 for the $M_{\text{gNB}} = 4, M_{\text{UE}} = 4$ and $M_{\text{gNB}} = 64, M_{\text{UE}} = 4$ configurations. The reason is that, when considering larger arrays, even the configuration with $\Delta_f = 240$ kHz and no repetitions has an average SNR which is high enough to reach small misdetection probability values.

3.4.3.2 Reactiveness Results

Analysis – For initial access, reactiveness is defined as the delay required to perform a full iterative search in all the possible combinations of the directions. The gNB and the UE need to scan respectively $N_{\theta,\text{gNB}}N_{\phi,\text{gNB}}$ and $N_{\theta,\text{UE}}N_{\phi,\text{UE}}$ directions to cover the whole horizontal and vertical space. Moreover, they can transmit or receive respectively $K_{\text{BF,gNB}}$ and $K_{\text{BF,UE}}$ beams simultaneously. Notice that, as mentioned in Sec. 3.4.2, for digital and omnidirectional architectures $K_{\text{BF}} = \min\{N_\theta N_\phi, M\}$, for hybrid $K_{\text{BF}} = \min\{N_\theta N_\phi, M\}/\nu$, where ν is a factor that limits the number of directions in which it is possible to transmit or receive at the same time, and for analog $K_{\text{BF}} = 1$ [69].

Then the total number of SS blocks needed is[3]

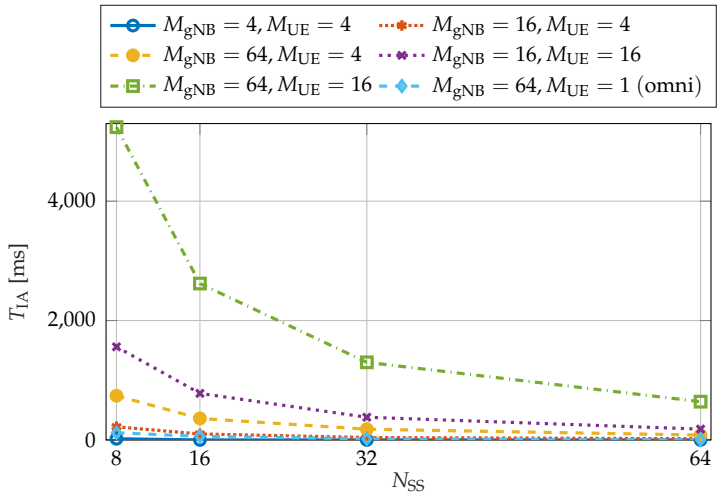
$$S_D = \left\lceil \frac{N_{\theta,\text{gNB}}N_{\phi,\text{gNB}}}{K_{\text{BF,gNB}}} \right\rceil \left\lceil \frac{N_{\theta,\text{UE}}N_{\phi,\text{UE}}}{K_{\text{BF,UE}}} \right\rceil. \tag{3.2}$$

Given that there are N_{SS} blocks in a burst, the total delay from the beginning of an SS burst transmission in a gNB to the completion of the sweep in all the possible directions is

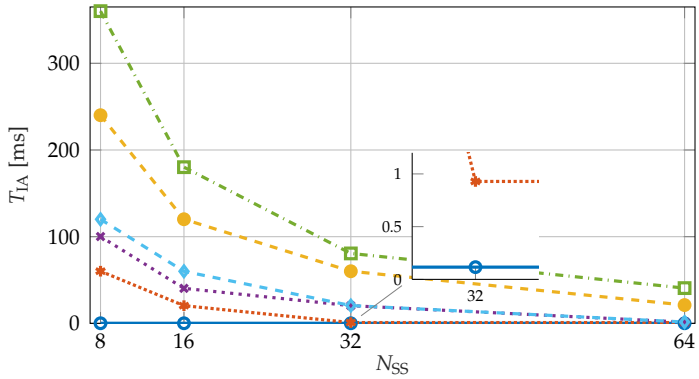
$$T_{\text{IA}} = T_{\text{SS}} \left(\left\lceil \frac{S_D}{N_{SS}} \right\rceil - 1 \right) + T_{last}, \tag{3.3}$$

where T_{last} is the time required to transmit the remaining SS blocks in the last burst (notice that there may be just one burst, thus the first term in Eq. (3.3) would be 0). This term depends on the subcarrier

[3]We recall that hybrid or digital architectures consume more power than analog ones, if the same number of bits in the Analog to Digital Converters (ADCs) is used, and thus are more likely to be implemented only at the receiver side. Nevertheless, some ADC configurations enable energy efficient digital beamforming (e.g., 3 bits ADC [117]), with a power consumption comparable to that of an analog implementation.



(a) gNB Analog, UE Analog



(b) gNB Analog, UE Digital (DL-based configuration)

Fig. 3.7: T_{IA} as a function of N_{SS} with $T_{SS} = 20$ ms.

spacing and on the number of remaining SS blocks which is given by

$$N_{SS,left} = S_D - N_{SS}\left(\left\lceil \frac{S_D}{N_{SS}} \right\rceil - 1\right). \tag{3.4}$$

Then, T_{last} is

$$T_{last} = \begin{cases} \frac{N_{SS,left}}{2} T_{slot} - 2T_{symb} & \text{if } N_{SS,left} \bmod 2 = 0 \\ \left\lfloor \frac{N_{SS,left}}{2} \right\rfloor T_{slot} + 6T_{symb} & \text{otherwise,} \end{cases} \tag{3.5}$$

The two different options account for an even or odd remaining number of SS blocks. In the first case, the SS blocks are sent in $N_{SS,left}/2$ slots, with total duration $N_{SS,left}/2T_{slot}$, but the last one is actually received in the $12th$ symbol of the last slot, i.e., 2 symbols before the end of that slot, given the positions of the SS blocks in each slot described in [11, 118]. If instead $N_{SS,left}$ is odd, six symbols of slot $\lfloor N_{SS,left}/2 \rfloor + 1$ are also used.

A selection of results is presented in the next paragraphs.

Number of SS blocks per burst and beamforming technology – In Fig. 3.7 we consider first the impact of the number of SS blocks in a burst, with a fixed SS burst periodicity $T_{SS} = 20$ ms and for different beamforming strategies and antenna configurations. In particular in Fig. 3.7a, in which both the UE and the gNB use analog beamforming, the initial access delay heavily depends on the

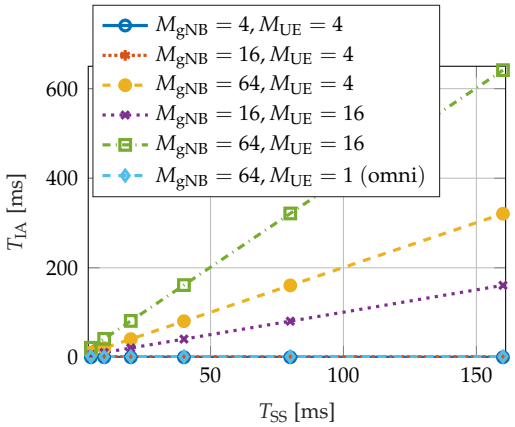


Fig. 3.8: T_{IA} as a function of T_{SS}, with analog gNB and hybrid UE. $N_{\text{SS}} = 64$

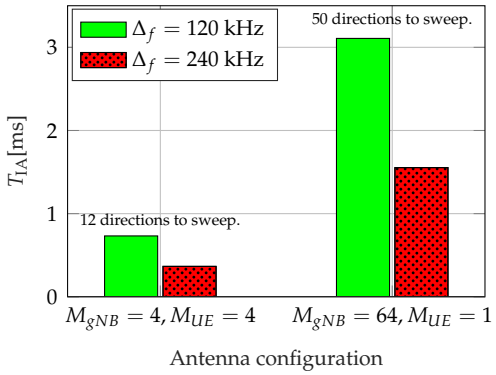


Fig. 3.9: T_{IA} for different antenna configurations and subcarrier spacing Δ_f, with gNB Analog, UE Analog.

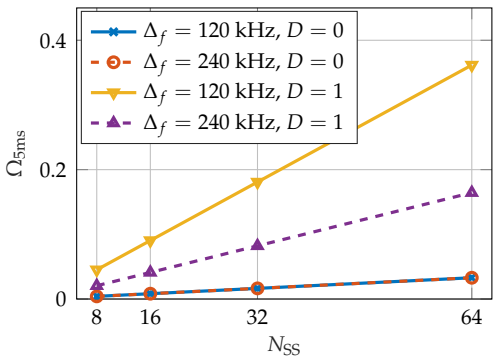
number of antennas at the transceivers since all the available directions must be scanned one by one. It may take from 0.6 s (with $N_{\text{SS}} = 64$) to 5.2 s (with $N_{\text{SS}} = 8$) to transmit and receive all the possible beams, which makes the scheme infeasible for practical usage. A reduction in the sweeping time can be achieved either by using an omnidirectional antenna at the UE or by decreasing the number of antennas both at the UE and at the gNB. In this case, the only configurations that manage to complete a scan in a single SS burst are those with 4 antennas at both sides and $N_{\text{SS}} \geq 16$, or that with $M_{\text{gNB}} = 64$, an omnidirectional UE and $N_{\text{SS}} = 64$. Another option is the usage of digital beamforming at the UE. Fig. 3.7b shows T_{IA} when the UE receives from all available directions at any given time. This leads to an increased number of configurations which are able to complete a sweep in an SS block, even with a large number of antennas at the gNB and the UE.

SS burst periodicity – For the setup with hybrid beamforming at the UE, that generally requires more than one SS burst periodicity, we show in Fig. 3.8 the dependency of T_{IA} and T_{SS}. It can be seen that the highest periodicities are not suited for a mmWave deployment, and that in general it is better to increase the number of SS blocks per burst in order to try to complete the sweep in a single burst.

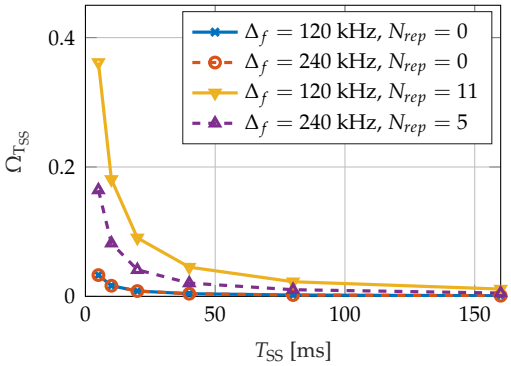
Subcarrier spacing – Another parameter that has an impact on T_{IA} is the subcarrier spacing Δ_f. As shown in Fig. 3.9, when the larger spacing is used the OFDM symbols have a shorter duration and the transmission of the SS blocks in the directions of interest can be completed earlier.

Impact of beam reporting – For initial access, in addition to the time required for directional sweeping, there is also a delay related to the allocation of the resources in which it is possible to perform initial access, which differs according to the architecture being used. As introduced in Sec. 3.3.2, 3GPP advocates, in Step 4 of the IA procedure, the implicit reporting of the chosen direction, e.g., the strongest SS block index, through contention-based random access messages, agreeing that the network should allocate multiple RACH transmissions and preambles to the UE for conveying the optimal SS block index to the gNB [9, 119]. When considering an SA configuration, beam reporting might require an additional sweep at the gNB side while, if an NSA architecture is preferred, the beam decision is forwarded through the LTE interface (and requires just a single RACH opportunity) which makes the beam reporting reactiveness equal to the latency of a legacy LTE connection. Assuming a 0% BLER data channel, the uplink latency in legacy LTE, including scheduling delay, ranges from 10.5 ms to 0.8 ms, according to the latency reduction techniques being implemented [120].

In Table 3.3, we analyze the impact of the number of SS blocks (and, consequently, of RACH opportunities) in a burst, with a fixed burst periodicity $T_{\text{SS}} = 20$ ms and for a subcarrier spacing of



(a) Ω_{5ms} as a function of N_{SS}, for different subcarrier spacings Δ_f and repetition strategies.



(b) $\Omega_{T_{SS}}$ as a function of T_{SS}, for different subcarrier spacings Δ_f and repetition strategies. N_{SS} is set to the maximum value, i.e., 64.

Fig. 3.10: Overhead for initial access, introduced by the transmission of the SS blocks. Notice that the number of repetitions for the different subcarrier spacings Δ_f is chosen to send as many repetitions of the SS blocks as possible.

$\Delta_f = 120$ KHz. The results are independent of the antenna configuration at the UE side, since the mobile terminal steers its beam through the previously determined optimal direction and does not require a beam sweeping operation to be performed. It appears clear that the SA scheme presents very good reactiveness for most of the investigated configurations and, most importantly, outperforms the NSA solution even when the LTE latency is reduced to 0.8 ms. The reason is that, if the network is able to allocate the needed RACH resources within a single SS burst, then it is possible to limit the impact of beam reporting operations on the overall initial access reactiveness, which is instead dominated by the beam sweeping phase. In particular, when considering small antenna factors and when digital beamforming is employed, beam reporting can be successfully completed through a single RACH allocation, thus guaranteeing very small delays.

3.4.3.3 Overhead Results

In this section, we characterize the overhead for IA in terms of the ratio between the time and frequency resources that are allocated to SS bursts and the maximum duration of the SS burst (i.e., 5 ms), or the entire T_{SS} interval.

Analysis – The total number of time and frequency resources R_{SS} scheduled for the transmission

Table 3.3: Reactiveness performance for beam reporting operations considering an SA or an NSA architecture. Analog or digital beamforming is implemented at the gNB side, while the UE configures its optimal beamformed direction. $T_{SS} = 20$ ms, $\Delta_f = 120$ KHz.

	$T_{BR,SA}$ [ms]			
	$N_{SS} = 8$		$N_{SS} = 64$	
M_{gNB}	Analog	Digital	Analog	Digital
4	0.0625	0.0625	0.0625	0.0625
16	0.5	0.0625	0.5	0.0625
64	40.56	0.0625	1.562	0.0625
$T_{BR,NSA} \in \{10, 4, 0.8\}$ ms, according to [120].				

Table 3.4: Overhead for beam reporting operations considering an SA architecture. Analog or digital beamforming is implemented at the gNB side, for different antenna array structures.

	$\Omega_{\text{BR,SA}} \cdot 10^{-3}$			
	$\Delta_{f,\text{RACH}} = 60$ kHz		$\Delta_{f,\text{RACH}} = 120$ kHz	
M_{gNB}	Analog	Digital	Analog	Digital
4	0.0894	0.0894	0.0894	0.0894
16	0.7149	0.0894	0.7149	0.0894
64	2.2341	0.0894	2.2341	0.0894

of N_{SS} SS blocks, each spanning 4 OFDM symbols and 240 (or multiple of 240) subcarriers, is given by

$$R_{\text{SS}} = N_{\text{SS}}\, 4T_{\text{symb}}\, 240 N_{rep} \Delta_f, \tag{3.6}$$

where T_{symb} is expressed in ms and Δ_f in kHz. The overhead for the 5 ms time interval with the SS burst transmission and total bandwidth B (in Hz) is then given by

$$\Omega_{\text{5ms}} = \frac{N_{\text{SS}}\, 4T_{\text{symb}}\, 240 N_{rep} \Delta_f}{5B}, \tag{3.7}$$

and the overhead considering the total burst periodicity T_{SS} is

$$\Omega_{\text{T}_{\text{SS}}} = \frac{N_{\text{SS}}\, 4T_{\text{symb}}\, 240 N_{rep} \Delta_f}{T_{\text{SS}} B}. \tag{3.8}$$

Subcarrier spacing and frequency diversity – Fig. 3.10 reports the overhead related to the maximum duration of the SS burst (i.e., 5 ms) for different subcarrier spacings and repetition strategies. It can be seen that if no repetitions are used (i.e., $D = 0$) then the overheads for the configurations with $\Delta_f = 120$ kHz and $\Delta_f = 240$ kHz are equivalent. In fact, when configuring large subcarrier spacings (i.e., $\Delta_f = 240$ kHz), the OFDM symbols used for the SS blocks have half the duration, but they occupy twice the bandwidth of the systems with narrower subcarrier spacings (i.e., $\Delta_f = 120$ kHz), given that the same number of subcarriers are used. Instead, when a repetition strategy is used (i.e., $D = 1$), the overhead is different. As mentioned in Sec. 3.4.2, we consider 5 repetitions for $\Delta_f = 240$ kHz and 11 for $\Delta_f = 120$ kHz. Therefore, the actual amount of bandwidth that is used is comparable, but since the OFDM symbols with $\Delta_f = 120$ kHz last twice as long as those with the larger subcarrier spacing, the overhead in terms of resources used for the SS burst is higher with $\Delta_f = 120$ kHz.

SS burst periodicity – Fig. 3.10b shows the dependency of the overhead for initial access on T_{SS}, which follows an inverse proportionality law. In particular, for very small T_{SS} (i.e., 5 ms) the impact of the SS bursts with repetitions in frequency is massive, with up to 43% of the resources allocated to the SS blocks. For $T_{\text{SS}} = 20$ ms or higher, instead, the overhead is always below 10%.

Impact of beam reporting – For the SA case, as reported in Table 3.4, the completion of the beam reporting procedure for initial access may require an additional overhead, due to the need for the system to allocate possibly multiple RACH resources[4] for the reporting operations. Conversely, for the NSA case, the beam decision is forwarded through the LTE overlay and requires a single RACH opportunity, with a total overhead of $0.0894 \cdot 10^{-3}$. Nevertheless, from Table 3.4, we notice that the SA additional reporting overhead is quite limited due to the relatively small number of directions that need to be investigated at this stage, especially when designing digital beamforming solutions.

[4]According to the 3GPP agreements [121], a bandwidth of 10 MHz (for $\Delta_{f,\text{RACH}} = 60$ kHz) or a bandwidth of 20 MHz (for $\Delta_{f,\text{RACH}} = 120$ kHz) is reserved for the RACH resources.

3.5 Conclusions and Design Guidelines

In this chapter, we compared the performance of different IA solutions for 5G cellular networks operating at mmWave frequencies. Overall, it is possible to identify some guidelines for the configuration of the IA framework and the deployment of an NR network. The goal is to highlight which are the main trade-offs between accuracy, responsiveness and overhead and the design parameters for IA.

Subcarrier spacing Δ_f When using a smaller subcarrier spacing (i.e., $\Delta_f = 120$ kHz) it is possible to achieve a higher accuracy (i.e., smaller misdetection probability), either because the impact of noise is less relevant, when frequency diversity is not used, or because it is possible to allocate a larger number of repetitions, when frequency diversity is used. This last option comes however at the price of an increase in the overhead in the order of 2 times, while the accuracy gain for the configuration with $\lambda = 30$ gNB/km^2 and the 4×4 antenna arrays is in the order of 23%, according to Fig. 3.6. A smaller subcarrier spacing has also a negative effect on the reactiveness, as shown in Fig. 3.9, since the OFDM symbols last longer and the SS blocks sweep takes more time.

Frequency diversity The repetition in frequency of multiple SS signals for the same OFDM symbol results in an increased accuracy (e.g., up to 45%, when $\lambda = 60$ gNB/km^2 and considering the 4×4 array configuration). The overhead is, however, from 5 to 11 times higher in our setup (according to the Δ_f used), thus there is a trade-off between the amount of resources to allocate to the users that are already connected (which is higher if frequency diversity is not used) and the opportunity to discover new users (which increases with frequency diversity for the SS blocks). However, notice that the accuracy gain reduces when increasing the array dimension (e.g., when $\lambda = 60$ gNB/km^2 and considering the 64×4 array configuration, a gain of just 15% is achieved, as seen from Fig. 3.6). In those circumstances, it may not be desirable to adopt a frequency diversity scheme which would inevitably increase the overhead while only providing marginal accuracy gain.

Number of SS blocks in a burst N_{SS} This parameter has a fundamental impact on the reactiveness, since a higher number of SS blocks per burst increases the probability of completing the sweep in a single burst and thus prevents T_{IA} from being dependent on T_{SS}. The number of SS blocks per burst, however, increases also the overhead linearly. N_{SS} has a strict relationship with the number of directions to be swept, i.e., with both the beamforming architecture and the number of antennas: if, for example, hybrid or digital beamforming is used at the receiver, a larger number of antennas (i.e., narrower beams) can be supported even with a smaller N_{SS}, as shown in Fig. 3.7

SS burst periodicity T_{SS} The periodicity of a burst has an impact on the reactiveness for initial access, since a smaller T_{SS} enables a larger number of opportunities in which a UE can receive synchronization signals. However, if the beam sweeping procedure is completed in a single burst, T_{SS} does not impact T_{IA} as previously defined.

gNB density λ_b As the network density increases, the accuracy and the average received power increase, and this allows a larger number of users to be served by a mmWave network. Besides the cost in terms of equipment and energy, a higher density has also a negative effect on the interference [55].

Beamforming architecture K_{BF} A digital beamforming architecture at the receiver side would improve the reactiveness of the IA scheme and decrease the overhead, without penalizing the accuracy. The same improvement in terms of reactiveness and overhead can be achieved with an omnidirectional receiver, but the accuracy would decrease with a loss of around 30% (when $\lambda = 30$ gNB/km^2) with respect to the $M_{gNB} = 64$ configuration, as displayed in Fig. 3.5. The complexity of the transceiver implementation and the energy consumption [122] are, however, two important parameters that must be taken into account. A hybrid configuration could represent a trade-off between an improved reactiveness and a simpler and less consuming transceiver design.

Antenna Arrays M_{gNB}, M_{UE} The antenna array is one of the parameters that has the largest impact on the accuracy. A larger number of antennas enable narrower beams and higher accuracy, since the received power at the UE increases. The width of the beam has, however, an inverse relationship with the number of directions to scan, thus configurations that provide a higher accuracy perform worse in terms of reactiveness and overhead. Notice that the choice of the antenna array and of the beam design is strictly tied to the beamforming architecture (if digital or hybrid beamforming is used then narrower beams can be supported without penalizing reactiveness and overhead) and the configuration of the SS bursts (a large number of directions to be swept with a limited number of SS blocks per bursts has a negative impact on the reactiveness).

SA vs. NSA The implementation of a standalone scheme generally guarantees more reactive access capabilities. The reason is that faster beam reporting operations are ensured if multiple SS blocks and RACH opportunities can be allocated within a single SS burst. On the other hand, a non-standalone framework may be preferable to reduce the impact of the overhead in the beam reporting phase. Moreover, a non-standalone architecture is also better than an SA one when it is not possible to allocate in the same SS burst the SS blocks for the first sweep and the subsequent RACH opportunities, because for example there are too many directions to monitor at the gNB. Finally, NSA enables a centralized beam decision: unlike in traditional attachment policies based on pathloss measurements, by leveraging on the presence of an eNB operating at sub-6 GHz frequencies, an NSA-based beam association can be performed by taking into account the instantaneous load conditions of the surrounding cells, thereby promoting fairness in the whole cellular network [123].

Overall, it is possible to identify some guidelines for the configuration of the IA framework and the deployment of an NR network at mmWave frequencies. First, a setup of N_{SS}, the RACH resources, the beamforming and the antenna array architectures that allows the completion of the beam sweeping and reporting procedures in a single burst is preferable, so that it is possible to increase T_{SS} (e.g., to 20 or 40 ms).

Second, the adoption of a frequency diversity scheme for the SS blocks depends on the load of the gNBs: if many users are connected to a certain gNB, this could disable the frequency diversity to both reduce the overhead and avoid discovering new users. Third, with low network density, larger antenna arrays make it possible to detect farther users, and provide a wider coverage but, as λ_b increases, it is possible to use a configuration with wide beams for SS bursts (so that it is more likely to complete a sweep in a single burst).

Finally, when considering static and dense scenarios which are marginally affected by the variability of the mmWave channel, a standalone architecture is preferable for the design of fast initial access procedures, since it enables rapid beam reporting operations.

Chapter 4

Mobility Management in 5G Millimeter Wave Cellular Networks

4.1 Introduction

We reviewed in Sec. 2.1 how mmWave spectrum offers the availability of huge bandwidths and thus has been considered as a means to satisfy ever increasing mobile users' traffic demands. We also reviewed in Sec. 2.2 how, to overcome the high isotropic propagation loss experienced at these frequencies, highly directional antennas are required to achieve sufficient link budget in wide area networks. The consequence is the need for precise alignment of the transmitter and the receiver beams at every stage of the communication. In Chapter 3 we presented the implications that directionality has on idle users accessing the network for the first time, i.e., when performing initial access. However, the dynamics of the mmWave channel, especially in high-speed scenarios, imply that the directional path to any cell can deteriorate rapidly [67]. Hence, the beam selected during initial access could rapidly change, thus requiring each mobile terminal to constantly monitor each potential directional link. This is achieved through a set of operations, including *beam tracking*, *handover*, and *radio link failure recovery*, generally known as *mobility management*, which are periodically triggered to help maintain connectivity. Such procedures, however, introduce some latency which lowers the rate at which the network can adapt, and can be a major obstacle in providing robust service in the face of variable link quality.

4.1.1 Mobility Management in 5G Networks: Limitations

In addition to the limitations described in Sec. 3.1.2 for idle users, the design of mobility management procedures for connected users in 5G mmWaves networks pose new challenges as a consequence of the channel dynamics that are typically experienced in mobility scenarios.

Small cell deployments The combination of the high propagation loss and the blockage phenomenon at mmWaves advocates for a high-density deployment of gNBs. With such network architecture, mmWave base stations would be deployed as small cell. This may exacerbate frequent

This chapter is based on the contributions presented in **[J2, J3, J5, J6, C2, B1]**. Part of the results included in this chapter is also based on joint work with Michele Polese.

handovers between adjacent cells, which is a potential drawback of mmWave systems, even for fixed UEs, due to their vulnerability to random obstacles.

Topology changes Most existing mobility management approaches are based on periodical beam sweeping, where the transmitting and the receiving nodes scan the angular space to search for the best beam pair. However, these solutions are limited by the potentially high speed of terminals, since suitable beam pair may not last long enough to allow the completion of a data exchange, thus resulting in transmission errors. Unpredictable topology changes is another potential drawback of mmWave systems.

Doppler effect The increased Doppler effect in mobility scenarios could make the assumption of channel reciprocity not valid and could impair the feedback over mmWave links, which is a potential point of failure for beam sweeping.

Transport-layer characteristics The fast channel dynamics and the delays associated to mobility management operations, first and foremost packet forwarding during handovers, can have significantly severe effects throughout the whole protocol stack [124]. Some recent work, e.g., [56], has demonstrated that, if TCP is used at the transport layer, the rapid variations in channel quality and the frequent transitions from LOS to NLOS, as typical in mmWave systems, can result in bufferbloat and slow growth of the congestion window, leading to high delays and dramatic under-utilization of the channel. Channel outages and radio link failures can also be source of retransmission timeouts and connection resets.

4.1.2 Motivations and Chapter Structure

Given the limitations outlined in Sec. 4.1.1, it is important to design and dimension mobility management solutions that are specifically tailored to address the requirements and characteristics of mmWave cellular networks. In this regard, the main research issues are 1) how to adapt and manage the beam steering direction of connected users towards their respective serving cells and 2) how to implement fast and efficient handover mechanisms when a mere beam adaptation operation is not sufficient to maintain connectivity.

In this chapter, both aspects are investigated. In Sec. 4.2 we will review related work on mobility management procedures at mmWave frequencies. In Sec. 4.3, we will provide an overview of the beam management frameworks that have been standardized by 3GPP as part of NR specifications. We focus on NSA and SA architectures, according to whether the control plane is managed with the support of an LTE overlay or not, respectively. We will analyze the reactiveness (i.e., how quickly the frameworks are able to detect an updated channel condition and properly adapt the beam), and the overhead (i.e., how many time and frequency resources should be allocated for the measurement operations). When beam management operations alone cannot recover acceptable QoS, handovers must then be triggered. In Sec. 4.4 we will illustrate a novel NSA uplink measurement system that, with the joint help of a local coordinator operating in the legacy band, guarantees continuous monitoring of the channel propagation conditions and allows for the design of fast and efficient handover for connected users. In Sec. 4.5, we face the challenge of resolving radio link failures, which occur when the quality of an associated control channel falls below a certain threshold (one of the most critical and common network events in a mmWave environment). DL and UL, as well as NSA and SA configurations, are investigated. Finally, in Sec. 4.6 we identify guidelines for the design of mobility management strategies at mmWaves.

Table 4.1: Literature on mobility management strategies for mmWave networks.

Topic	Relevant References
Beam Management [53]	[126–128] Mobility-aware strategies. [73,74,123,129–131] Multi-connectivity solutions.
Handover [110]	[132,133] Below-6 GHz solutions [134–136] Above-6 GHz standalone solutions [110,123,137–140] Above-6 GHz multi-connectivity solutions

4.2 Related Work

Measurement reporting for mobility management is quite straightforward in LTE [125]: the DL channel quality is estimated from an omnidirectional signal called the Cell Reference Signal (CRS), which is regularly monitored by each UE in connected state to create a wideband channel estimate that can be used both for demodulating downlink transmissions and for estimating the channel quality [67]. However, when considering mmWave networks, in addition to the rapid variations of the channel, CRS-based estimation is challenging due to the directional nature of the communication, thus requiring the network and the UE to constantly monitor the direction of transmission of each potential link. In addition, the UE and the gNB may only be able to listen to one direction at a time, thus making it hard to receive the control signaling necessary to switch paths.

Beam management The issue of designing efficient beam management solutions for mmWave networks was addressed in [126], in which the author designed a mobility-aware user association strategy to overcome the limitations of the conventional power-based association schemes in a mobile 5G scenario. Other relevant papers on this topic include [127], in which the authors proposed smart beam tracking strategies for fast mmWave link establishment and maintenance under node mobility. In [128], the authors proposed the use of an extended Kalman filter to enable a static base station, equipped with a digital beamformer, to effectively track a mobile node equipped with an analog beamformer after initial channel acquisition, with the goal of reducing the alignment error and guarantee a more durable connectivity. Recently, robust beam tracking schemes have been designed by leveraging out-of-band information to estimate the mmWave channel. In [74,110,123,129] an approach where 5G cells operating at mmWaves (offering much higher rates) and traditional 4G cells below 6 GHz (providing much more robust operation) are employed in parallel have been proved to enable fast and resilient tracking operations. In [73], a framework which integrates both LTE and 5G interfaces was proposed as a solution for mobility-related link failures and throughput degradation of cell-edge users, relying on coordinated transmissions from cooperating cells are coordinated for both data and control signals. In [130], a novel approach for analyzing and managing mobility in joint sub-6GHz–mmWave networks was proposed by leveraging device caching along with the capabilities of dual-mode base stations to minimize handover failures, reduce inter-frequency measurement, reduce energy consumption, and provide seamless mobility in emerging dense heterogeneous networks. Moreover, the authors in [131] illustrated how to exploit spatial congruence between signals in different frequency bands and extract mmWave channel parameters from side information obtained in another band. Despite some advantages, the use of out-of-band information for the 5G control plane management poses new challenges that remain unsolved and which deserve further investigation.

Handover Although the literature on handover in more traditional sub-6 GHz heterogeneous networks is quite mature, papers on handover management for mmWave 5G cellular are very recent, and research in this field has just started. The survey in [132] presented multiple vertical handover decision algorithms that are essential for heterogeneous wireless networks, while article [133] investigated the management of the handover process between macro, femto and pico cells, proposing a theoretical model to characterize the performance of a mobile user in heterogeneous scenarios as a function of various handover parameters. However, these works focused on low frequency legacy cellular systems. When dealing with mmWaves, frequent handover, even for fixed UEs, is a potential drawback that needs to be addressed. In [134], the handover rate in 5G systems was investigated and in [135] a scheme for handover management in high-speed railway was proposed by employing the received signal quality from measurement reports. The authors of [136] presented an architecture for mobility, handover and routing management.

Handover mechanisms can also be designed in such a way that mmWave operations are supported by the LTE overlay, to provide additional robustness. In [137, 140] the impact of user mobility in multi-tier heterogeneous networks was analyzed and a framework was proposed to solve the dynamic admission and mobile association problem in a wireless system with mobility. Zang *et al.*, in [138], proposed a new efficient handover decision algorithm based on a Markov Decision Process to optimize the overall service experience of users in mmWave heterogeneous networks. In [139], a mmWave-aware handover scheme was proposed to use selective control/user-plan split combined with a non-coherent cooperative multi-point joint transmission. The combination of this strategy with fine-tuning of the measurement interval resulted in reduction of handover failure rates and throughput improvement. As part of our previous contributions, in [110] we provided the first comprehensive end-to-end evaluation of handover mechanisms in mmWave cellular systems. The simulation framework included detailed measurement-based channel models to realistically capture spatial dynamics of blocking events, as well as the full details of MAC, RLC and transport protocols. Compared to conventional handover mechanisms, the study had revealed significant benefits of the proposed method under several metrics. Additionally, in [123] we proposed an MC UL approach to realize energy-efficient handover for mobile terminals.

4.3 Beam Management in 5G Networks: the 3GPP NR Approach

In line with the 3GPP design for NR, beam management is required for users in connected mode to maintain alignment of the transmitter and receiver, an operation that is also defined as *tracking*. In this context, after reviewing in Sec. 4.3.1 the most relevant measurement signals for beam management in 3GPP NR, in Sec. 4.3.2 we present the beam management frameworks we consider, while numerical performance evaluation of such schemes is provided in Sec. 4.3.3.

4.3.1 3GPP NR Measurement Signals for Beam Management

The LTE Channel State Information - Reference Signals (CSI-RSs) allow connected UEs to regularly estimate the channel conditions and report Channel Quality Information (CQI) to their serving base station. Likewise, in 3GPP NR, these signals can be used as Radio Resource Management (RRM) measurements and for mobility in connected mode [9]. However, the UEs must know in which time and frequency resources the CSI-RS signals will be sent. In Sec. 3.3.1, we saw that SS blocks and bursts can be used for periodic synchronization signal transmissions from the gNBs to perform initial access.

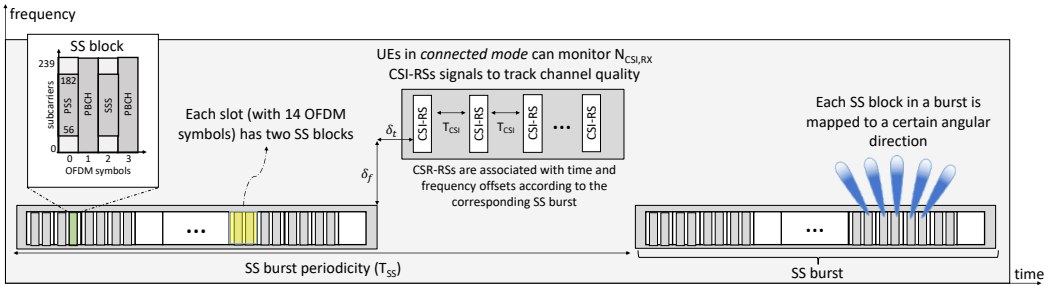
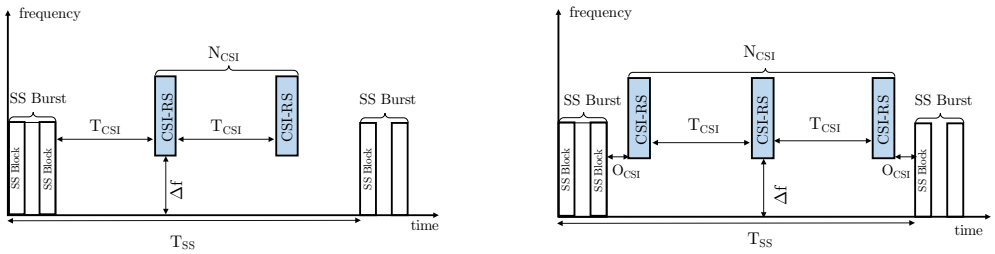


Fig. 4.1: Downlink measurement signals for NR systems. SS blocks are sent every T_{SS}, and embed time and frequency offsets (δ_t and δ_f, respectively) for the allocation of CSI-RSs within the frame structure. CSI-RS signals are sent every T_{CSI}.

As long as each CSI-RS is represented by a unique identifier, it is possible to configure multiple CSI-RSs using the same SS burst, so that the UE can first synchronize with a certain cell after initial access (using the SS bursts), and then use those to identify the CSI-RS resources [109,141]. This is illustrated in Fig. 4.1. The CSI-RS configuration, which is UE-specific, should indicate the time/frequency offsets with respect to the associated SS burst and the possible CSI-RS periodicity. Fig. 4.2 shows the two options we consider for the time offset of the CSI-RS transmissions. The first option, shown in Fig. 4.2a, allows the transmission of the first CSI-RS T_{CSI} ms after the end of an SS burst. The second one, shown in Fig. 4.2b, has an additional parameter, i.e., an offset in time O_{CSI}, which represents the time interval between the end of the SS burst and the first CSI-RS. Although an NR network may transmit CSI-RS for measurements in the full bandwidth, the reference signals may also be broadcast through a subset of the available frequency resources (with a minimum of 50 resource blocks [142]) if this is sufficiently large for a proper channel estimation at the receiver [142, 143]. CSI-RSs may span $N = 1$, 2 or 4 OFDM symbols [11, 144]. Moreover, the 3GPP defines different activation methodologies for the CSI-RS measures. For *periodic* or *semi-persistent* CSI-RS transmissions, the following periodicities (in slots) are supported: $T_{\text{CSI}} \in \{5, 10, 20, 40, 80, 160, 320, 640\}$ [11]. For *semi-persistent* and *aperiodic* CSI-RS transmissions, the resources are configured and pre-allocated by the higher layers, while the periodic configuration broadcasts measurement signals with regularity. For *aperiodic* transmissions, a single set of CSI-RS triggering states is higher-layer configured, therefore a UE is not expected to receive more than one aperiodic CSI-RS in a given slot, i.e., it will not transmit more than one aperiodic CQI report to its serving cell [142]. Notice that CSI-RSs may have a significantly higher time/frequency density compared to that of the SS blocks, thus implying higher overhead but, at the same time, more responsive tracking.

As we assessed previously in this chapter, when considering directional communications, the



(a) The first CSI-RS is sent T_{CSI} ms after an SS burst.

(b) The first CSI-RC is sent O_{CSI} ms after an SS burst.

Fig. 4.2: Examples of CSI-RS measurement window and periodicity configurations. SS blocks embed time and frequency offsets indicating the time and frequency allocation of CSI-RS signals.

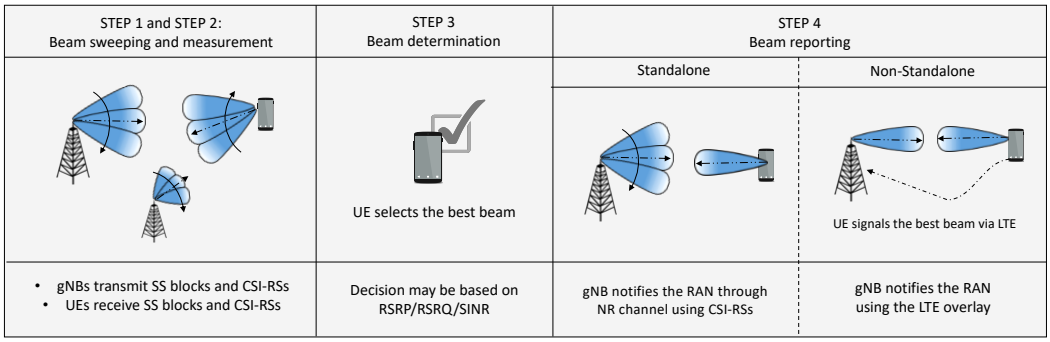


Fig. 4.3: Representation of the beam management procedures defined by the 3GPP in the NR specifications, with SA and NSA architectures.

best directions for the beams of the transceiver need to be periodically identified (e.g., through beam management operations), in order to maintain the alignment between the communicating nodes. For this purpose, SS- and CSI-based measurements can be jointly used to reflect the different coverage which can be achieved through different beamforming architectures [111, 145]. As far as CSI signals are concerned, the communication quality can be derived by averaging the signal quality from the $N_{\text{CSI,RX}}$ best beams among all the available ones, where the value of $N_{\text{CSI,RX}}$ can be configured to 1 or more than 1 [109, 141][1]. Nevertheless, to avoid the high overhead associated with wide spatial domain coverage with a huge number of very narrow beams, on which CSI-RSs are transmitted, it is reasonable to consider transmitting only subsets of those beams, based on the locations of the active UEs. This is also important for UE power consumption considerations [109, 146]. For example, the measurement results based on SS blocks (and referred to a subset of transmitting directions) can be used to narrow down the CSI-RS resource sets based on which a UE performs measurements for beam management, thereby increasing the energy efficiency.

4.3.2 3GPP NR Beam Management Frameworks

3GPP NR specifications include a set of basic procedures for beam management [64], as illustrated in Fig. 4.3, which are based on the reference signals described above. We consider both SA and NSA deployments, according to whether or not the LTE overlay is exploited to support control operations. For the SA deployment, the beam management scheme proposed in 3GPP NR has a very similar structure to the IA procedures described in Sec. 3.3.2 and is composed of the following steps:

- **Step 1 (Beam sweeping)**: gNBs and UEs sequentially transmit and receive, respectively, synchronization and reference signals.
- **Step 2 (Beam measurements)**: UEs evaluate the quality of the signals collected from the SS bursts and the CSI-RSs transmitted in the previous step. CSI-RSs, in particular, cover a set of directions which may or may not cover the entire set of available directions according to the users' needs, as explained in Sec. 4.3.1.
- **Step 3 (Beam determination)**: UEs select the beam(s) through which they experienced the maximum SNR, if above a predefined threshold.

[1] In [109] it is specified that, for the derivation of the quality of a cell, the UEs should consider an absolute threshold, and average the beams with quality above the threshold, up to $N_{\text{CSI,RX}}$ beams. If there are no beams above threshold, then the best one (regardless of its absolute quality) should be selected for the cell quality derivation.

- **Step 4 (Beam reporting)**: UEs send to the RAN information on the quality of the received beamformed signals and on the decision made in the previous step. While, for IA (Sec. 3.3.2) the UE has to wait for the gNB to schedule the RACH opportunity towards the best direction that it just determined (this may require an additional complete directional scan of the gNB), for beam management in connected mode the UE can directly provide feedback using the mmWave control channel it has already established.

When the quality of the received control signals falls below a predefined threshold (e.g., in the case of beam failure), beam reporting operations cannot be performed since no directions can be recovered for the feedback using CSI-RS. Full-stack end-to-end recovery operations, including initial access using the SS bursts, must be triggered instead, which may take up to several tens of milliseconds. We claim that faster and more efficient beam reporting can be guaranteed if an NSA architecture is preferred [13]. In this case, beam management reports can be sent through the robust LTE overlay, thereby allowing the feedback to be received even while the user experiences a service unavailability in the mmWave bands. With such an approach, the delay for beam reporting would correspond to the latency of a traditional LTE connection, which may be significantly lower than the time it takes to perform a recovery operation in an SA deployment. The LTE radio may also serve the UE's traffic requests until the mmWave directional communication is successfully restored, thereby offering service continuity.

4.3.3 Performance Results

Along the lines of the performance analysis we presented in Sec. 3.4 for IA, we now evaluate the performance of the beam management architectures presented in the previous subsection. The performance is assessed in terms of *reactiveness* (T_{tr}) , i.e., the time required to receive the first CSI-RS after an SS burst, and thus react to channel variations or mobility in order to eventually switch beams, and *overhead*, i.e., the amount of time and frequency resources allocated to the framework with respect to the total amount of available resources, taking into account both SS block and CSI-RS transmissions.

The performance is affect by the parameters listed in Sec. 3.4.1, namely (i) the 3GPP NR frame structure, (ii) the frequency diversity scheme, (iii) the SS block configuration, (iv) the array geometry (i.e., the number of antenna elements, M_{UE} and M_{gNB}, at the UE and gNB, respectively), (v) the beamforming architecture (i.e., analog, digital or hybrid), (vi) the network deployment (i.e., the number of users $\mathcal{N}$ per sector of the gNBs and the density of base stations λ_b). In addition, the impact of the CSI-RS configuration is evaluated. In fact, there are different options for the configuration of the CSI-RS structure. These options include (i) the number N_{CSI} of CSI-RS per SS burst period, (ii) the CSI-RS periodicity $T_{\text{CSI,slot}} \in \{5, 10, 20, 40, 80, 160, 320, 640\}$ slots, and (iii) the offset O_{CSI} with respect to the end of an SS burst. We will use the parameter $T_{\text{CSI}} = T_{\text{CSI,slot}} T_{\text{slot}}$ to represents the absolute CSI-RS periodicity in ms, where T_{slot} is the slot duration in ms as is given in Eq. (3.1). These settings will be specified by the system information carried by the SS blocks of each burst. Other CSI-related parameters are the number of symbols of each CSI-RS transmission, i.e., $N_{\text{symb,CSI}} \in \{1, 2, 4\}$, and the portion of bandwidth ρB allocated to the CSI-RSs. Moreover, the user will listen to $N_{\text{CSI,RX}}$ CSI-RSs through an equivalent number of directions, when in connected state. We will consider $N_{\text{CSI,RX}} \in \{1, 4\}$.

4.3.3.1 Reactiveness Results

Analysis – Assume that $\mathcal{N}$ UEs are uniformly distributed in the space covered by $k = N_{\theta,\text{gNB}} N_{\phi,\text{gNB}}$ beams available at the gNB. Moreover, each UE has to monitor $N_{\text{CSI,RX}}$ directions. Given that a UE

may or may not be in LOS, it is not obvious that these directions will be associated to the closest beams with respect to the one selected during the initial access. Therefore, we also assume that this scenario is equivalent to a scenario with $n = \mathcal{N} N_{\text{CSI,RX}}$ uniformly distributed UEs, each of them monitoring a single direction. We will refer to n as the number of measures. Consequently, on average there are n/k measurements for the area belonging to each beam, if the beams divide the space into equally sized regions. Therefore, if $n \geq k$, a CSI-RS is needed in each beam, otherwise it is sufficient to send at least n CSI-RSs, and thus the total number of CSI-RS that need to be transmitted is on average $Z_{\text{CSI}} = \min\{n, k\}$. Depending on the combination of T_{SS} (i.e., the SS burst periodicity), $T_{\text{CSI}} = T_{\text{CSI,slot}} T_{\text{slot}}$ and Z_{CSI}, it may not be possible to allocate all the CSI-RS transmissions between two consecutive SS bursts. Notice that after the end of an SS burst, there are $T_{tot,\text{CSI}} = T_{\text{SS}} - D_{\text{max,SS}}$ ms available for the CSI-RS transmission, where $D_{\text{max,SS}}$ is the maximum duration of an SS burst. Then, the number N_{CSI} of CSI-RS that can be allocated between two SS bursts may depend on which of the options shown in Fig. 4.2 is chosen.

Option 1: the first CSI-RS is transmitted T_{CSI} ms after the transmission of the SS burst. In this case, $N_{\text{CSI}} = \lfloor T_{tot,\text{CSI}} / T_{\text{CSI}} \rfloor$, and single periodicity is not enough if $Z_{\text{CSI}} > N_{\text{CSI}}$. For option 1, the metric $T_{\text{tr,opt1}}$ is given by

$$T_{\text{tr,opt1}} = \frac{\sum_{p=0}^{\left\lfloor \frac{Z_{\text{CSI}}}{N_{\text{CSI}}} \right\rfloor - 1} \left(\sum_{i=1}^{N_{\text{CSI}}} (p T_{\text{SS}} + i T_{\text{CSI}}) \right) + \sum_{i=1}^{Z_{\text{CSI}} \bmod N_{\text{CSI}}} \left(\left\lfloor \frac{Z_{\text{CSI}}}{N_{\text{CSI}}} \right\rfloor T_{\text{SS}} + i T_{\text{CSI}} \right)}{Z_{\text{CSI}}} \tag{4.1}$$

The last sum accounts for the case $Z_{\text{CSI}} < N_{\text{CSI}}$ and for the CSI-RS in the last SS burst periodicity when $Z_{\text{CSI}} > N_{\text{CSI}}$. The sum over p, instead, accounts for $Z_{\text{CSI}} \geq N_{\text{CSI}}$.

Option 2: thanks to the additional parameter O_{CSI} it is possible to transmit $N_{\text{CSI}} = \lceil T_{tot,\text{CSI}} / T_{\text{CSI}} \rceil$, as shown in Fig. 4.2b. The offset is computed as

$$O_{\text{CSI}} = \frac{T_{tot,\text{CSI}} - (N_{\text{CSI}} - 1) T_{\text{CSI}}}{2}. \tag{4.2}$$

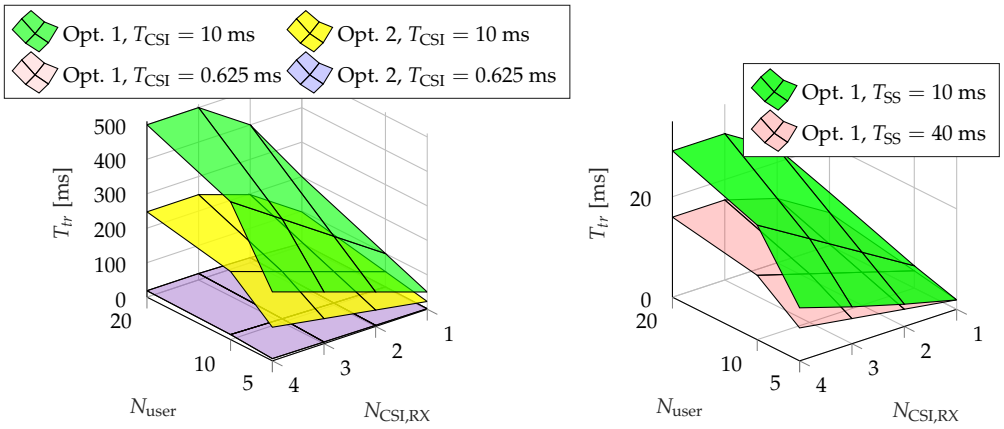
The metric $T_{\text{tr,opt2}}$ takes into account also O_{CSI} and is computed as

$$T_{\text{tr,opt2}} = \frac{\sum_{p=0}^{\left\lfloor \frac{Z_{\text{CSI}}}{N_{\text{CSI}}} \right\rfloor - 1} \left(\sum_{i=0}^{N_{\text{CSI}}-1} (p T_{\text{SS}} + i T_{\text{CSI}} + O_{\text{CSI}}) \right) + \sum_{i=0}^{Z_{\text{CSI}} \bmod N_{\text{CSI}} - 1} \left(\left\lfloor \frac{Z_{\text{CSI}}}{N_{\text{CSI}}} \right\rfloor T_{\text{SS}} + i T_{\text{CSI}} + O_{\text{CSI}} \right)}{Z_{\text{CSI}}} \tag{4.3}$$

Notice that if $Z_{\text{CSI}} > N_{\text{CSI}}$, a signal in a certain direction could be either received as SS block in the next burst, or as CSI-RS, depending on how the transmission of SS blocks and CSI-RSs is scheduled.

Scheduling options, number of users and CSI-RS periodicity – Fig. 4.4a shows the value of T_{tr} for different parameters, such as the different scheduling option 1 or 2, the number of users per gNB $\mathcal{N}$ and of directions of interest $N_{\text{CSI,RX}}$, for SS burst periodicity $T_{\text{SS}} = 20$ ms and 64 antennas at the gNB. The fundamental parameter is the periodicity of the CSI-RS transmission: only a small CSI-RS periodicity makes it possible to sweep all the directions to be covered during a relatively short interval, and to avoid the dependency on T_{SS}. Moreover, if the periodicity is small (i.e., $T_{\text{CSI}} =$ 0.625 ms, or 5 slots with $\Delta_f = 120$ kHz), then there is no difference between the two scheduling options, while this becomes notable for $T_{\text{CSI}} = 10$ ms, as expected.

SS burst periodicity – Fig. 4.4b compares two different T_{SS} periodicities, i.e., 10 and 40 ms, using the smallest $T_{\text{CSI,slot}}$ available (i.e., 5 slots, or 0.625 ms at $\Delta_f = 120$ kHz). It can be seen that using a higher T_{SS} would allow a decreased T_{tr}, since more CSI-RSs can be scheduled between two SS bursts and consequently a larger number of directions can be swept. For the sake of completeness, Fig. 4.5 shows the number of CSI-RSs that can be scheduled in between two SS bursts as a function of T_{SS} and of the different scheduling options and periodicities. Since in a mmWave scenario there may



(a) $M_{gNB} = 64$, analog, $T_{SS} = 20$ ms

(b) $M_{gNB} = 64$, analog, $T_{CSI} = 0.625$ ms

Fig. 4.4: Reactiveness performance of beam management using CSI-RSs for Option 1 and Option 2, as described in Fig. 4.2, as a function of different parameters (e.g., T_{CSI}, T_{SS}), for $\Delta_f = 120$ kHz.

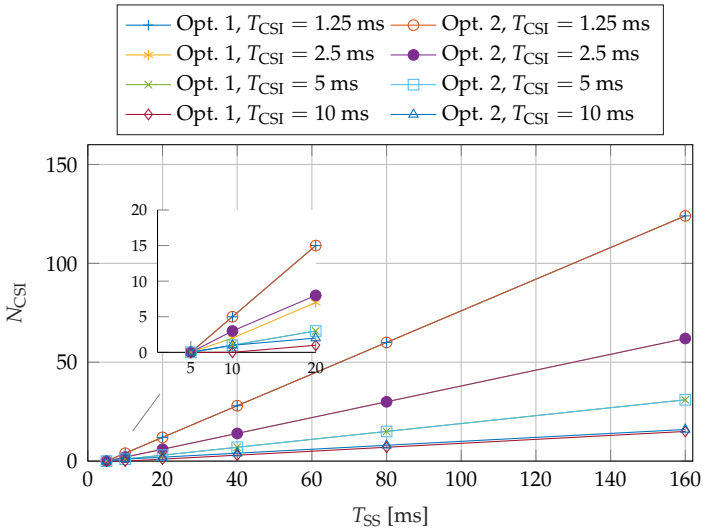


Fig. 4.5: N_{CSI} as a function of the T_{SS} and T_{CSI} periodicities.

be a need to scan a large number of CSI-RSs, it is advisable to either use an adaptive scheme for the scheduling of CSI-RSs, which adapts the periodicity according to the number of users in the different directions, or adopt a conservative approach and use a short T_{CSI} interval.

Limits on the CSI-RS periodicity – Since the CSI-RSs that a user receives from multiple base stations should not overlap in time and frequency (otherwise the SNR value would be over-estimated), there is a maximum number of neighboring cells that a gNB can support. According to [11], there are 4 symbols per slot in which a CSI-RS can be sent (additional symbols are under discussion), and a CSI-RS can last 1, 2 or 4 symbols, each with bandwidth ρB. The total number of orthogonal CSI-RS transmission opportunities is

$$N_{CSI,\perp} = \frac{T_{SS} - D_{max,SS}}{T_{slot}} \frac{4}{N_{symb,CSI}} \left\lfloor \frac{1}{\rho} \right\rfloor, \tag{4.4}$$

where the first ratio is the number of slots in the time interval in which CSI-RSs can be scheduled,

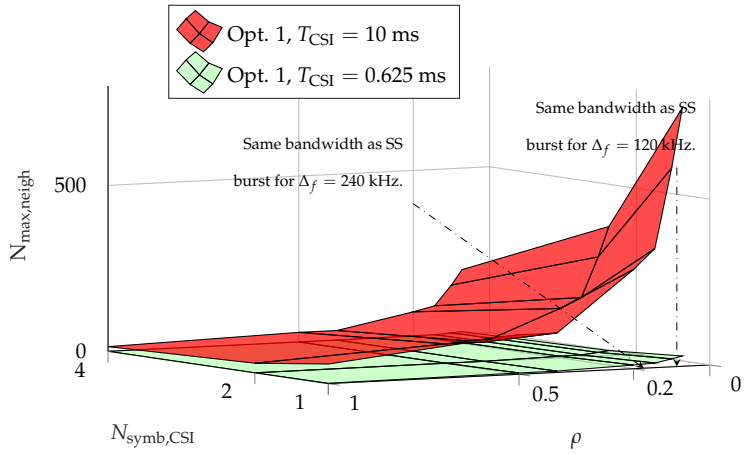


Fig. 4.6: $N_{\text{max,neigh}}$ as a function of $N_{\text{symb,CSI}}$ and ρ for different T_{CSI} periodicities, with $T_{\text{SS}} = 20$ ms and $\Delta_f = 120$ kHz.

and the second and third express the number of CSI-RSs per slot (there are at most 4 OFDM symbols per slot for CSI-RSs). Then, the maximum number of neighbors that a gNB can support is

$$N_{\text{max,neigh}} = \left\lfloor \frac{N_{\text{CSI},\perp}}{N_{\text{CSI}}} \right\rfloor - 1, \tag{4.5}$$

with N_{CSI} computed as in the previous paragraphs. Fig. 4.6 reports the value of $N_{\text{max,neigh}}$ for a different number of OFDM symbols for the CSI-RSs and bandwidth scaling factor ρ, which ranges from 0.1 to 1, and represents also the bandwidth values corresponding to 240 subcarriers with $\Delta_f \in \{120, 240\}$ kHz, i.e., the bandwidth occupied by an SS burst. Notice that for the frequencies in the mmWave spectrum it is advisable not to use the entire bandwidth for CSI-RSs [143], and the number of neighbors of a mmWave gNB will be limited, given the short propagation distance typical of these frequencies. If $T_{\text{CSI}} = 10$ ms, then even when using 4 OFDM symbols and the whole bandwidth it is possible to support only 14 neighbors. Instead, when $T_{\text{CSI}} = 0.625$ ms it is not feasible to use the whole bandwidth and 4 symbols, but more conservative configurations should be adopted. For example, with $\rho = 0.072$ (i.e., 240 subcarriers with $\Delta_f = 120$ kHz) it is possible to support 15 or 31 neighbors, respectively with 2 or 1 OFDM symbols.

SA vs NSA – Notice that when the SA scheme is used and the UE experiences a link failure on all the $N_{\text{CSI,RX}}$ directions it is monitoring, then the UE has no choice but using the SS blocks in the SS burst to perform either a link recovery or a new initial access, and meanwhile it is not able to transmit or receive data or control information [110]. When an NSA architecture is used, instead, the UE could signal this event to the RAN on the lower-frequency control link, and the data plane can be switched to the sub-6 GHz RAT, and faster recovery options could be designed, for example, by instructing the UE to monitor additional CSI-RSs.

4.3.3.2 Overhead Results

Analysis – Besides the overhead introduced by the transmission of SSs blocks ($\Omega_{T_{\text{SS}}}$, evaluated for IA in Sec. 3.4.3.3), additional overhead is introduced by the transmission of CSI-RSs after the SS burst. The value of the overhead Ω_{CSI} depends on the number of symbols $N_{\text{symb,CSI}}$, the symbol duration T_{symb} (defined in Eq. (3.1)), and the bandwidth ρB for each CSI-RS, as well as on the number of

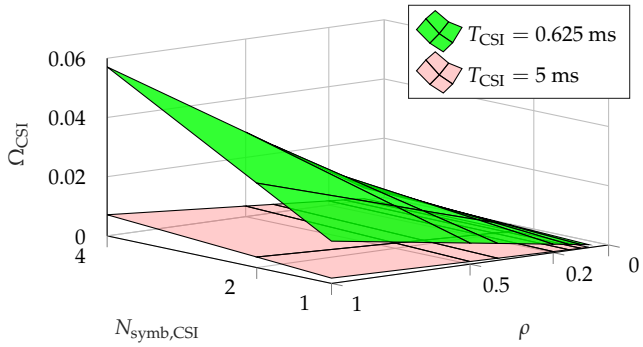


Fig. 4.7: Overhead Ω_{CSI} as a function of $N_{symb,CSI}$ and ρ, for different T_{CSI} periodicities, with $T_{SS} = 20$ ms.

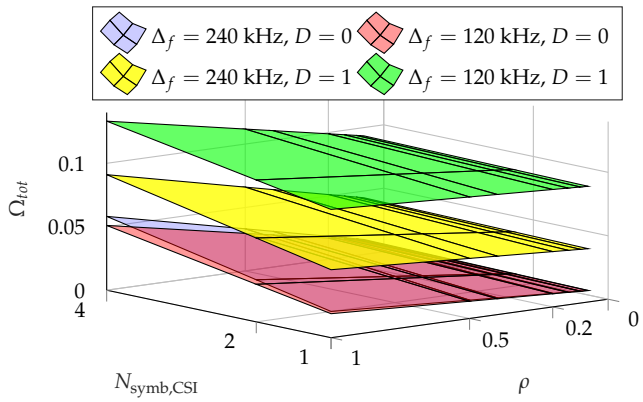


Fig. 4.8: Overhead Ω_{tot} as a function of $N_{symb,CSI}$ and ρ, for different subcarrier spacings Δ_f and repetition strategies. N_{SS} is set to the maximum value, i.e., 64, and $T_{CSI,slot} = 5$ slot.

CSI-RSs N_{CSI} computed as in Sec. 4.3.3.1 for the two CSI-RS scheduling options:

$$\Omega_{CSI} = \frac{N_{CSI} N_{symb,CSI} T_{symb} \rho B}{(T_{SS} - D_{max,SS}) B} = \frac{N_{CSI} N_{symb,CSI} T_{symb} \rho}{(T_{SS} - D_{max,SS})}. \tag{4.6}$$

The total overhead Ω, taking into account both the SS bursts and the CSI-RSs in T_{SS}, can then be computed as:

$$\Omega_{tot} = \frac{N_{CSI} N_{symb,CSI} T_{symb} \rho B + R_{SS}}{T_{SS} B}, \tag{4.7}$$

where R_{SS} is the total number of time and frequency resources scheduled for the transmission of N_{SS} SS blocks, as given in Eq. (3.6).

CSI-RS periodicity – The overhead due to the transmission of CSI-RSs is shown in Fig. 4.7 for different T_{CSI} periodicities and time and frequency resource allocation to the CSI-RSs. It is always below 0.008 with $T_{CSI} = 5$ ms, and below 0.06 for $T_{CSI} = 0.625$ ms. However, for practical values of the configuration of the CSI-RSs, in which the bandwidth for the reference signal is smaller than half of the bandwidth, then also for $T_{CSI} = 0.625$ ms the overhead is very small, i.e., below 0.028.

Impact of initial access and beam management – Ω_{tot} in Eq. (4.7) accounts both for IA and beam management operations, and its trend is shown in Fig. 4.8. It can be immediately seen that the largest impact is given by the term R_{SS} (which is due to IA) at the numerator and not by the CSI-RS-related overhead. The parameters on the x and y axes have indeed a limited effect on the gradient of the

surfaces, which are almost horizontal. The main difference is introduced by the different subcarrier spacings and repetition strategies. Notice that, contrary to what is shown in Fig. 3.10a for the sole IA-related overhead, there is a difference between the two different subcarrier spacings for the total overhead Ω_{tot} and for the CSI-RS-related overhead Ω_{CSI}, because we consider a different T_{symb} in Eq. (4.6), but the same ρ factor, thus a different number of subcarriers for the different values of Δ_f.

4.4 Handover in 5G Networks: a Novel Uplink Multi-Connectivity Scheme

Beam management operations allow the mobile terminals to adjust their beamforming direction to realign with their current serving gNB, as a consequence of network topology changes or channel variations. In case of high mobility, or when the users experience service unavailability, beam management operations alone might not be able to maintain connectivity. In this case, handing over to a stronger and more robust gNB could help recover acceptable communication capabilities, upon notifying the network.

In legacy LTE networks, the UE always connects to a single cell only, i.e., the switching of a UE's connection from a source cell to a target cell is a *Hard Handover (HH)*. In other words, the target gNB generates a Radio Resource Control (RRC) message that orders the UE to perform the handover, and the message is transparently forwarded by the source gNB to the UE [83]. However, due to the asynchronous nature of the LTE handover functionality with RA at every cell change, there is an undesirable temporary interruption gap at every handover. Field measurements reveal that the data interruption time at each handover ranges from, at best, 20-30 ms and up to 100 ms (or even more) for some networks [147], which is unsuitable in view of the strict latency constraints of future 5G networks (see Chapter 1). To overcome this issue, 5G systems will likely adopt a time-synchronized and RA-less handover functionality. Similar solutions have also been discussed in 3GPP, e.g., in [148], even though no decision has yet been taken at the time of writing.

Along these lines, in Sec. 4.4.1 we present a novel uplink multi-connectivity framework [123], which we will refer to as UL-ML scheme, that supports the design of faster network handover procedures – namely *Fast Switching (FS)* and *Secondary Cell Handover (SCH)*, described in Sec. 4.4.2 –, with respect to the standard standalone HH scheme [110]. The UL control enables the network to monitor the angular directions of communication to the UEs on all possible links simultaneously, so that, when a handover is necessitated, no directional search needs to be performed (this approach greatly saves switch time, since directional scanning dominates the delay in establishing a new link [85]). The MC feature, at the same time, enables a better resource allocation and mobility management compared to a standalone configuration. In fact, LTE connectivity can offer a ready backup in case the mmWave links suffer an outage and can be used to forward the scheduling and handover decisions to the user if the main propagation path is unavailable. We evaluate the proposed switching and handover protocols following the evaluation methodology we have developed in [56, 110, 129, 149, 150]. Network Simulator 3 (ns-3) [151] is used as a simulator, thereby making it possible to use detailed measurement-based channel models that can account for both the spatial characteristics of the channel and the channel dynamics arising from blocking and other large-scale events, which is important for a detailed and realistic assessment. In addition, ns-3 implements a 3GPP-like cellular protocol stack and features a MAC layer with Hybrid Automatic Repeat reQuest (HARQ), all the network-layer signaling, and an end-to-end transport protocol. Our study reveals several important findings on the interaction of transport layer mechanisms, buffering, and its relation with physical-layer handover delays. We demonstrate in Sec. 4.4.3 that the proposed FS and

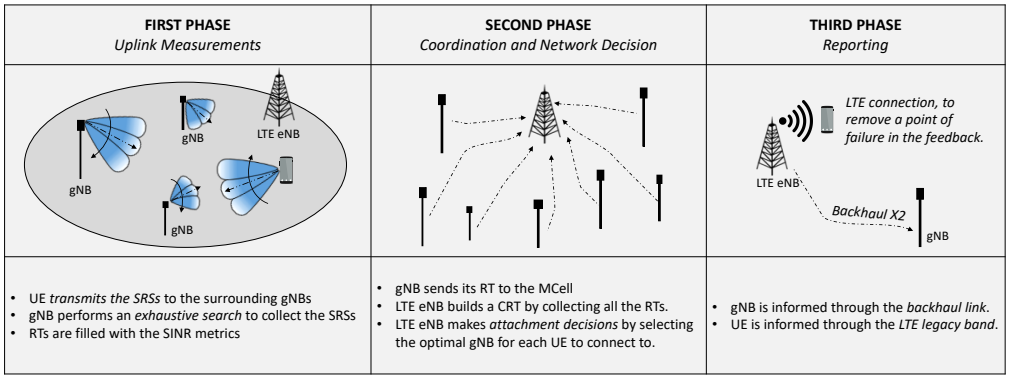


Fig. 4.9: Illustrative representation of the UL-MC scheme proposed for handover management.

SCH mechanisms offer significant performance improvements compared to the hard handover solution, including (i) reduced packet loss, (ii) reduced control signaling, (iii) reduced latency, and (iv) higher throughput stability. Moreover, we show that a dynamic Time-to-Trigger (TTT) adaptation should be preferred for handover management, since it can deliver non-negligible improvements in specific mobility scenarios in which state-of-the-art methods fail.

4.4.1 Proposed Uplink Multi-Connectivity (UL-MC) Scheme

In this subsection and as illustrated in Fig. 4.9, we present an innovative framework (proposed for the first time in [74,123]) to perform efficient handover operations in 5G networks. Motivated by the fact that the increasing heterogeneity in cellular networks is making the role of the uplink much more important [152], we consider an UL framework in which the measurements are based on reference signals transmitted by the mobile terminals rather than by the eNBs as in traditional cellular systems. Moreover, we consider a MC framework in which gNBs operating at mmWaves use an LTE eNB as a support for the control plane management [153] and UEs maintain multiple possible connections (i.e., LTE and mmWave overlays) to different cells to provide connectivity in case of failure in one of the network interfaces (as described in Sec. 2.3).

We consider a scenario in which each UE directionally broadcasts a Sounding Reference Signal (SRS) in a time-varying direction that continuously sweeps the angular space. Each potential serving gNB scans all its angular directions and monitors the strength of the received SRSs , building a Report Table (RT) based on the channel quality of each receiving direction, to better capture the dynamics of the channel. A centralized coordinator (which may reside in the LTE eNB) obtains complete directional knowledge from all the RTs sent by the potential cells in the network to make the optimal serving cell selection and scheduling decisions. In particular, due to the knowledge gathered on the signal quality in each angular direction for each gNB-UE pair, the coordinator is able to match the beams of the transmitter and of the receiver to provide maximum performance. We assume that nodes select one of a finite number of directions for measuring the signal quality, and we let N_{gNB} and N_{UE} be the number of directions at each gNB and UE, respectively. Supposing that $\mathcal{M}$ mmWave gNBs (interconnected via an X2 link, which may be a wired or wireless backhaul) are deployed within the coverage of the coordinator, the procedure works as follows.

- **Step 1 (Uplink Measurements)**: Each UE directionally broadcasts uplink SRSs in dedicated slots, steering through directions $d_1, \ldots, d_{N_{\text{UE}}}$, one at a time, to cover the whole angular space. The SRSs are scrambled by locally unique identifiers (e.g., C-RNTI) that are known to the mmWave gNBs and can be used for channel estimation. According to the 3GPP [109,154], the gNBs signal to the UEs the set of resources and directions to use for the transmission of the

RT (mmWave gNB$_j$)	
UE$_1$	SINR$_{1,j}$
UE$_2$	SINR$_{2,j}$
...	...
UE$_{\mathcal{N}}$	SINR$_{\mathcal{N},j}$

Complete Report Table (CRT)			
UE	mmWave gNB$_1$	...	mmWave gNB$_{\mathcal{M}}$
UE$_1$	SINR$_{1,1}$	...	SINR$_{1,\mathcal{M}}$
UE$_2$	SINR$_{2,1}$	...	SINR$_{2,\mathcal{M}}$
...	...	...	...
UE$_{\mathcal{N}}$	SINR$_{\mathcal{N},1}$	...	SINR$_{\mathcal{N},\mathcal{M}}$

Table 4.2: An example of RT (left) and CRT (right), referred to $\mathcal{N}$ users and $\mathcal{M}$ available mmWave gNBs in the network. We suppose that the UE can send the sounding signals through N_{UE} angular directions and each mmWave gNB can receive them through N_{gNB} angular directions. Each pair is the maximum SINR measured in the best direction between the gNB and the UE.

SRSs, that can span 1 to 4 OFDM symbols, and a portion of the entire bandwidth available at the UE [11]. If analog beamforming is used, each mmWave gNB scans through directions $D_1, \ldots, D_{N_{\text{gNB}}}$ one at a time or, if digital beamforming is applied, collects measurements from all of them at once. Each mmWave gNB fills a RT, as in Table 4.2 left, whose entries represent the highest Signal to Interference plus Noise Ratio (SINR) between UE$_i$, $i = 1, \ldots, \mathcal{N}$, transmitting through its best direction $d_{\text{UE,opt}} \in \{d_1, \ldots, d_{N_{\text{UE}}}\}$, and gNB$_j$, $j = 1, \ldots, \mathcal{M}$, receiving through its best possible direction $D_{\text{gNB,opt}} \in \{D_1, \ldots, D_{N_{\text{gNB}}}\}$:

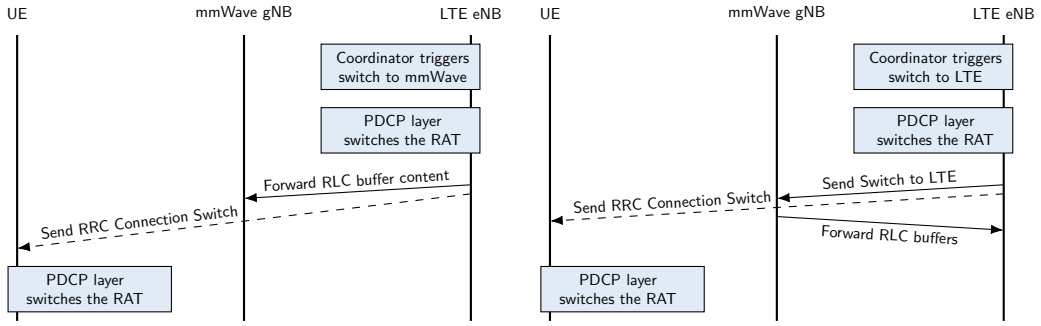
$$\text{SINR}_{i,j} = \max_{\substack{d_{\text{UE}}=d_1,\ldots,d_{N_{\text{UE}}} \\ D_{\text{gNB}}=D_1,\ldots,D_{N_{\text{gNB}}}}} \text{SINR}_{i,j}(d_{\text{UE}}, D_{\text{gNB}}) \tag{4.8}$$

- **Step 2 (Coordination and Network Decision)**: Once the RT of each mmWave gNB has been filled for each UE, each mmWave cell sends this information, through the X2 link, to the coordinator[2] which, in turn, builds a complete report table (CRT), as depicted in Table 4.2 right. When accessing the CRT, the optimal mmWave gNB (with its optimal direction $D_{\text{gNB,opt}}$) is selected for each UE (with optimal direction $d_{\text{UE,opt}}$), considering the absolute maximum SINR in each CRT's row.
- **Step 3 (Reporting)** The coordinator reports to the UE, on a legacy LTE connection, which mmWave gNB yields the best performance, together with the optimal direction $d_{\text{UE,opt}}$ in which the UE should steer its beam, to reach the candidate serving mmWave gNB in the optimal way. The choice of using the LTE control link is motivated by the fact that the UE may not be able to receive from the optimal mmWave link if not properly configured and aligned. Moreover, since handover events in the mmWave regime are commonly due to link failures, the control link to the serving mmWave cell may not be available. Finally, the coordinator also notifies the designated mmWave gNB, through the X2 link, about the optimal direction $D_{\text{gNB,opt}}$ in which to steer the beam, for serving each UE.

4.4.2 Proposed Handover Procedures

The UL-MC scheme described in Sec. 4.4.1 allows to design handover procedures that are faster than the HH solution, thus improving the mobility management in mmWave networks. The HH architecture will be the baseline for the performance evaluation of Sec. 4.4.3: the UE is connected to

[2]The complexity of this framework resides in the central coordinator, which has to aggregate the RT from the $\mathcal{M}$ mmWave gNBs that are under its control and perform for each of the $\mathcal{N}$ UEs a search operations among $\mathcal{M}$ entries. As the number of the mmWave gNBs $\mathcal{M}$ increases, the search space increases linearly.



(a) Switch from LTE RAT to mmWave RAT. **(b)** Switch from mmWave RAT to LTE RAT.

Fig. 4.10: Fast Switching (FS) procedures.

either the LTE or the mmWave RAT and, in order to switch from one to the other, it has to perform a complete handover, or, if the mmWave connectivity is lost, an initial access to LTE from scratch. Besides, in order to perform a handover between mmWave gNBs, the UE has to interact with the MME in the core network, introducing additional delays. The UL ML scheme, instead, allows to perform FS between the LTE and mmWave RATs and SCH across mmWave gNBs.

Fast Switching (FS) – Fig. 4.10 It is used when all the mmWave gNBs for a certain UE are in outage. Since the handling of the state of the user plane for both the mmWave and the LTE RATs is carried out by the LTE RRC, it is possible to correctly modify the state of the PDCP layer and perform a switch from the mmWave to the LTE RAT. The proposed switch procedure simply requires an RRC message (RRC Connection Switch command) to the UE, sent on the LTE link, and a notification to the mmWave gNB via X2 if the switch is from mmWave to LTE, in order to forward the content of the RLC buffers to the LTE eNB. The UL-MC framework therefore allows to have an uninterrupted connection to the LTE anchor point.

Secondary Cell Handover (SCH) – Fig. 4.11. It allows to switch from a secondary mmWave gNB to a different mmWave gNB with a procedure which is faster than a standard intra RAT handover, since it does not involve the interaction with the core network. RA procedure is needed at this stage, even though the UL-MC framework described in Sec. 4.4.1 allows to identify the best beam to be used by the UE in short time and avoids the need for the UE to perform an initial beam search. Moreover, if the UE is capable of maintaining timing control with multiple mmWave gNBs, the RA procedure in the target mmWave gNB can be skipped. We also propose an algorithm for SCH, based on the SINR measurements reported by the mmWave gNBs to the coordinator and on a threshold in time (i.e., the TTT). When a mmWave gNB has a better SINR than the current one (and neither of the two is in outage), the LTE coordinator checks for TTT seconds if the condition still holds, and eventually triggers the SCH. Notice that, if during the TTT the SINR of a third cell becomes better than that of the target cell by less than 3 dB, the handover remains scheduled for the original target gNB, while, if the original cell SINR becomes the highest, then the SCH is canceled. The TTT is computed in two different ways. With the *fixed* TTT option (the benchmark solution) it always has the same value (i.e., $f_{TTT} = 150$ ms), while for the *dynamic* TTT case (our proposed solution) we introduce a dependency on the difference Δ between the SINRs of the best and of the current cell:

$$f_{TTT}(\Delta) = TTT_{max} - \frac{\Delta - \Delta_{min}}{\Delta_{max} - \Delta_{min}}(TTT_{max} - TTT_{min}) \tag{4.9}$$

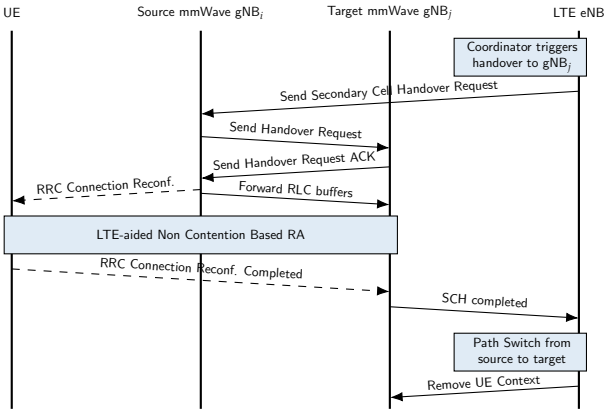


Fig. 4.11: Secondary Cell Handover (SCH) procedure.

so that the actual TTT value is smaller when the difference in SINR between the current gNB and the target is higher. We assume $TTT_{max} = 150$ ms, $TTT_{min} = 25$ ms, $\Delta_{min} = 3$ dB, $\Delta_{max} = 8$ dB. Finally, if at a given time all the mmWave gNBs are in outage, then the UE is instructed to switch to the LTE eNB. If instead only the current mmWave gNB is in outage, the UE immediately performs a handover to the best available mmWave gNB, without waiting for a TTT.

4.4.3 Performance Results

In this subsection, we present some numerical results aiming at comparing the proposed FS and SCH procedures versus the traditional standalone HH architecture for handover management in mmWave networks.

The performance of the proposed handover architecture is assessed through ns-3-based system-level simulations. This approach has the advantage of including many more details than would be allowed by an analytical model (which, for such a complex system, would have to introduce many simplifying assumptions), and makes it possible to evaluate the system performance accounting for realistic (measurement-based) channel behaviors and detailed (standard-like) protocol stack implementations. Existing ns-3 modules, for both LTE [155] and mmWave [156] operations, have been extended to implement the handover procedures described in Sec. 4.4.2, and the source code of the UL-MC framework is publicly available[3].

The reference scenario (depicted in Fig. 4.12) is a urban grid having area 200×115 meters, where 4 non-overlapping buildings of random size and height are deployed, in order to randomize the channel dynamics (in terms on LOS-NLOS transitions) for the moving user. Three mmWave gNBs are located at coordinates $\text{gNB}_2 = (0;50)$, $\text{gNB}_3 = (200;50)$ and $\text{gNB}_4 = (100;110)$, at a height of 10 meters. The LTE eNB_1 is co-located with gNB_4. We consider a single UE that is at coordinates $(50;-5)$ at the beginning of the simulation. It then moves along the x-axis at speed v m/s, until it arrives in position $(150;-5)$. The simulation duration T_{sim} therefore depends on the UE speed v and is given by $T_{\text{sim}} = \frac{l_{\text{path}}}{v} = 20$ s, where $l_{\text{path}} = 100$ m is the length of the path of the UE during the simulation and the default value of the mobile speed has been taken to be $v = 5$ m/s.

Our results are derived through a Monte Carlo approach, where multiple independent simulations are repeated, to get different statistical quantities of interest. In each experiment: (i) we randomly deploy the obstacles; (ii) we apply the UL-MC framework described in Sec. 4.4.1 to collect one

[3] `https://github.com/nyuwireless/ns3-mmwave/tree/new-handover`.

Table 4.3: Simulation parameters for handover performance evaluation.

Parameter	Value	Description
mmWave B_{mmW}	1 GHz	Bandwidth of mmWave gNBs
mmWave f_c	28 GHz	mmWave carrier frequency
mmWave P_{TX}	30 dBm	mmWave transmission power
LTE B_{LTE}	20 MHz	Bandwidth of the LTE eNB
LTE f_c	2.1 GHz	LTE carrier frequency
LTE DL P_{TX}	30 dBm	LTE DL transmission power
LTE UL P_{TX}	25 dBm	LTE UL transmission power
NF	5 dB	Noise figure
Γ_{th}	-5 dB	Minimum SINR threshold
gNB antenna	8×8	gNB UPA MIMO array size
UE antenna	4×4	UE UPA MIMO array size
N_{gNB}	16	gNB scanning directions
N_{UE}	8	UE scanning directions
v	5 m/s	UE speed
B_{RLC}	10 MB	RLC buffer size
D_{X2}	1 ms	One-way delay on X2 links
D_{MME}	10 ms	One-way MME delay
τ_{UDP}	$\{20, 80\}$ μs	UDP packet interarrival time
P_{UDP}	1024 byte	UDP payload size
D	$\{1.6, 12.8, 25.6\}$ ms	CRT intergeneration delay

CRT every D seconds; and (iii) we eventually employ one of the handover algorithms presented in Sec. 4.4.2. The simulations are based on realistic system design parameters which are summarized in Table 4.3. The value of the delay to the MME node (D_{MME}) is chosen in order to model both the propagation delay to a node which is usually centralized and far from the access network, and the processing delays of the MME server. We also model the additional latency D_{X2} introduced by the X2 connections between each pair of gNBs, which has an impact on (i) the forwarding of PDCP PDUs from the LTE eNB to the mmWave ones; (ii) the exchange of control messages for the measurement reporting framework and (iii) the network procedures which require coordination among gNBs. Thus, the latency D_{X2} may delay the detection at the LTE eNB coordinator of an outage with respect to the current mmWave link. In order to avoid performance degradation, the value of D_{X2} should be smaller than 2.5 ms, as recommended by [157]. We consider an SINR threshold $\Gamma_{\text{th}} = -5$ dB, assuming that, if the perceived SINR is below Γ_{th}, no control signals are collected by gNB$_j$ at

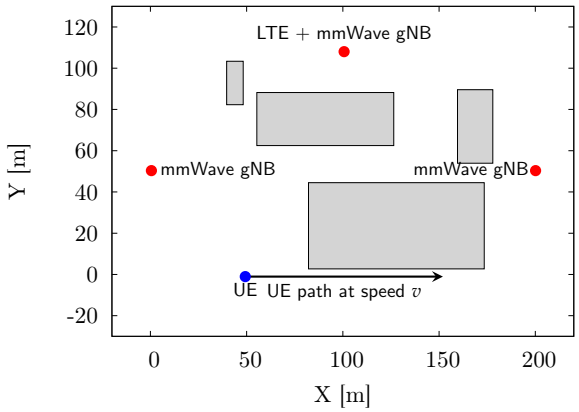
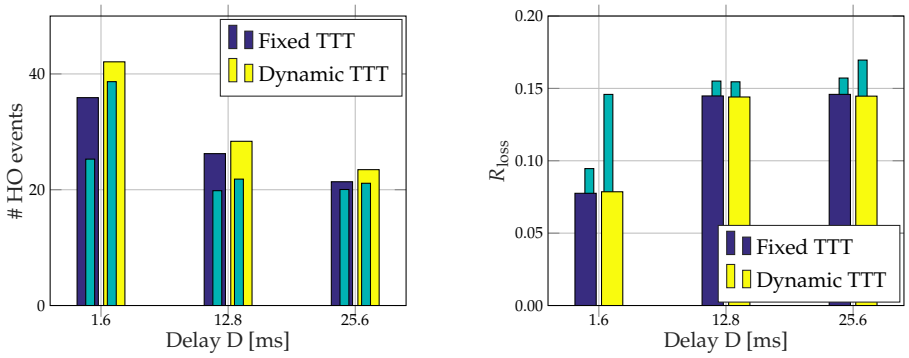


Fig. 4.12: Random realization of the simulation scenario. The grey rectangles are 4 randomly deployed non-overlapping buildings.



(a) Number of handover events during T_{sim}.

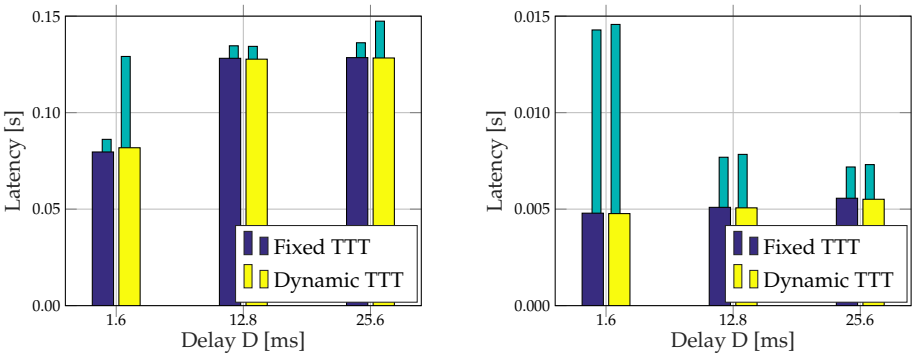
(b) UDP packet loss ratio.

Fig. 4.13: Average number of handover events and packet loss ratio, for different values of the delay D, for a fixed and dynamic TTT handover algorithm. Narrow bars refer to a HH configuration, while wide colored bars refer to a FS/SCH implementation. The RLC buffer size is $B_{RLC} = 10$ MB and the interarrival packet time is $\tau_{udp} = 20\ \mu s$.

time t when the UE is transmitting its SRSs. Reducing Γ_{th} allows the user to be potentially found by more suitable mmWave cells, at the cost of designing more complex (and expensive) receiving schemes, able to detect the intended signal in more noisy channels. gNBs are equipped with a UPA of 8×8 elements, which allow them to steer beams in $N_{gNB} = 16$ directions, whereas UEs have a UPA of 4×4 antennas, steering beams through $N_{UE} = 8$ angular directions. Finally, we assume that CRTs are collected at the LTE eNB every $D = \{1.6, 12.8, 25.6\}$. In general, D can be reduced by configuring hybrid or digital beamforming architectures, thereby allowing UEs and gNBs to look at multiple directions simultaneously in the first step of the proposed UL-MC procedure [110]. The behavior of the User Datagram Protocol (UDP) transport protocol (whose interarrival packet generation time is τ_{UDP}) is tested, to check whether our proposed FS and SCH frameworks for handover management offer good resilience in mobility scenarios. Only downlink traffic is considered.

Packet Loss and Handover In Fig. 4.13a we plot the *average number of handover* (or switch) events for different values of the CRT intergeneration delay D. As expected, we notice that this number is much higher when considering the FS/SCH configuration. The reason is that, since the FS and SCH procedures are faster than the traditional standalone hard handover, the UE has more chances to change its current cell and adapt to the channel dynamics in a more responsive way. Moreover, when increasing the delay D, i.e., when reducing the CRT generation periodicity, the number of handovers reduces, since the UE may have fewer opportunities to update its serving cell, for the same simulation duration. Finally, we see that a dynamic handover procedure requires, on average, a larger number of handover events, to account for the situations in which TTT$<$ 150 ms, when the UE may change its serving cell earlier than it would have done if a fixed TTT algorithm had been applied.

Another element to consider in this performance analysis is the *packet loss ratio* R_{loss}, plotted in Fig. 4.13b, and defined as the ratio between lost and sent packets, averaged over the different iterations for each set of parameters. Since the UDP source constantly injects packets into the system, with interarrival time τ_{UDP}, it can be computed as $R_{loss} = 1 - r\tau_{UDP}/T_{sim}$ where r is the total number of received packets and T_{sim} is the duration of each simulation. We first notice that, with the use of the FS/SCH solution, fewer packets are lost. In fact, there are mainly two elements that contribute to the losses: (i) some UDP packets, which are segmented in the RLC retransmission buffer, cannot



(a) Latency, for $\tau_{\text{UDP}} = 20\ \mu s$.

(b) Latency, for $\tau_{\text{UDP}} = 80\ \mu s$.

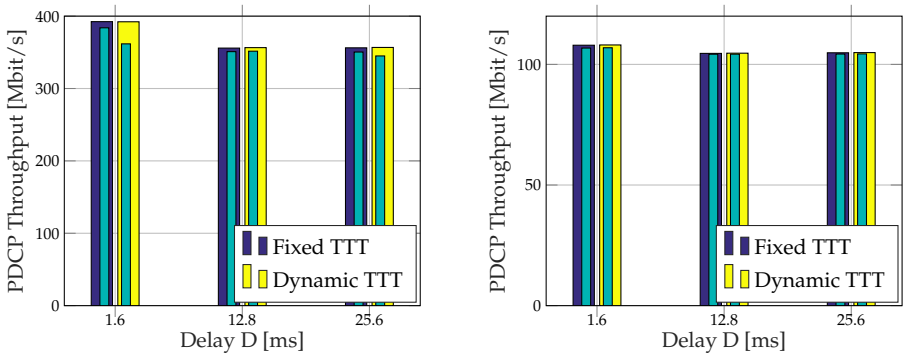
Fig. 4.14: Average latency, for different values of the delay D and the UDP packet interarrival time τ_{UDP}, for a fixed and dynamic TTT handover algorithm. Narrow bars refer to a HH configuration, while wide colored bars refer to a FS/SCH implementation. The RLC buffer size is $B_{\text{RLC}} = 10$ MB.

be reassembled at the PDCP layer and are therefore lost; (ii) during handover, the target gNB RLC transmission buffer receives both the packets sent by the UDP application with interpacket interval τ_{UDP} and the packets that were in the source gNB RLC buffer. If the latter is full, then the target gNB buffer may overflow and discard packets. Both these phenomena are stressed by the fact that the standalone HH procedure takes more time than both the FS and SCH procedures. Moreover, during a complete outage event, with the HH solution, until the UE has completed the Non Contention Based RA procedure with the LTE eNB, packets cannot be sent to the UE and must be buffered at the RLC layer. This worsens the overflow behavior of the RLC buffer. Instead, with FS/SCH, the UE does not need to perform RA since it is already connected and, as soon as packets get to the buffer of the LTE eNB, they are immediately transmitted to the UE. Fig. 4.13b also shows that the packet loss ratio increases when D increases since, if handover or switch events are triggered less frequently, the RLC buffer occupancy increases, and so does the probability of overflow.

Finally, almost no differences are registered when considering a dynamic or fixed TTT handover algorithm, nor when increasing the CRT delay from $D = 12.8$ ms to $D = 25.6$ ms (this aspect will be explained in more detail later).

Latency The latency is measured for each packet, from the time it leaves the PDCP layer of the LTE eNB to when it is successfully received at the PDCP layer of the UE. Therefore, it is the latency of only the correctly received packets, and it accounts also for the forwarding latency D_{X2} on the X2 link. Moreover, this metric captures the queuing time in the RLC buffers, and the additional latency that occurs when a switch or handover happens, before the packet is forwarded to the target gNB or RAT. Fig. 4.14 shows that the FS/SCH frameworks outperform the standalone hard handover: in fact, handovers (which dominate the HH configuration) take more time than the FS and SCH procedures, and therefore the UE experiences a reduced latency and no service interruptions. This result is even more remarkable when realizing that, from Fig. 4.13, the absolute number of handover (or switch) events is higher when considering FS/SCH: despite this consideration, the overall latency is still higher for a system where hard handover is implemented[4]. Furthermore, the latency increases

[4]The latency gap is even more remarkable when considering a dynamic TTT handover algorithm. In fact, although the UE experiences, on average, almost 15% more handovers than in the fixed TTT configuration, the overall latency of the two configurations shown in Fig. 4.14 is comparable, due to the fact that with dynamic TTT some SCHs are more timely.



(a) PDCP throughput in Mbit/s, for $\tau_{\text{UDP}} = 20\ \mu s$. **(b)** PDCP throughput in Mbit/s, for $\tau_{\text{UDP}} = 80\ \mu s$.

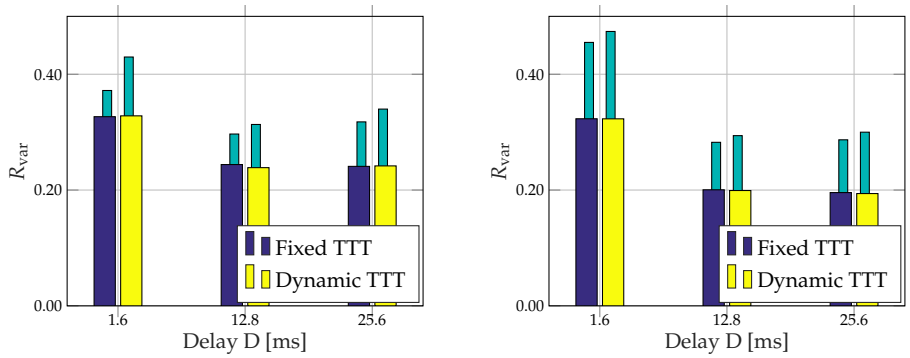
Fig. 4.15: Average PDCP throughput in Mbit/s, for different values of the delay D and the UDP packet interarrival time τ_{UDP}, for fixed and dynamic TTT handover algorithm. Narrow bars refer to a HH configuration, while wide colored bars refer to a FS/SCH implementation. The RLC buffer size is $B_{\text{RLC}} = 10$ MB.

as D increases. In fact, when reducing the intergeneration time of the CRT, the UE is attached to a suboptimal mmWave gNB (or to the LTE eNB) for a longer period of time: this increases the buffer occupancy, thus requiring a stronger effort (and longer time) for forwarding many more packets to the new candidate cell, once the handover (or switch) is triggered. Finally, there are no remarkable differences between $D = 12.8$ and $D = 25.6$ ms.

According to Fig. 4.14b, the latency gap between the HH and FS/SCH configurations is much more impressive when considering $\tau_{\text{UDP}} = 80\ \mu s$. In fact, with this setup, the RLC buffer is empty most of the time and, when a handover (or a switch) is triggered, very few UDP packets need to be forwarded to the destination mmWave or LTE eNB, thus limiting the impact of latency.

We finally recall that the *handover interruption time* (HIT, i.e., the time in which the user's connectivity is interrupted during the handover operations) takes different values, according to the implemented handover scheme (either FS/SCH or HH). When considering a switch to LTE, the HIT is negligible if a FS/SCH approach is used, since the UE is already connected to both the LTE and the mmWave RATs. There may be an additional forwarding latency for the switch from mmWave to LTE, which however is already accounted for in Fig. 4.14. On the other hand, when referring to the baseline HH architecture, the UE has to perform a complete handover to switch from one RAT to the other, thus introducing a significant additional delay. When considering the handover between mmWave gNBs, instead, the HIT is comparable for both the FS/SCH and the HH schemes. However, in the first case, the procedure does not involve any interaction with the core network and the UE is informed about the new mmWave gNB to handover to and the best angular direction to set through an LTE message.

PDCP Throughput The throughput over time at the PDCP layer is measured by sampling the logs of received PDCP PDUs every $T_s = 5$ ms and summing the received packet sizes to obtain the total number of bytes received $B(t)$. Then the throughput $S(t)$ is computed in bit/s as $S(t) = B(t) \times 8/T_s$. In order to get the mean throughput S_{PDCP} for a simulation, these samples are averaged over the total simulation time T_{sim}, and finally over all the simulations, to obtain the parameter $\mathbb{E}[S_{\text{PDCP}}]$. Notice that the PDCP throughput (which is mainly a measure of the rate that the radio access network can offer, given a certain application rate), is mostly made up of the transmission of new incoming packets, but it may also account for the retransmissions of already transmitted ones.



(a) Variance/Mean ratio, for $\tau_{\text{UDP}} = 20\ \mu s$.

(b) Variance/Mean ratio, for $\tau_{\text{UDP}} = 80\ \mu s$.

Fig. 4.16: Average ratio R_{var}, for different values of the delay D and the UDP packet interarrival time τ_{UDP}, for a fixed and dynamic TTT handover algorithm. Narrow bars refer to a HH configuration, while wide colored bars refer to a FS/SCH implementation. The RLC buffer size is $B_{\text{RLC}} = 10$ MB.

In Fig. 4.15, it can be observed that the throughput achievable with the FS/SCH solution is slightly higher than with HH. The reason is that, when relying on the LTE eNB for dealing with outage events, the UE experiences a non-zero throughput, in contrast to the hard handover configuration which cannot properly react to a situation where no mmWave gNBs are within reach. Moreover, the difference in throughput increases as the application rate increases, in accordance with the results on packet loss described in the previous section. As expected, the PDCP throughput decreases as D increases, since the CRT are generated less frequently and the beam pair between the UE and its serving mmWave gNB is monitored less intensively. This means that, when the channel conditions change, the communication quality is not immediately recovered and the throughput is affected by portions of time where suboptimal network settings are chosen.

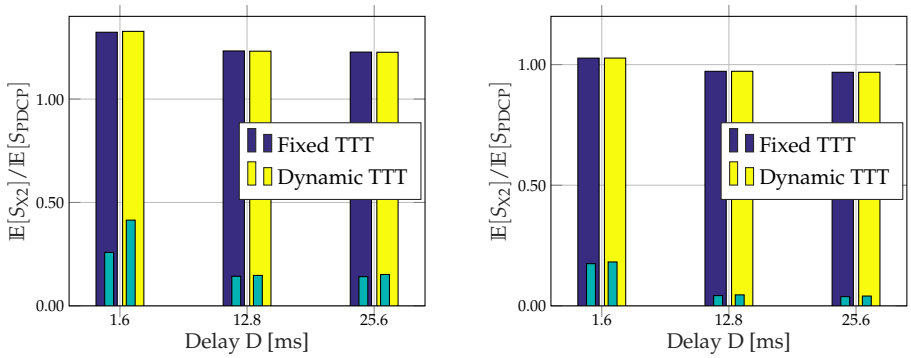
Also a lower UDP rate, according to Fig. 4.15b, presents comparable PDCP throughput gains with respect to the HH option. Finally, it is interesting to notice that, when the system implements a FS/SCH architecture for handover management, the traditional trade-off between latency and throughput no longer holds. In fact, despite the increased number of handover and switch events shown in Fig. 4.13a, with respect to the baseline HH configuration, the UE experiences both a reduced latency and an increased PDCP throughput, thus enhancing the overall network quality of service.

Variance Ratio In order to compare the variance of the rate experienced in time by a user, according to the different handover algorithms implemented, we used the ratio

$$R_{\text{var}} = \frac{\sigma_{S_{\text{PDCP}}}}{\mathbb{E}[S_{\text{PDCP}}]}, \tag{4.10}$$

where $\mathbb{E}[S_{\text{PDCP}}]$ is the mean value of the PDCP throughput measured for each handover configuration and $\sigma_{S_{\text{PDCP}}}$ is its standard deviation, obtained over all repetitions. High values of R_{var} reflect remarkable channel instability, thus the rate would be affected by local variations and periodic degradations.

Let $R_{\text{var,FS/SCH}}$ and $R_{\text{var,HH}}$ be the variance ratios of Eq. (4.10) for the FS/SCH and HH configurations, respectively. From Fig. 4.16, we observe that $R_{\text{var,HH}}$ is higher than $R_{\text{var,FS/SCH}}$, for each value of the delay D, the handover metric and the UDP packet interarrival time τ_{UDP}, making it clear that the LTE eNB employed in a FS/SCH configuration can stabilize the rate, which is not subject to significant variations. In fact, in the portion of time in which the UE would experience zero gain if



(a) $\mathbb{E}[S_{X2}]/\mathbb{E}[S_{PDCP}]$, for $\tau_{UDP} = 20\ \mu s$.

(b) $\mathbb{E}[S_{X2}]/\mathbb{E}[S_{PDCP}]$, for $\tau_{UDP} = 80\ \mu s$.

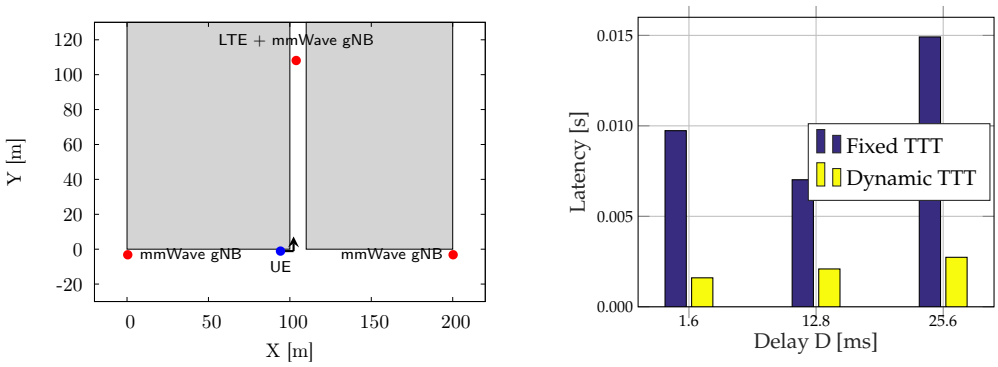
Fig. 4.17: Average ratio of X2 and PDCP throughput, for different values of the delay D and of the UDP packet interarrival time τ_{UDP}, for a fixed and dynamic TTT handover algorithm. Narrow bars refer to a HH configuration, while wide colored bars refer to a FS/SCH implementation. The RLC buffer size is $B_{RLC} = 10$ MB.

a hard handover architecture were implemented (due to an outage event), the rate would suffer a noticeable discrepancy with respect to the LOS values, thus increasing the rate variance throughout the simulation. This is not the case for the FS/SCH configuration, in which the UE can always be supported by the LTE eNB, even when a blockage event affects the scenario. This result is fundamental for real-time applications, which require a long-term stable throughput to support high data rates and a consistently acceptable Quality of Experience (QoE) for the users.

Furthermore, it can be seen that R_{var} increases when the CRT are collected more intensively. In fact, even though reducing D ensures better monitoring of the UE's motion and faster reaction to the channel variations (i.e., LOS/NLOS transitions or periodic modification of the small and large scale fading parameters of the channel), the user is affected by a higher number of handover and switch events, as depicted in Fig. 4.13a: in this way, the serving cell will be adapted regularly during the simulation, thereby causing large and periodic variation of the experienced throughput. For the same reason, R_{var} is higher when applying a dynamic TTT handover algorithm, since the handovers and switches outnumber those of a fixed TTT configuration.

Finally, to compare the FS/SCH and the HH architectures, we can consider the ratio $\frac{R_{var,FS/SCH}}{R_{var,HH}}$. We can therefore affirm that the ratio assumes values lower than 1 for every parameter combination, reflecting the lower variance of a FS/SCH configuration, with respect to the baseline HH option, and (ii) although the dynamic TTT handover approach shows an absolute higher variance than the fixed TTT one, the hard handover baseline suffers much more because of the aggressiveness of the dynamic TTT configuration than the FS/SCH architecture.

X2 Traffic One drawback of the FS/SCH architecture for handover management is that it needs to forward PDCP PDUs from the LTE eNB to the mmWave gNB, besides forwarding the content of RLC buffers during switching and SCH events. On the other hand, the HH option only needs the second kind of forwarding during handovers. Therefore, the load on the X2 links connecting the different gNBs is lower for the HH solution, as can be seen in Fig. 4.17, which shows the ratio between the average $\mathbb{E}[S_{X2}]$ of the sum of the throughput S_{X2} in the six X2 links of the scenario and the average PDCP throughput $\mathbb{E}[S_{PDCP}]$. It can be seen that for the FS/SCH architecture the ratio is close to 1, therefore the X2 links for such configuration must be dimensioned according to the target PDCP throughput for each mmWave gNB. For both architectures the ratio is higher for



(a) Corner case. The grey rectangles are buildings.

(b) Latency.

Fig. 4.18: Average latency, for different values of the delay D and for $\tau_{\text{UDP}} = 20\,\mu s$, comparing a fixed and dynamic TTT handover algorithm. The colored bars refer to a FS/SCH implementation for handover management. The RLC buffer size is $B_{\text{RLC}} = 10$ MB and a *corner scenario* is implemented, for a user moving at speed v.

the lower UDP interarrival time, since there are more packets buffered at the RLC layer that must be forwarded, and also for lower delay D, since there are more handover events. Nevertheless, the forwarding cost (in terms of inbound traffic to the mmWave gNB) of the FS/SCH architecture is similar to that of HH.

Fixed vs. Dynamic TTT In the previous paragraphs, we showed that the proposed dynamic TTT approach never results in a performance degradation for any of the analyzed metrics. Now, we claim that it may also deliver tangible improvements in some specific scenarios where the traditional methods fail, such as the one shown in Fig. 4.18. In this *corner scenario*, the UE turns left at a T-junction and loses LOS with respect to both mmWave gNBs at the bottom. However, the mmWave gNB on top of the scenario is now in LOS, thus the handover should be triggered as quickly as possible. From the result in Fig. 4.18b, we observe that in this case a dynamic and more aggressive approach is able to massively reduce latency compared to the fixed configuration, since a reduced TTT may be vital in this specific scheme, in which the user experiences a degraded rate until the handover to the LOS mmWave gNB is completed. We indeed state that, since the dynamic TTT algorithm never underperforms the fixed TTT approach but is able to greatly improve the performance in specific scenarios, it should be preferred for handover management.

4.5 Facing Radio Link Failures in 5G Networks

One of the key challenges that systems operating in the mmWave bands have to cope with is the rapid channel dynamics. When the quality of an associated control channel falls below a certain threshold, i.e., in the case of Radio Link Failure (RLF) [109, 158], mechanisms to recover acceptable communication capabilities need to be quickly triggered upon notifying the network. Most literature on RLF-related topic refers to challenges that have been recently analyzed in the 60 GHz IEEE 802.11ad WLAN and WPAN scenarios, e.g., in [159–161]. It is however unclear how reactiveness of mobility management schemes for 5G networks is affected by the time it takes to recover from RLFs.

Table 4.4: RLF recovery delay considering the SA or the NSA measurement frameworks, for different values of N_{SS} (the number of SS blocks in a burst), T_{SS} (the time between two consecutive bursts), and for different beamforming configurations. $\Delta_f = 120$ kHz. ABF stands for Analog Beamforming, and DBF for Digital.

Antenna		$T_{\text{RLF,SA}}$ [ms]		
M_{gNB}	M_{UE}	$N_{\text{SS}} = 8$, $T_{\text{SS}} = 20$ *gNB* ABF, UE ABF	$N_{\text{SS}} = 64$, $T_{\text{SS}} = 40$ *gNB* DBF, UE ABF	$N_{\text{SS}} = 64$, $T_{\text{SS}} = 80$ *gNB* DBF, UE ABF
4	4	30.2322	20.3572	40.3572
64	1	130.1072	20.0535	40.0535
64	16	5250	22.6072	42.6072
$T_{\text{RLF,NSA}} \in \{10, 4, 0.8\}$ ms, according to the considerations in [162].				

4.5.1 Solution 1: Use SS/CSI-RS Measurements

Natural candidates for monitoring the link quality and detect the link failure are the SS blocks in a burst introduced in Sec. 3.3.1, and CSI-RSs , introduced in Sec. 4.3.1 [109, 163]. Assume that an object blocks the propagation path of the transceiver at time $T \sim \mathcal{U}[t, t + T_{\text{SS}}]$, i.e., on average at time $\bar{T} = T_{\text{SS}}/2$ within two consecutive SS bursts.

- When implementing an SA architecture, as soon as an impairment is detected, the UE may no longer be able to communicate with its serving gNB since the optimal directional path connecting the endpoints is affected by the failure. The recovery phase is most likely triggered at the beginning of the subsequent SS burst (i.e., on average after $T_{SS} - \bar{T} = T_{\text{SS}}/2$ seconds) and at least after the completion of an IA operation of duration T_{IA} seconds. Table 4.4 reports the RLF recovery delay $T_{\text{RLF,SA}}$ for some network configurations when an SA architecture is implemented. We observe that the latency is quite high for all the investigated settings and is dominated by the IA delay, as illustrated in Fig. 3.7. Moreover, in some circumstances (e.g., $N_{\text{SS}} = 8$, $T_{\text{SS}} = 20$ ms, $M_{gNB} = 64$, $N_{gNB} = 16$ and when analog beamforming is implemented), the RLF recovery delay assumes unacceptably high values.

- Much more responsive RLF recovery operations may be prompted if the failure notification is forwarded through the LTE overlay (i.e., by leveraging the MC feature of an NSA deployment), which may also serve the UE's traffic requests until the mmWave directional communication is successfully restored. If an NSA architecture is designed, the RLF recovery delay $T_{\text{RLF,NSA}}$ is equal to the latency of a traditional LTE connection (which depends on the implemented latency reduction technique, as assessed in [162]). Alternatively, the gNB can autonomously declare an RLF event (without the user's notification) and react accordingly by monitoring the SRS messages. Without loss of generality, assuming that SRSs are uniformly allocated within two SS bursts with periodicity T_{SRS}, an RLF is detected as soon as the gNB is not able to correctly receive N_{SRS} consecutive SRSs from its reference user. In this case, the reactiveness of the RLF recovery operation depends on the periodicity of the sounding signals and is equal to

$$T_{\text{RLF,NSA}} = \frac{T_{\text{SRS}}}{2} + (N_{\text{SRS}} - 1)\, T_{\text{SRS}}. \tag{4.11}$$

  From the results in Table 4.4, it appears that fast and efficient RLF recovery operations can be guaranteed if an NSA solution is preferred over an SA one for all the investigated network configurations.

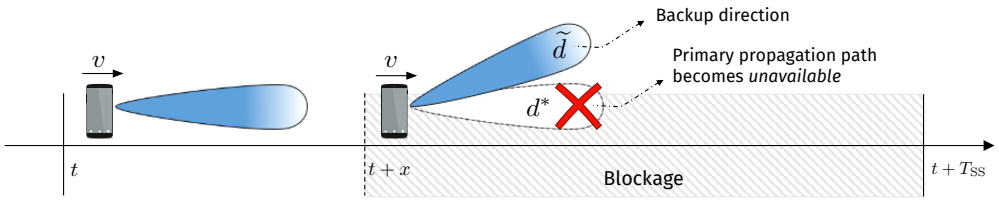


Fig. 4.19: RLF recovery procedure based on the use of backup directions. At time $t+x$ a blockage is detected and the UE, moving at speed v, loses the connection with its current serving mmWave gNB. The UE can promptly react to the channel failure by exploiting its backup direction $\widetilde{d}$.

4.5.2 Solution 2: Use Backup Directions

The UL-MC scheme proposed in Sec. 4.4.1 can be employed to partially overcome the RLF with no delay. We use Fig. 4.19 as a reference. Assume that, at time t, the UE, moving at constant speed v, is connected to mmWave gNB$_{j^*}$ through direction d^*. As soon as a blockage is detected, e.g., at time $t+x$, the UE may lose the connection to its serving cell due to the failure. If no actions are taken, RLF would be recovered at the beginning of the subsequent SS burst (at most at time $t+T_{SS}$) before a new IA operation of duration T_{IA} can be triggered (as described in Sec. 4.5.1). One practical way to promptly react to the path impairment is by configuring the UE to communicate to gNB$_{j^*}$ through its second best direction $\widetilde{d}$ as a sort of *backup solution* before the transceiver fully recovers the optimal beam configuration [123]. Although $\widetilde{d}$ represents a suboptimal solution, having a second available link when the primary link is obstructed adds diversity and robustness to the communication and allows the UE to experience a possibly higher throughput than it would have achieved if no actions were taken. Moreover, it does not require to wait for initial access and/or beam management operations to be completed.

We define $\mathcal{R}^{(d^*)}$ as the optimal rate experienced when no obstacles affect the signal propagation (the UE will communicate through its optimal direction d^*), and $\mathcal{R}^{(\widetilde{d})}$ as the suboptimal rate experienced when the backup beam pair $\widetilde{d}$ is selected, as the primary path is not available. Assume that a blockage event is detected at time $T \sim \mathcal{U}[t, t+T_{SS}]$, and lasts for T_B s. We aim at finding the *rate gain* ($\mathcal{R}_G$), namely the ratio between the rate experienced when a backup beam pair between the UE and its serving gNB is established after a blockage is detected ($\mathcal{R}_{WB}$), and the rate perceived when no actions are taken ($\mathcal{R}_{OB}$). Then, the rate $\mathcal{R}_{WB}$ can be computed (for a fixed time window T_{SS}), as:[5]

$$\mathcal{R}_{WB} = \frac{\mathcal{R}^{(d^*)}T + \mathcal{R}^{(\widetilde{d})}T_B + \mathcal{R}^{d^*}(T_{SS} - T - T_B)}{T_{SS}} = \frac{\mathcal{R}^{(d^*)}(T_{SS} - T_B) + \mathcal{R}^{(\widetilde{d})}T_B}{T_{SS}} \tag{4.12}$$

If no actions are taken, after the obstacle has been detected, the rate $\mathcal{R}_{OB}$ is:

$$\mathcal{R}_{OB} = \frac{\mathcal{R}^{(d^*)}T + 0T_B + \mathcal{R}^{(d^*)}(T_{SS} - T - T_B)}{T_{SS}} = \frac{\mathcal{R}^{(d^*)}(T_{SS} - T_B)}{T_{SS}} \tag{4.13}$$

The average rate gain (R_G) between the two options is finally defined as:

$$\mathcal{R}_G = \frac{\mathcal{R}_{WB}}{R_{OB}} - 1 = \frac{\mathcal{R}^{(\widetilde{d})}}{\mathcal{R}^{(d^*)}} \cdot \frac{T_B}{T_{SS} - T_B} \tag{4.14}$$

In Fig. 4.20, we first notice, as expected, that $\mathcal{R}_G > 0$ for all values of T_{SS} and T_B, making it clear that having a second available link (in case the primary one is blocked) guarantees improved

[5]We consider a time window of duration T_{SS} since, if no actions are taken, RLF recovery will still be achieved at the beginning of the subsequent SS burst, i.e., every $t+T_{SS}$, through IA, as explained in Sec. 4.5.1.

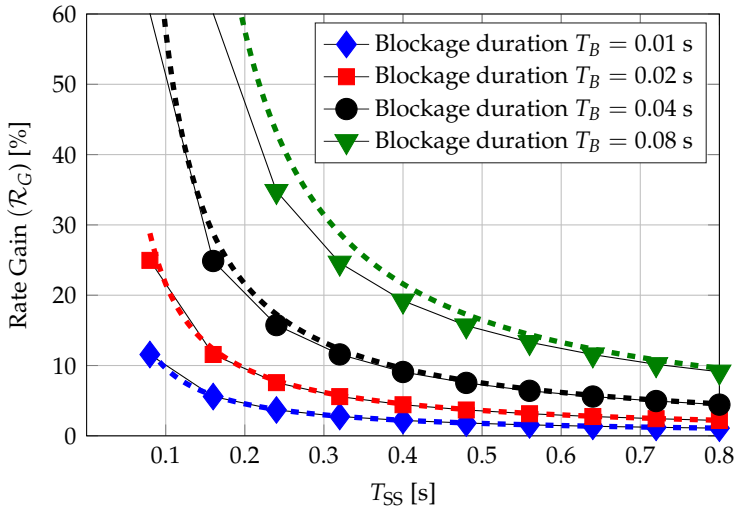


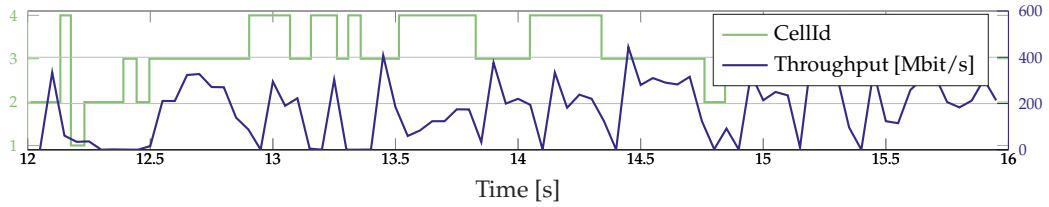
Fig. 4.20: Rate gain experienced when applying a backup procedure for the RLF recovery vs. T_{SS}, for different blockage scenarios. The obstacle duration is T_B.

communication throughput performance with respect to a traditional scheme in which a backup configuration is not available. Furthermore, when T_{SS} is sufficiently large, e.g., when $T_{SS} \gg 2T_B$, the simulation curves asymptotically overlap with the dashed lines plotting Eq. (4.14), thereby validating our analysis. Fig. 4.20 shows also that, for a fixed blockage duration T_B, as T_{SS} increases, the rate gain R_G decreases. In fact the portion of time in which the user would experience zero throughput (if no actions are taken when the primary path is obstructed) proportionally decreases with T_{SS}. This, in turn, results in a larger overhead for beam management, as demonstrated in Sec. 3.4.3.3 and Sec. 4.3.3.2. Finally, we see that, when T_B increases, the rate gain $\mathcal{R}_G$ increases as well, due to the increased enhancement provided by the use of a suboptimal beam pair after a blockage event occurs, with respect to the baseline algorithm in which no actions are taken until a new complete initial access opportunity.

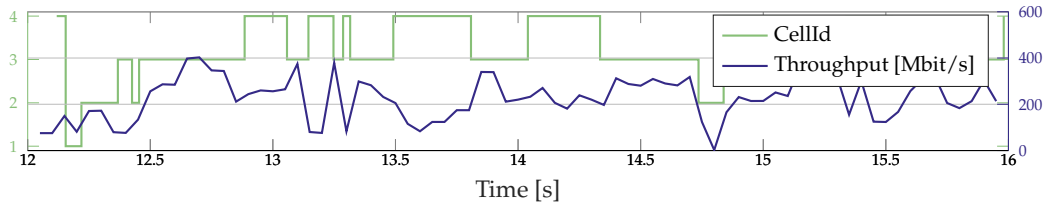
4.6 Conclusions and Design Guidelines

In this chapter, we compared the performance of different mobility management solutions for 5G cellular networks operating at mmWave frequencies. In particular, we simulated beam management, handover, and radio link failure recovery procedures, and we drew the following conclusions.

Beam Management A short T_{CSI} allows an improved reactiveness for the beam management of connected users. In particular, when the number of users per gNB is high then a short CSI-RSs periodicity enables a much shorter T_{tr}. On the other hand, the overhead related to the CSI-RSs is small if compared with that of the SS bursts. The impact of the number of CSI-RSs to be monitored at the UE side $N_{\text{CSI,RX}}$ on the reactiveness is related to both the number of users per gNB and the total number of directions to be swept with the reference signals. If there is a limited number of directions and a large number of users, uniformly distributed in the available directions, then the monitoring of additional CSI-RSs does not impact T_{tr} or the overhead at the network side. The UE may, however, be impacted by the energy consumption related to the monitoring of too many directions, i.e., by a needlessly high $N_{\text{CSI,RX}}$.



(a) Evolution of PDCP throughput for HH.



(b) Evolution of PDCP throughput for FS and SCH procedures.

Fig. 4.21: Evolution, for a specific simulation of duration $T_{\text{sim}} = 20$ seconds, of the PDCP throughput and of the UE's instantaneous mmWave gNB association. We compare both the HH (above) and the FS/SCH (below) configurations, for the fixed TTT handover algorithm and a delay $D = 1.6$ ms. The RLC buffer size is $B_{\text{RLC}} = 10$ MB. The green line represents the current cell over time, where cells from 2 to 4 are mmWave gNBs and cell 1 is the LTE eNB.

Handover We demonstrated that, in general, the proposed FS and SCH architectures for handover perform better than the baseline HH configuration. The main benefit is the short time it takes to change radio access network and its enhancements are shown in terms of mainly: (i) *latency*, which is reduced up to 50% because the FS and SCH procedures are in general much faster than traditional handovers (although the number of handover or switching events may be higher), as observed in Fig. 4.14 and Fig. 4.13a; (ii) *packet loss*, which is reduced since PDUs are less frequently buffered, thus reducing the overflow probability, as shown in Fig 4.13b. This is shown by the lower PDCP throughput of Fig. 4.21a, referred to the HH configuration, with respect to that of the FS/SCH architecture of Fig. 4.21(b); (iii) *throughput variance*, where smaller rate variations are registered, with a reduction of R_{var} of up to 40%, as observed in Fig. 4.16. As an example, Fig. 4.21a shows periodic wide fluctuations of the throughput (which sometimes is even zero, when outages occur), while it settles on steady values when FS/SCH is applied, as in Fig. 4.21b. We also showed that, when the system implements the FS/SCH configuration, despite the increased number of handovers and switches, the UE can *jointly* achieve both a reduced latency and an increased PDCP throughput, enhancing its overall quality of service. We also examined the main cost of the FS/SCH architecture, showing in Fig. 4.17 that the X2 traffic for the FS/SCH option is higher than for the HH configuration because of the forwarding of packets from the LTE eNB to the mmWave gNBs. However, we recall that, with the HH solution, the mmWave gNBs receive the packets from the core network through the S1 link, which is not used for the mmWave gNBs in the FS/SCH configuration. When considering the overall inbound traffic to the mmWave gNBs on both the X2 and the S1 links, the costs of the two architectures may be equivalent. Given these considerations, we argue that the use of multi-connectivity for mobility management is to be preferred to the traditional hard handover approach.

We also observed that the general behaviors of different UDP interarrival times are similar for most metrics. However, the latency is much lower when $\tau_{\text{UDP}} = 80\,\mu s$, i.e., when considering less congested scenarios, since RLC buffers are empty most of the time and fewer packets need to be forwarded during handover. This justifies the wider gap between FS/SCH and HH architectures,

with respect to the $\tau_{\text{UDP}} = 20\,\mu$s case.

Regarding the value of the CRT intergeneration delay D, we noticed remarkable differences between $D = 1.6$ and $D = 25.6$ ms (validating the choice of designing a *digital* BF architecture, more complex but more efficient in terms of both latency and throughput) but almost no distinction between the $D = 12.8$ and $D = 25.6$ ms configurations: we conclude that a *hybrid* BF system at the mmWave gNB side is not to be preferred to an *analog* one, since the complexity is increased while the overall performance is almost equivalent.

Radio Link Failure Recovery We proved that an NSA framework may be preferable to implement, for UEs in connected mode, efficient and reactive recovery operations, thanks to the fact that the failure notification can be forwarded through the LTE overlay as soon as an RLF event is detected, with no need to wait for a new initial access procedure to be completed. Moreover, we made the case that, in some circumstances, the UE can autonomously react to an RLF by selecting an alternative direction of communication, as a sort of backup solution before the transceiver fully recovers the optimal beam configuration.

Chapter 5

Integrated Access and Backhaul 5G Millimeter Wave Cellular Networks

5.1 Introduction

As described in Chapter 1, the 3GPP has recently completed, as part of its Release 15, the standardization of 3GPP NR that supports, in addition to a flexible frame structure and a revised core network design, transmissions at mmWave frequencies (up to 52.6 GHz) [9]. The large available spectrum at mmWaves, with much wider bandwidths than in previous network generations, offers the potential of orders of magnitude higher transmission speeds than when operating in the congested bands below 6 GHz, even though communication suffers from severe path and penetration losses [10].

One promising approach to overcome such limitations is using high gain antennas to help close the link, thus introducing directionality in the communication, with electronic beamforming to support mobile users (see Sec. 2.2). Network densification is also used to improve the performance by reducing inter-site distance to establish stronger access channels. An ultra-dense deployment, however, involves high capital and operational expenditures (capex and opex) for network operators [164], because high capacity backhaul connections have to be provided to a larger number of cellular base stations than in networks operating at lower frequencies.

Network *disaggregation* (i.e., the separation of the layers of the protocol stack into different physical equipments) [165] and *virtualization* (i.e., the usage of software- and not hardware-based protocol stack implementations) [166] can lower capex and opex by reducing the complexity of individual base stations. Some researchers have also started investigating the feasibility of *Integrated Access and Backhaul (IAB)*, in which only a fraction of gNBs connect to traditional fiber-like infrastructures, while the others wirelessly relay the backhaul traffic, possibly through multiple hops and at mmWave frequencies [167]. The importance of the IAB framework as a cost-effective alternative to the wired backhaul has been recognized by the 3GPP, which recently completed a Study Item (SI) for NR Release 16 [168] to investigate architectures, radio protocols, and physical layer aspects for sharing radio resources between access and backhaul links. Although the 3GPP LTE and LTE-Advanced standards already provide specifications for base stations with wireless backhauling capabilities, the SI on IAB foresees a more advanced and flexible solution, which includes the support of multi-hop

This chapter is based on the contributions presented in **[J7, C7, C8]**. Part of the results included in this chapter is also based on joint work with Michele Polese.

Table 5.1: Literature on IAB strategies for mmWave networks.

Topic	Relevant References
IAB [168, 171]	[167, 172–175] Below-6 GHz solutions. [170, 176–181] Above-6 GHz performance evaluation. [169, 182–184] Advanced solutions for IAB deployments.

communications, dynamic multiplexing of the resources, and a plug-and-play design to reduce the deployment complexity.

5.1.1 Motivations and Chapter Structure

Despite the consensus about IAB's ability to reduce costs, designing an efficient and high-performance IAB network is still an open challenge. Along these lines, in this chapter, we evaluate the feasibility of realizing an IAB architecture in realistic mmWave deployments. In particular, in Sec. 5.3 we shed light on the most recent 3GPP standardization activities on IAB.

In Sec. 5.4 we investigate how to efficiently forward the backhaul traffic from the wireless IAB-nodes to the core network [169]. In particular, we compare how different greedy policies perform with respect to the number of hops and the bottleneck SNR, i.e., the SNR of the weakest wireless backhaul link, relying only on local information, without the need for a centralized coordinator. We also discuss the use of a function that biases the link selection towards base stations with wired backhaul to the core network, and show that, for a certain set of parameters for this bias, it is possible to decrease the number of hops without affecting the average bottleneck SNR.

In Sec. 5.5 we evaluate the performance of the IAB architecture in an end-to-end environment in terms of experienced throughput and latency, considering realistic traffic models. More specifically, we compare network scenarios in which a percentage of gNBs (i.e., the IAB-nodes) use wireless backhaul connections to a few gNBs (i.e., the IAB-donors) with a wired connection to the core network against two baseline solutions, i.e., a network with only the IAB-donors, and one in which all the gNBs have a wired connection. To do so, we conduct system-level simulations using ns-3, an open-source network simulator which has recently been extended to feature a detailed 3GPP-like protocol stack implementation of IAB at mmWaves [170]. We consider both traditional UDP services and more realistic applications including HyperText Transfer Protocol (HTTP) for web browsing and Dynamic Adaptive Streaming over HTTP (DASH) for high-quality video streaming. Our results demonstrate that, while wired backhaul implementations deliver improved overall throughput in conditions of highly saturated traffic, the IAB configuration promotes fairness for the worst users by associating to relay nodes (IAB-nodes) the UEs which otherwise would have a poor connection to the wired donor.

Finally, in Sec. 5.6 we identify potential and challenges on IAB-related topics.

5.2 Related Work

Research on wireless backhaul solutions has been carried out in the past at frequencies below 6 GHz, e.g., in the WLAN domain [172] and as part of the LTE standardization activity with a single wireless backhaul hop [173]. The practical implementation of wireless multi-hop networks, however, never really turned into a commercial deployment due to practical limitations including, but not limited to, (i) scalability issues [167], (ii) the scheduling constraints between hops [174], and (iii) the large overhead for maintaining multi-hop routes [175].

Nonetheless, with the recent advancements in mmWave communication and leveraging highly directional beamforming, the integration of wireless backhaul and radio access is being considered as a promising solution for 5G cellular networks. In [176], the authors demonstrated that the noise-limited nature of large-bandwidth mmWave networks offer interference isolation, thereby providing an opportunity to incorporate self-backhauling in a mesh small-cell deployment without significant throughput degradation. Ge *et al.* [177] showed that wireless backhaul over mmWave links can meet the expected increase in mobile traffic demands, while Mesodiakaki *et al.* [178] evaluated the energy efficiency of mmWave backhaul at different frequencies and demonstrated that 140 GHz represents a very promising spectrum solution mainly due to the possibility of highly directive beams as well as its high bandwidth availability. The authors in [179] further evaluated the performance of the integration between access and backhaul and determined the maximum total network load that can be supported using the IAB architecture. Ometov *et al.* [180] assessed the aggregated packet loss probability of backhauling technologies for 3GPP NR systems as a function environmental conditions, mmWave system specifics, and generated traffic volume. They showed that the autocorrelation in aggregated traffic provides a significant impact on service characteristics of mmWave backhaul, that can be partially compensated by increasing either emitted power or the number of antenna array elements. In any case, performance analyses for IAB topics are typically focused on PHY or MAC layer protocols, whereby the impact of upper layers is neglected. Conversely, Tian *et al.* [181] recently evaluated the system performance of IAB node deployment in terms of coverage and user throughput by field experiment using mmWave bands, and proved that coverage ratio can be significant improved by through wireless backhaul, in particular when dynamic resource allocation is implemented. In our previous contribution [170, 171], we also assessed the end-to-end performance of IAB networks through ns-3 simulations, and provided guidelines on how to design optimal wireless backhaul solutions in the presence of resource-constrained and traffic-congested mmWave scenarios.

More sophisticated IAB architectures have also been discussed in the literature. The authors in [182], for example, considered out-of-band backhaul for mmWave networks, i.e., the access and backhaul tiers do use the same frequency band, and implemented a scheduling algorithm to increase throughput performance in certain scenarios. In [183] a novel joint resource allocation design for IAB-based user provided networks was modeled as a Nash bargaining problem, and simulations showed that the proposed implementation can effectively improve both user experience and network throughput. IAB deployment performance can be also improved by leveraging the flying capabilities of Unmanned Aerial Vehicles (UAVs) as hovering IAB-nodes, as proposed in [184].

5.3 Integrated Access and Backhaul in 3GPP NR

The 3GPP recently finalized the SI on IAB [168], whose main objective was to assess the feasibility of integrated access and wireless backhaul over NR (i.e., the 5G radio interface), and to propose potential solutions to ensure efficient backhauling operations. This SI led to a work item, and is expected to be integrated in future releases of the 3GPP specifications.

The SI considered fixed wireless relays with both in-band (i.e., the access and the backhaul traffic are multiplexed over the same frequency band) and out-of-band backhauling capabilities (i.e., the access and the backhaul traffic use separate frequency bands), with a focus on the former, which is more challenging in terms of network design and management but maximizes the spectrum utilization. According to [168], IAB operations are spectrum agnostic, thus the relays can be deployed either in the above-6 GHz or sub-6 GHz spectrum, and can operate both in SA (connected to the 5G core network) or NSA modes (connected to the 4G EPC). The possible topologies for an IAB network are (i) a Spanning Tree (ST), in which each IAB-node is connected to a single parent, or (ii) a Directed

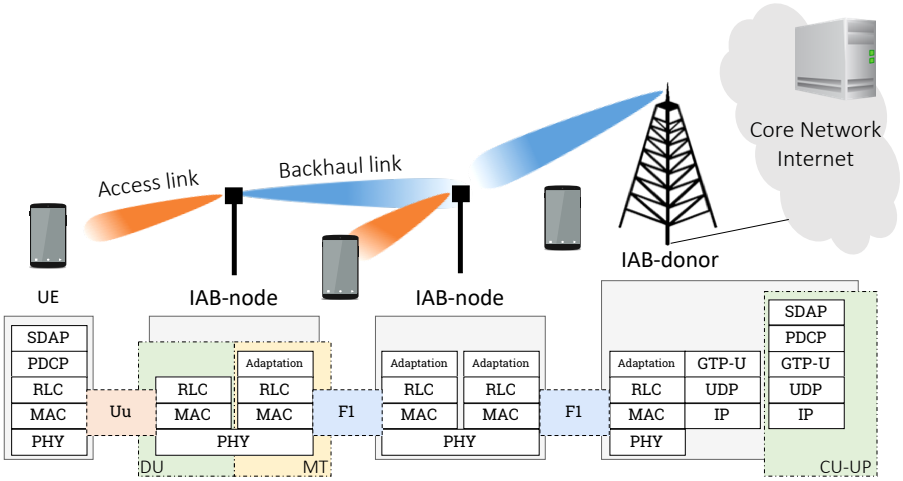


Fig. 5.1: Protocol stack and basic architecture of an IAB network. The Uu interface represents the interface between the UE and the DU in the IAB-node, while the F1* interface is used between the IAB DU and the upstream CU.

Acyclic Graph (DAG), in which each IAB-node may be connected to multiple upstream nodes.

In the following, we will review the main innovations introduced in [168] for the network architecture, procedures for network management, and resource multiplexing through scheduling.

5.3.1 Architecture

As shown in Fig. 5.1, the logical architecture of an IAB network is composed of multiple IAB-nodes, having wireless backhauling capabilities and can serve UEs as well as other IAB-nodes, and IAB-donors, having fiber connectivity towards the core network and can serve UEs and IAB-nodes.

The SI initially proposed five different configuration options for the architecture, with different levels of decentralization of the network functionalities and different solutions to enable backhauling. The final version, selected for future standardization, was preferred because it had limited impact on the core network specifications, had lower relay complexity and processing requirements, and had more limited signaling overhead.

According to the chosen architecture, each IAB-node hosts two NR functions: (i) a Mobile Termination (MT), used to maintain the wireless backhaul connection towards an upstream IAB-node or IAB-donor, and (ii) a Distributed Unit (DU), to provide access connection to the UEs or the downstream MTs of other IAB-nodes. The DU connects to a CU hosted by the IAB-donor by means of the NR F1* interface running over the wireless backhaul link. Therefore, in the access of IAB-nodes and donors there is a coexistence of two interfaces, i.e., the Uu interface (between the UEs and the DU of the gNBs) and the aforementioned F1* interface. With this choice it is possible to exploit the functional split of the radio protocol stack: the CU at the IAB-donor holds all the control and upper layer functionalities, while the lower layer operations are delegated to the DUs located at the IAB-nodes. The split happens at the RLC layer, therefore RRC, Service Data Adaptation Protocol (SDAP) and PDCP layers reside in the CU, while RLC, MAC and PHY are hosted by the DUs. An additional adaptation layer is added on top of RLC, which routes the data across the IAB network topology, hence enabling the end-to-end connection between DUs and the CU.

5.3.2 Network Procedures and Topology Management

An important element to be considered in an IAB deployment is the establishment and management of the network topology. This is because the end-to-end performance of the overall network strongly depends on the number of hops between the donor and the end relay, on how many relays the donor needs to support, and strategies adopted for procedures such as network formation, route selection and resource allocation. To ensure efficient IAB operations, it is necessary to optimize the performance of various network procedures involving topology and resource management.

The topology establishment is performed during the IAB-node setup, and is a critical step. When an IAB-node becomes active, it first selects the upstream node to attach to. To accomplish this, the MT performs the same initial access procedure as a UE, i.e., it makes use of the synchronization signals transmitted by the available cells (formally called SS blocks in NR, as described in Sec. 3.3.1) to estimate the channel and select the parent. Moreover, although not currently supported by the specifications, we argue that it would be beneficial if the MT could retrieve additional information (e.g., the number of hops to reach the donor, the cell load, etc.), and then select the cell to attach to, based on more advanced path selection metrics [53] than just the Received Signal Strength (RSS), as will be discussed in Sec. 5.4. Then, the IAB-node configures its DU, establishes the F1* connection towards the CU in the remote IAB-donor, and is ready to provide services to UEs and other IAB-nodes. During this initial phase, the IAB-node may transmit information to the IAB-donor about its topological location within the IAB network.

The topology management function then dynamically adapts the IAB topology in order to maintain service continuity (e.g., when a backhaul link is lost), or for load balancing purposes. In addition to the information provided during the initial setup procedure, the IAB-nodes may also transmit periodic information about traffic load and backhaul link quality. This allows the CU to be aware of the overall IAB topology, find the optimal configuration, and adapt it by changing network connectivity (i.e., the associations between the IAB-nodes) accordingly. In case the IAB-nodes support a DAG topology with MC towards multiple upstream nodes, it is also possible to provide greater redundancy and load balancing. In this case, the addition/removal of redundant routes is managed by the CU based on the propagation conditions and traffic load of each wireless backhaul link.

5.3.3 Scheduling and Resource Multiplexing

For in-band IAB operations, the need to multiplex both the access and the backhaul traffic within the same frequency band forces half-duplex operations. This constraint has been acknowledged in the 3GPP SI report [168], although full-duplex solutions are not excluded. Therefore, the radio resources must be orthogonally partitioned between the access and the backhaul, either in time (Time Division Multiplexing (TDM), which is the preferred solution in [168]), frequency (Frequency Division Multiplexing (FDM)), or space (Space Division Multiplexing (SDM)), using a centralized or decentralized scheduling coordination mechanism across the IAB-nodes and the IAB-donor.

Despite the limitations imposed by the half-duplex constraint, the IAB network is required to address the access traffic requirements of all the users. For this reason, the available resources should be allocated fairly, taking into account channel measurements and topology-related information possibly exchanged between the IAB-nodes. Furthermore, both hop-by-hop and end-to-end flow control mechanisms should be provided to mitigate the risk of congestion on intermediate hops, which might arise in case of poor propagation conditions.

Table 5.2: Comparison of different link selection policies for IAB.

Policy	Metric	Link selection rule	Pros & Cons
HQF	SNR	Select the link with max SNR	High bottleneck SNR High outage probability
WF	SNR	Select the wired gNB, if available, otherwise apply HQF	Low number of hops Low bottleneck SNR
PA	SNR	Select the link with max SNR (if parents are close to a wired gNB)	Low number of hops Possible ping-pong effects
MLR	Rate	Select the link with the max rate	High bottleneck rate High outage probability

5.4 Path Selection Policies for IAB

In this section, we present some distributed link selection policies that we proposed in [169] to efficiently forward the backhaul traffic (possibly through multiple hops) from a wireless gNB to a wired gNB connected to the core network. In the following paragraphs, we will use the term (i) *wired* gNB or *donor* to identify gNBs which are connected to the core network with a wired backhaul; (ii) IAB node or *relay* to label gNBs which do not have a wired backhaul link; and (iii) *parent* gNB to name a gNB which provides a wireless backhaul link to an IAB node. The parent can be itself a wireless IAB node, or a wired gNB. For all of the policies, the IAB node that has to find the path towards the core network initiates the procedure by applying the selection policy on the first hop, and then the procedure continues iteratively at each hop until a suitable wired gNB is reached. Therefore, the strategies we evaluate are greedy, i.e., consider local information to perform the hop-by-hop link selection decisions, and do not need a centralized controller. These policies can be used to re-route backhaul traffic on the fly, in case of a link failure, and to connect (possibly via multiple hops) an IAB node which is joining the network for the first time to a suitable wired gNB in an autonomous and non-coordinated fashion.

5.4.1 Description of Path Selection Policies

The considered policies differ from one another because of the metric used to measure the link quality (SNR or rate), and because of the ranking criterion of the different available links at each hop. For every policy, and at each hop, we consider an SNR threshold Γ_{th}, i.e., for the link selection, we compare only backhaul connections with an SNR Γ higher than or equal to this threshold. If Γ_{th} is small, then it is possible to select and compare a larger number of base stations as parent candidates, and possibly increase the probability of successfully reaching a wired gNB, at the price of a lower data rate on the bottleneck link. For the access network, Γ_{th} is usually set to -5 dB [53], i.e., access links with an SNR smaller than -5 dB are usually considered in outage. However, this choice is not valid in a backhaul context, where the link is required to reliably forward high-data-rate traffic from the relay to its parent gNB. Therefore, we select a higher value for Γ_{th}, i.e., 5 dB, which corresponds to a theoretically achievable Shannon rate of 830 Mbit/s, on a single carrier with a bandwidth $B = 400$ MHz [11]. Moreover, we avoid loops, i.e., if an IAB node was used as a relay in a previous hop, it cannot be selected again. Table 5.2 sums up the main properties of each policy, which are described in detail in the following sections.

5.4.1.1 Highest-Quality-First (HQF) policy

At each hop, the HQF strategy compares the SNR Γ of the available links towards each possible parent gNBs (either wired or wireless), and selects that with the highest SNR, without considering any additional information. It is a very simple selection rule, which can be implemented only by measuring the link quality using synchronization signals. Moreover, by always selecting the best SNR, the bottleneck link, i.e., the link with the lowest SNR among the hops towards the wired gNB, will have a high SNR when compared to other policies. On the other hand, given that this policy follows a greedy approach, it may happen that the parent gNB with the best SNR leads further away from a wired gNB, thus increasing the number of hops. Moreover, in some cases, the highest SNR leads to the choice of another relay gNB which however is not within reach of any other possible wireless parent or wired donor, thus failing to connect to a wired gNB.

5.4.1.2 Wired-first (WF) policy

The WF policy is designed to reduce as much as possible the number of hops needed to reach a wired gNB. Indeed, if at a given hop one of the available backhaul links is toward a wired gNB, i.e., if a wired gNB is reachable from the current IAB node with an SNR higher than the threshold Γ_{th}, then the wired gNB is selected even if it is not associated to the connection with the highest SNR. If instead no wired gNB is available, then the HQF policy is applied. The IAB node would need to know which candidate parents are wired or wireless, and this can be done by extending the information directionally broadcast (using SS blocks [109]) by each gNB in the MIB or SIB. While this policy increases the probability of reaching a wired gNB, even with a greedy approach, it may cause a degradation in the quality of the bottleneck link.

5.4.1.3 Position-Aware (PA) policy

This strategy uses additional context information related to the position of the IAB node that has to perform the link selection and the wired gNB in the scenario. This information can be available in advance and pre-configured in the relays (especially if non-mobile relays are considered [185]), or shared on directional broadcast messages. The goal is to avoid selecting a parent gNB that is more distant from the closest wired gNB than the current IAB node. Therefore, the IAB node divides the neighboring region into two half-planes, identified by a line which (i) passes through the position of the IAB node and (ii) is perpendicular to the line that passes through the positions of the IAB node and the closest wired gNB. Then, it considers for its selection only the candidate parents which are in the half-plane containing the wired gNB, and selects that with the highest SNR. This policy should strike a balance between HQF and WF.

5.4.1.4 Maximum-Local-Rate (MLR) policy

The MLR policy does not consider the SNR as a metric, but at each hop selects the candidate parent with the highest achievable Shannon rate. Consider IAB node i, and the candidate parent j, and let $\mathcal{N}_j$ be the number of users and IAB nodes currently attached to j. Then, given a bandwidth B and the SNR $\Gamma_{i,j}$ between the IAB node and the candidate parent, the Shannon rate is computed as $\mathcal{R}_j = B/\mathcal{N}_j \log_2(1 + \Gamma_{i,j})$. Finally, the IAB node selects the parent with the highest achievable rate $\mathcal{R}$. Once again, we assume that the information on the load (in terms of number of users $\mathcal{N}_j$) of candidate parent j is known to the IAB node, for example through extension of the MIB or SIB, or with a passive estimation of the power ratio between the resources allocated to synchronization

signals and to data transmissions. This strategy is designed to take into account the load information in the decision, but has the same drawbacks of the HQF policy, i.e., it may yield a high number of hops and/or connection failures.

5.4.2 Wired Bias Function

For multi-hop scenarios, one of the Key Performance Indices (KPIs) considered in the 3GPP SI for IAB is the number of hops from a certain wireless IAB node to the first wired gNB it can reach. However, as discussed in the previous section, some of the proposed policies may need a high number of hops, or even never reach the target wired gNB. In order to solve this issue, it is possible to apply a *Wired Bias Function (WBF)* to the SNR of the wired gNBs during the evaluation of the metric for the link selection. Consequently, a wired gNB may be chosen as parent even though it is not the candidate with the highest considered metric. The bias is not fixed, but is a function $W(N_h)$ of the number of hops N_h traveled from the IAB node that is trying to connect to a wired gNB. The idea is that as N_h increases, it becomes more and more convenient to select as a parent a wired gNB with respect to another wireless IAB node (that would otherwise add up to the number of hops) even though the wired gNB is not the best according to the metric considered. The WF policy is a particular case of a decision with bias, with $W(N_h)$ large enough so that the wired gNB is always selected if above Γ_{th}.

We compare two different WBFs, which are respectively polynomial and exponential in the number of hops N_h. The first is defined as follows:

$$W_p(N_h) = \left(\frac{N_h}{N_{h,t}}\right)^k \Gamma_{gap} + \Gamma_H, \tag{5.1}$$

where k is the degree of the polynomial, $N_{h,t}$ is a threshold on the number of hops, Γ_{gap} a tolerable SNR gap, and Γ_H an SNR hysteresis. The idea is that, if N_h is smaller than $N_{h,t}$, then the SNR gap parameter Γ_{gap} is multiplied by a number smaller than 1, and the WBF $W(N_h)$ does not impact too much the link choice. When the number of hops N_h reaches the threshold $N_{h,t}$, then $W(N_h)$ takes values which are greater than or equal to Γ_{gap}, increasing the weight of the bias in the link selection. The SNR hysteresis Γ_H is set to 2 dB, and slightly offsets the choice towards a wired gNB in case the best wireless relay candidate and the wired gNB have a very similar SNR. Very conservative WBF would use a large $N_{h,t}$, and small k and Γ_{gap}, and vice versa for an aggressive parameter tuning.

Similarly, the exponential WBF is defined as

$$W_e(N_h) = \gamma^{\left(\frac{N_h}{N_{h,t}}\right)} \Gamma_{gap} + \Gamma_H. \tag{5.2}$$

Notice that γ must be greater than or equal to 1, otherwise $\gamma^{\left(\frac{N_h}{N_{h,t}}\right)}$ would decrease with the number of hops. Moreover, for any γ, the exponential WBF $W_e(N_h)$ is larger than the polynomial $W_p(N_h)$, for the same choice of the other parameters. For example, if $N_{h,t} = 6$ and $N_h = 1$, with $\gamma = 1.5$ we have $\gamma^{\left(\frac{N_h}{N_{h,t}}\right)} = 1.07$, while with $k = 1$ we have $\left(\frac{N_h}{N_{h,t}}\right)^k = 0.17$.

5.4.3 Performance Evaluation

The performance evaluation for the path selection policies presented in Sec. 5.4.1 is done via Monte Carlo simulations with 20000 independent repetitions for each configuration. In particular, we compare the CDFs of (i) the number of hops required to forward the backhaul traffic from a wireless to a wired gNB, and (ii) the bottleneck SNR, i.e., the SNR of the weakest link. The main parameters for the simulations are reported in Table 5.3.

Table 5.3: Simulation parameters for IAB path selection policies.

Parameter	Value	Description
B	400 MHz	Bandwidth of mmWave gNBs
f_c	28 GHz	mmWave carrier frequency
P_{TX}	30 dBm	mmWave transmission power
NF	5 dB	Noise figure
M_{gNB}	$\{8 \times 8, 16 \times 16\}$	gNB UPA MIMO array size
S	3	Number of sectors for each gNB
λ_b	$\{30, 60\}$ gNB/km^2	gNB density
p_w	$\{0.1, 0.3\}$	Fraction of wired gNB

Table 5.4: Wired Bias Function parameters for IAB.

Configuration	Parameters
Aggressive $W_e(N_h)$	$N_{h,t} = 1$, $\gamma = 3$, $\Gamma_{gap} = 15$ dB, $\Gamma_H = 2$ dB
Conservative $W_e(N_h)$	$N_{h,t} = 6$, $\gamma = 1.5$, $\Gamma_{gap} = 5$ dB, $\Gamma_H = 2$ dB
Aggressive $W_p(N_h)$	$N_{h,t} = 1$, $k = 3$, $\Gamma_{gap} = 15$ dB, $\Gamma_H = 2$ dB
Conservative $W_p(N_h)$	$N_{h,t} = 6$, $k = 1$, $\Gamma_{gap} = 5$ dB, $\Gamma_H = 2$ dB

The gNBs (both wired and wireless) are deployed according to a PPP with density $\lambda_b \in \{30, 60\}$ gNB/km^2, and a fraction $p_w \in \{0.1, 0.3\}$ is configured with a wired backhaul link to the core network. Therefore, the density of the wired gNBs is $\lambda_{w,b} = p_w \lambda_b$ gNB/km^2, while the IAB nodes have a density $\lambda_{i,b} = (1 - p_w)\lambda_b$ gNB/km^2. For the evaluation of the MLR policy, we also deploy UEs according to a PPP with density of λ_u UE/km^2, and associate them to the gNB with the smallest pathloss, in line with previous studies [179]. We assume that the IAB nodes are equipped with S uniform planar antenna arrays, with the same number $M_{gNB} \in \{64, 256\}$ of isotropic antenna elements at both endpoints of the connection. Each antenna array covers a sector of $2\pi/S$ degrees. Moreover, node i can monitor the link quality of gNB $j \in \mathcal{E}_i$, where $\mathcal{E}_i$ is the set of wired or wireless gNBs whose reference signals can be received by node i. The IAB node can then select the best beam to communicate with j using the standard beam management procedures of 3GPP NR. Table 5.4 finally summarizes the main parameters used for the WBF. In particular, we identify a conservative policy, with $N_{h,t} = 6$, $\Gamma_{gap} = 5$ dB and $k = 1$ or $\gamma = 1.5$ for the polynomial and the exponential policies, respectively, and an aggressive one, with $N_{h,t} = 1$, $\Gamma_{gap} = 15$ dB and $k = 3$ or $\gamma = 3$.

Antenna and deployment configurations In Fig. 5.2 we investigate how the relaying performance evolves as a function of different setup configurations, i.e., the number of antenna elements M_{gNB} each gNB is equipped with and the gNB density λ_b. The WF strategy is considered. Increasing the MIMO array size has beneficial effects on both the number of hops and the bottleneck SNR. In the first case, the narrower beams that can be steered and the resulting higher gains that are produced by beamforming enlarge the discoverable area of each gNB, thereby increasing the probability of detecting a wired gNB with sufficiently good signal quality and through a limited number of hops. In the second case, sharper beams guarantee better signal quality and thus stronger received power.

Similarly, enhanced backhauling performance is achieved by densifying the network since the gNBs are gradually closer and thus establish more precise alignment and, in general, connections with a higher link budget. Of course, increasing λ_b beyond a point has a negative impact on the performance due to higher interference from the surrounding base stations.

Finally, notice that the $M_{gNB} = 64$, $\lambda_b = 60$ gNB/km^2 and the $M_{gNB} = 256$, $\lambda_b = 30$ gNB/km^2

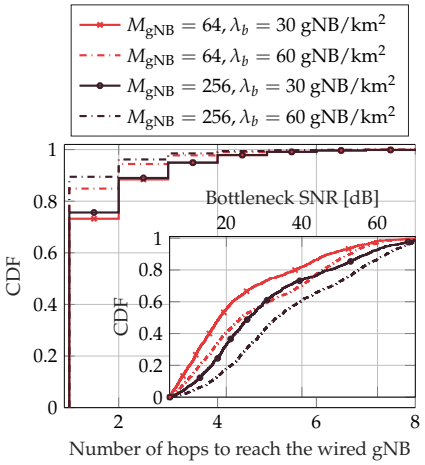


Fig. 5.2: Performance of the WF policy with different values of the number of antennas M_{gNB} and gNB density λ_b.

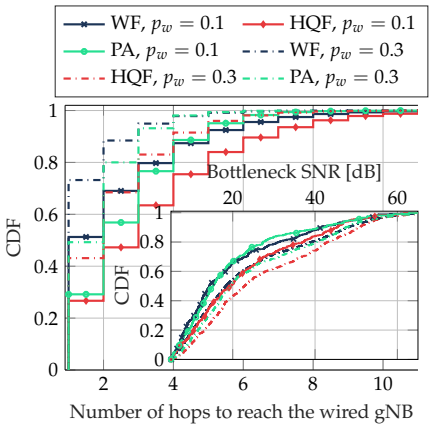


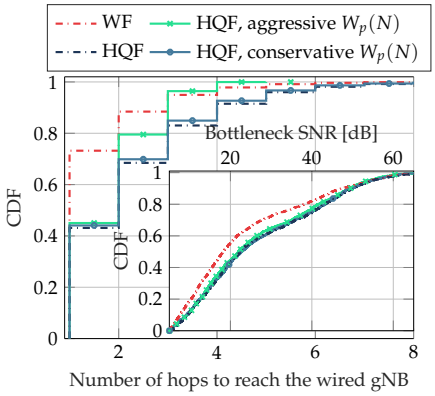
Fig. 5.3: Comparison of WF, HQF and PA policies, without WBF, for $M_{gNB} = 64$ antennas at the gNBs, $\lambda_b = 30$ gNB/km^2 and $p_w = 0.3$.

configurations show, on average, comparable performance in terms of bottleneck SNR. However, for low SNR regimes, i.e., when considering farther nodes and more demanding signal propagation characteristics, densification is more effective than directionality.

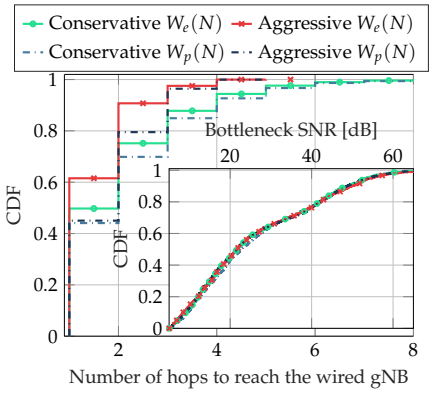
Path selection policies Fig. 5.3 compares the performance of the different path selection algorithms for different values of p_w, without WBF. In general, increasing p_w makes it possible to minimize the number of hops required to forward the backhaul traffic from a wireless node to the core network and, at the same time, guarantees more efficient relaying operations. However, the trade-off oscillates between more robust backhauling and more expensive network deployment and management. Moreover, although the HQF policy delivers the best bottleneck SNR performance, it exhibits the worst behavior in terms of number of hops, as it greedily selects the strongest available gNB as a relay regardless of the nature (i.e., wired or wireless) of the destination node. On the other hand, both WF and PA mechanisms have the potential to reduce the number of hops since the selection is biased by the availability of the wired gNB (independent of the quality of other surrounding cells) and by context information related to the position of the wired nodes, respectively. Conversely, both approaches degrade the quality of the bottleneck link as they may end up selecting a suboptimal node among all the candidate relays within reach.

Interestingly, we observe that, when the number of available wired gNBs is very low (i.e., $p_w = 0.1$ and for low SNR regimes), the PA policy performs better than WF in terms of both number of hops and bottleneck SNR. Fig. 5.3 indeed shows that the PA policy needs a smaller number of hops than WF (and also HQF) for the paths with 4 or more hops. In low SNR and $\lambda_{w,b}$ regimes, the WF scheme asymptotically operates as HQF and, therefore, the best choice is to select the parent which is geographically closest to a wired gNB with the PA strategy (for $p_w = 0.3$ this phenomenon is obviously less pronounced but still PA and WF reveal comparable performance in low SNR regimes).

WBF configurations In Fig. 5.4a we compare the behavior of the HQF and the WF policies when considering different WBF configurations to bias the path selection results. First, we see that, since the WF approach is designed to minimize the number of hops to reach a wired gNB, it generally outperforms any other architecture for the hop-count metric. However, the quality of the bottleneck



(a) Comparison of WF, and HQF policies with and without WBF.



(b) Comparison HQF policies with polynomial and exponential WBF.

Fig. 5.4: Impact of WBF (aggressive or conservative, polynomial or exponential) on the performance.

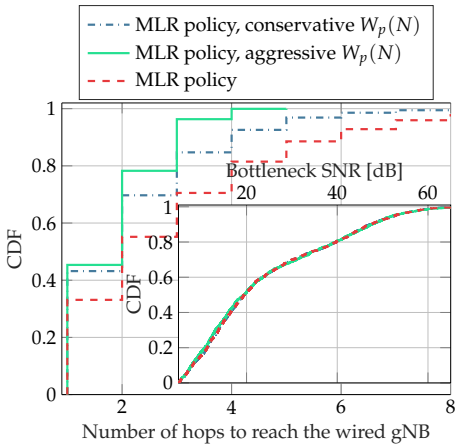


Fig. 5.5: Comparison of MLR policy with and without WBF.

link inevitably decreases (on average by more than 4 dB compared to its HQF counterpart), thereby increasing the risk of communication outage between the endpoints. Moreover, for bad SNR regimes (i.e., as the probability of detecting valid wired nodes reduces) the HQF scheme implementing aggressive WBF achieves the best performance in terms of both number of hops and bottleneck SNR.

Second, we observe that, although a conservative WBF applied to an HQF scheme does not provide any significant performance improvements with respect to a pure HQF approach, a more aggressive design of the bias function has the ability to remarkably reduce the number of hops required to forward the backhaul traffic to a wired gNB, without any visible degradation in terms of SNR. We deduce that it is highly convenient to configure very aggressive[1] WBF functions since, for a multi-hop scenario, they deliver more efficient relaying operations without affecting the communication quality.

Third, in Fig. 5.4b we compare the behavior of the HQF policy with polynomial and exponential WBFs. Based on the design choices presented in Tab. 5.4 and according to Eqs. (5.1) and (5.2), the

[1]Of course, if the WBF parameters are too aggressively configured, the HQF approach will more likely operate as a WF policy, with all that this implies (including, but not limited to, a detrimental degradation of the bottleneck SNR).

exponential bias function is more aggressive than the polynomial one for all values of N_h, i.e., the current number of hops. However, the exponentially-biased HQF approach, because of its inherently aggressive nature, is affected by SNR deterioration, though moderate (i.e., smaller than 1 dB on average), with respect to its polynomially-biased counterpart.

MLR performance While the IAB results presented in the previous paragraphs were based on SNR considerations, i.e., the candidate parent is chosen according to the instantaneous quality of the received signal, the CDF curves displayed in Fig. 5.5 analyze the performance of the MLR backhauling approach which relies on the instantaneous cell load and the Shannon rate as a metric for the path selection operations. We observe that Fig. 5.5 leads to the same conclusions previously set out, i.e., the design of aggressive polynomial bias functions has the potential to significantly reduce the number of hops without affecting the quality of the communication (in terms of bottleneck SNR). Aggressive exponential WBFs are able to further reduce the number of relaying events, though this may slightly undermine the quality of the weakest link.

5.5 End-To-End Evaluation of IAB Architectures

IAB has been investigated as an approach to relay access traffic to the core network wirelessly, thereby removing the need for all gNBs to be equipped with fiber backhaul. Along these lines, in this section we evaluate, through simulations, the practical feasibility of designing IAB solutions in the mmWave bands, and investigate the impact on the IAB paradigm on end-to-end network metrics.

5.5.1 An ns-3 Module for IAB Architectures

Discrete-event network simulators, e.g., ns-3, enable full-stack simulation of complex and realistic scenarios, and therefore represent a viable tool for accurate system-level analysis. IAB performance, therefore, is assessed extending the current ns-3 mmWave modules [156] with IAB functionalities, as will be described in the following subsections [170]. The IAB framework (for which the source code is publicly available[2]) features a new ns-3 `NetDevice`, the `MmWaveIabNetDevice` with a dual stack for access and backhaul, an extension of the ns-3 mmWave module schedulers, and network procedures to support IAB nodes in a simulation scenario. Moreover, we simulate the wireless relaying of both data and control plane messages, in order to accurately model the IAB operations.

5.5.1.1 IAB Node

As mentioned in [185], the IAB nodes should re-use the specifications for the access stack of NR as much as possible. At the moment, there are a few protocol stacks being discussed in the 3GPP [186]. All of them, however, include PHY, MAC and RLC layers, and differ because of the support of layer-2 (i.e., RLC or PDCP) or layer-3 relaying. Given the need for a flexible solution, able to adapt to the direction that the 3GPP will take, we have implemented a light layer-3 relaying solution, i.e., each backhaul radio bearer is set up locally, and an adaptation layer above the PDCP handles the forwarding of the packets from the access to the backhaul PDCPs. Fig. 5.6 shows the protocol stack for an IAB node and the classes that model it.

The main novelties are the `MmWaveIabNetDevice` and the `EpcIabApplication` classes. The first is an extension of the ns-3 `NetDevice` class, and, similarly to the `NetDevice` implementations of the UE

[2] `https://github.com/signetlabdei/ns3-mmwave-iab`.

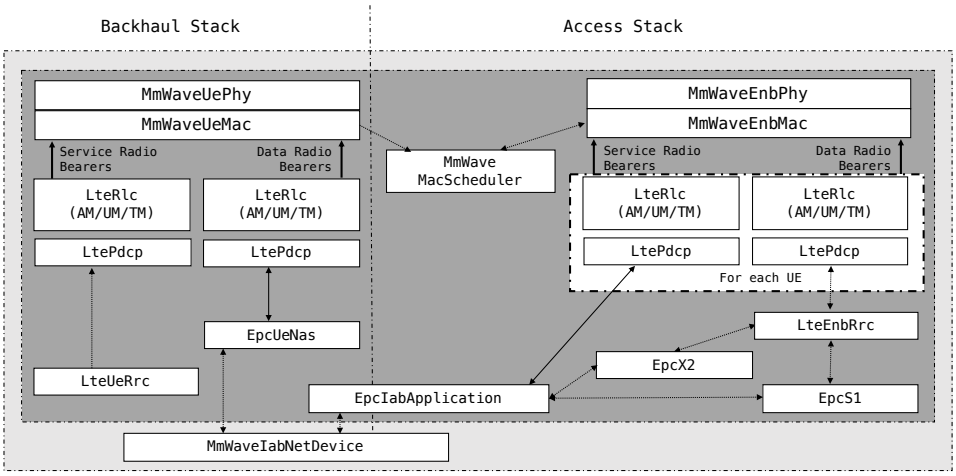


Fig. 5.6: Protocol stack and organization of the ns-3 classes for an IAB node.

and gNB, holds pointers to all the objects that model the other layers of the protocol stack. Moreover, it is internally used in the ns-3 model to forward packets between an instance of the `EpcUeNas` class in the backhaul stack and the `EpcIabApplication` in the access stack.

The `EpcIabApplication`, instead, implements the main logic related to the control and data plane management in the IAB node. In particular, for the data plane, the `EpcIabApplication` class is in charge of applying the forwarding rules for local UEs, i.e., those directly connected to the IAB node this class belongs to, and for remote UEs, i.e., those connected to downstream IAB nodes. In this case, the traffic will be forwarded to the local bearer mapped to the downstream IAB device. More details on how the routing is performed will be given in Sec. 5.5.1.2. This class is also responsible for the processing and forwarding of control packets for the interfaces toward the core network and the other neighboring gNBs. When a control message is received on either the access or the backhaul interface, the `EpcIabApplication` checks if it is a local message, i.e., if the destination is the RRC layer of the current IAB node, and, if this is the case, forwards the packet to the RRC. Otherwise, as done in the data plane, the packets are relayed via one of the downstream IAB nodes.

The other classes are the same as those used in the UE protocol stack (for the backhaul) and gNB protocol stack (for the access). The consequence is that, in the access, the UEs in the scenario consider the IAB node as a normal gNB, and, similarly, in the backhaul, the parent gNBs and/or IAB nodes consider the IAB child as a UE. Therefore, there is no need to adapt the UE and gNB ns-3 implementations to support the IAB feature. The only change is the extension of the gNB schedulers, to support the multiplexing of access and backhaul in the same resources, and the introduction of a new interface between the access and backhaul MAC layers. These extensions will be described in Sec. 5.5.1.3. Nonetheless, additional enhancements can be introduced in future releases, to improve the overall performance of the IAB protocol stack and track the 3GPP SI and specifications on IAB.

5.5.1.2 Single- and Multi-Hop Control Procedures

Given that the 3GPP is still considering IAB as an SI, there are no standard specifications yet on control procedures to support IAB networks. Nonetheless, the SI [185] specifies that both single- and multi-hop topologies should be considered, and that the IAB node should be able to autonomously connect to the network, adapt the access and backhaul resource partitioning and, eventually, independently update the parent node in case of blockage. All these features require specific control procedures, and, given the high level of detail of the ns-3 model, we implemented a number of re-

alistic control procedures, which involve an exchange of messages on the wireless backhaul links to set up and automatically configure the network. These can be easily updated to implement different procedures that the 3GPP may specify in the future.

In particular, we assume that the parent IAB node for a backhaul link terminates the NG control interface to the core network (i.e., the NR equivalent of the LTE S1 interface) [187], and that it takes care of forwarding the control messages towards the network servers that host the Access and Mobility Management Function (AMF). Moreover, the IAB node has a similar role with respect to the UEs connected to it, as would happen with a traditional wired gNB. Thanks to this design, the differences with respect to the 3GPP specifications for the access stack are minimized. This configuration makes it possible to seamlessly support both single- and multi-hop deployments, given that the architecture of the upstream portion of the network is transparent to each IAB node, which will simply relay all of its packets to the parent. Furthermore, for the purpose of packet transport in the backhaul network, we exploit GPRS Tunneling Protocol (GTP) tunnels from each IAB node to the relevant element in the core network (i.e., the server with control functions or the packet gateway). Each data bearer of all the UEs (and IAB nodes, for the backhaul part) is associated with a unique tunneling ID, and all the packets sent on backhaul links will be associated with a GTP header carrying that ID.

We also implemented realistic autonomous access and configuration procedures for the IAB nodes. When the IAB selects its parent node during the IA procedure (see Chapter 3 and [53,89]), the parent sends an initial message to the AMF, which will reply with the configuration for the backhaul bearer between the IAB node and its parent. These messages will be relayed by all the IAB nodes in the path between the parent and the donor gNB, and each of them will register the presence of an additional downstream IAB device. Notice that there may be multiple IAB children for each parent, therefore the parent has to match the new downstream node to the correct child to correctly route the other control and data packets. For the UEs, there is no difference between a wireless relay and a gNB with a wired connection to the core network. Therefore, the UE's IA procedure does not change, and the IAB node will take care of forwarding the relevant control messages to the AMF and the other network functions involved in the IA. Moreover, the upstream relays and the donor gNB will exploit the control messages for the UE's IA to associate to each IAB bearer the total number of downstream UEs. Finally, during the UE IA procedure, each gNB associates the GTP tunneling ID of the bearers of downstream UEs to a local IAB child, so that, when a backhaul packet is received, the gNB uses the information in the GTP header to correctly route the packet.

5.5.1.3 Backhaul-Aware Dynamic Scheduler

The MAC and the associated scheduler are a key component in the design of scheduled wireless relay architectures in which the resources between the access and the backhaul are shared. In order to avoid self-interference between access and backhaul, indeed, there is a need to multiplex the two interfaces. In our implementation, we consider Time Division Multiple Access (TDMA), but we plan to extend the support to spatial division multiplexing in future releases, to harness the directionality of mmWave communications. Moreover, the scheduler is usually not part of the 3GPP specifications, and, therefore, equipment vendors have the possibility of designing custom solutions in this domain.

We opted for a distributed scheduling solution, in order to minimize the difference in the scheduling mechanism with respect to a traditional access-only scenario, and to limit the amount of control overhead that a centralized solution would require. In the ns-3 IAB module, each gNB (either wired or wireless) indeed schedules the resources for its access interface (i.e., for both UEs and IAB children) independently of the other gNBs, as would happen in a traditional network without IAB. In a TDMA setup, however, the IAB node cannot schedule resources in the time and frequency slots already allocated to the backhaul by their parent. Therefore, if at time t the relay has to perform a

scheduling decision for subframe $t + \eta$, then it has to be already aware of the scheduling decision of its parent for $t + \eta$. Given a delay ϵ for the communication of scheduling information between the parent and the relay, then the parent should perform its scheduling decisions for $t + \eta$ at time $t - \epsilon$.

In order to efficiently address this issue, we implemented a *look-ahead backhaul-aware scheduling* mechanism. The backhaul-aware component is given by a new interface between the access and the backhaul MAC layers. The backhaul MAC layer is seen as a UE by the parent node, and thus will receive Downlink Control Informations (DCIs) with the scheduling and modulation and coding scheme information for η subframes in advance. Then, the backhaul MAC shares DCI with the scheduler of the IAB node (in the access stack), which registers the resources occupied by backhaul transmissions for the relevant subframe. The look-ahead mechanism, additionally, makes it possible to adjust the value of η according to the maximum number of downstream relaying hops N_h from the current gNB to the farthest IAB node: the gNB schedules ahead by $\eta = N_h + 1$ subframes[3], and propagates this information with a DCI to the UEs and IAB nodes connected to it. In turn, these IAB nodes will schedule ahead by at most $\eta = N_h$ subframes. Each of them will consider the time and frequency resources allocated for their downlink or uplink backhaul transmission as busy, and will schedule access resources for their UEs and, eventually, for IAB nodes in unallocated resources.

We added the look-ahead and backhaul-aware capabilities to two of the ns-3 mmWave module schedulers, i.e., the `MmWaveFlexTtiMacScheduler` class, modeling a Round Robin (RR) scheduler, and the `MmWaveFlexTtiPfMacScheduler` class, implementing a Proportional Fair (PF) scheduling algorithm. Moreover, in a TDMA setup, with shared resources between the access and the backhaul, it is important to make sure that the parent gNB does not schedule all of the available resources to a single IAB node (e.g., if it is the only active terminal connected to the parent). Otherwise, the child IAB node would not be able to allocate any resource to the access. We thus limit the maximum number of time/frequency resources allocated to an IAB device to half of the total available resources.

5.5.2 Simulation Setup

The setup of a simulation with the IAB feature resembles that of a simulation with traditional wired-only backhaul [156]. We added two auxiliary methods in the `MmWaveHelper` class, which hides from the ns-3 user much of the complexity related to the configuration of the mmWave RAN and core network. Similarly to the methods used to set up UEs and gNBs, the `InstallIabDevice` method returns a `NetDevice` properly configured, with the stack described in Fig. 5.6.

The initial attachment of each IAB node to its parent gNB is performed by the methods `AttachIabToClosestWiredEnb` or `AttachIabToBestNodeHQF`. The latter scans the signal quality of the available IAB nodes or wired donors, and selects that with the highest SNR. Moreover, it avoids the creation of loops in the network tree. These helper methods, moreover, automatically register the new IAB nodes to the control entities in the core network, and define the default radio bearer that will be used for the backhaul link. Finally, by default, the UEs in the ns-3 mmWave module perform the initial attachment as soon as the simulation starts, i.e., at simulation time $t_s = 0$. Therefore, we added the `AttachToClosestEnbWithDelay` method that delays by D seconds the initial attachment of UEs to the chosen gNBs, either wired or wireless. This method can be used to let the UEs perform IA only after the IAB nodes have completed their IA and backhaul bearer setup.

[3]The additional subframe with respect to N_h is needed because the farthest IAB node (without IAB children) has to schedule its resources at least one subframe in advance, in order to transmit the DCI beforehand to its UEs

Table 5.5: Simulation parameters for IAB end-to-end evaluation.

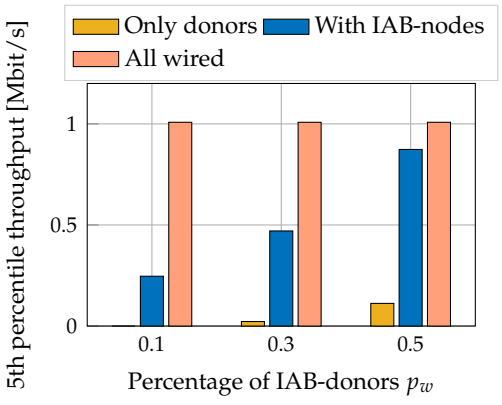
Parameter	Value	Description
B	1 GHz	Bandwidth of mmWave gNBs
f_c	28 GHz	mmWave carrier frequency
λ_b	$\{30, 45\}$ gNB/km^2	gNB density
M_{gNB}	64	gNB MIMO array size
M_{UE}	16	UE MIMO array size
p_w	$\{0.1, 0.3, 0.5\}$	Fraction of wired gNB
$\mathcal{R}_{\text{PHY}}$	3.2 Gbit/s	mmWave max PHY rate
$B_{\text{RLC,UE}}$	10 MB	RLC buffer size for UEs
$B_{\text{RLC,IAB}}$	40 MB	RLC buffer size for IAB nodes
R	220 Mbit/s	UDP application rate
P_{UDP}	1400 byte	UDP packet size

5.5.3 End-to-End Performance Evaluation

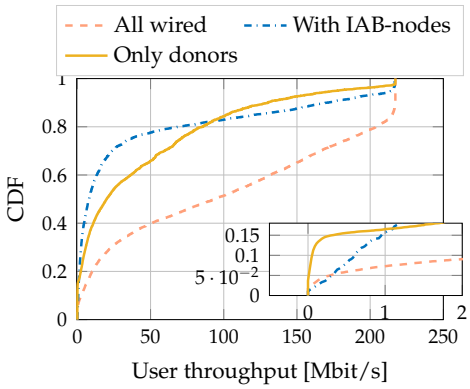
The IAB's performance is assessed in terms of end-to-end throughput and latency, considering different applications and traffic types. The main parameters for the simulations are reported in Table 5.5. In the scenario we investigate, $\mathcal{M}$ base stations are deployed following a PPP with density λ_b gNB/km^2, and a fraction $0 \leq p_w \leq 1$ of the $\mathcal{M}$ base stations have wired backhaul connections (i.e., the IAB-donors), while the others (i.e., the IAB-nodes) are wirelessly connected to the IAB-donors, perhaps over multiple hops. The network implements in-band backhaul, at 28 GHz, with TDM of the radio resources among the access and the backhaul links. We consider uniform rectangular antenna arrays in the base stations and UEs, with 64 and 16 elements, respectively, and the beamforming model described in [156]. The UEs are also deployed with a PPP with density $\lambda_u = 10\lambda_b$ UE/km^2 using the new ns-3 `OutdoorPositionAllocator` method, although we only evaluate the performance of the subset of users connected to a target gNB, which is either the first gNB deployed in a baseline scenario in which all nodes have a wired connection to the core network, or the first IAB-node that performs the initial access in an IAB scenario. The MAC layer performs HARQ retransmissions, and the RLC layer uses the Acknowledged Mode (AM) to provide additional reliability. The scheduler is `Round Robin`, with the look-ahead backhaul-aware mechanisms described in Sec. 5.5.1.3.

We test three different deployment scenarios. The best case is when all the $\mathcal{M}$ base stations in the network are equipped with a wired connection to the core network (i.e., the *all wired* scenario). This represents the most expensive solution, in terms of density of fiber drops, but permits the whole bandwidth to be used for access traffic. With the *IAB-nodes* option, $p_w\mathcal{M}$ base stations are IAB-donors, i.e., have a wired connection and $(1 - p_w)\mathcal{M}$ have wireless backhaul. Finally, the baseline is the one that 3GPP considers for comparisons with IAB solutions, described in [168], i.e., a deployment with only $p\mathcal{M}$ wired base stations and no IAB-nodes (the *only donors* configuration).

UDP user traffic In Fig. 5.7, we consider an IAB network where each user downloads content from a remote server with a constant bit-rate of $R = 220$ Mbps, using UDP as the transport protocol, thus introducing a full buffer source traffic model. The flow of each end-to-end connection does not self-regulate to the actual network conditions, thus congestion arises. This experiment aims to test the performance of an IAB setup in a saturation regime, where the access and backhaul links are constantly used. As expected, the best performance is provided by the all wired configuration, given that it provides the same access point density as the IAB setup, but avoids the multiplexing of resources between access and backhaul. On the other hand, it is possible to identify two advantages and one drawback of the IAB configuration with respect to the only donors one. A higher

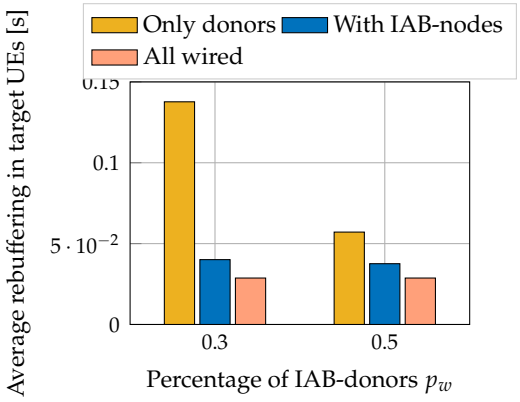


(a) Fifth percentile throughput.

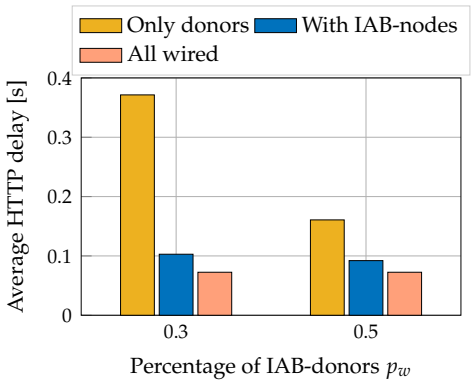


(b) Throughput CDF for $p_w = 0.3$.

Fig. 5.7: Fifth percentile and CDF of the throughput for UEs of a target IAB-node, varying the percentage of IAB-donors p_w and the deployment strategies, with $\lambda_b = 45$ gNB/km^2.



(a) Average rebuffering for DASH clients.



(b) Average delay to retrieve an HTTP web page.

Fig. 5.8: Performance for UEs of a target IAB-node, for different applications and $\lambda_b = 30$ gNB/km^2.

throughput for the worst users is achieved when using IAB-nodes, as shown by the fifth percentile throughput plot in Fig. 5.7a. In particular, for $p_w = 0.5$ (i.e., when the number of relays is equal to the number of IAB-donors), IAB has only 13% less fifth percentile throughput than the all wired configuration. Moreover, the usage of IAB-nodes likely offloads the worst users from the IAB-donors, and this frees up resources for users with the best IAB-donor channel quality, thereby enabling a higher throughput, as illustrated in Fig. 5.7b. The IAB solution, however, requires multiplexing of the wireless resources between access and backhaul. In a scenario where the links are always saturated, this results in a worse performance for the average users connected to the relays, which are throttled on the backhaul links by the round robin scheduler at the donors and have a smaller throughput than with the only donors setup.

DASH, HTTP user traffic The next set of results considers a more common use case, in which the users either stream video using DASH [188] or access web pages using HTTP from a remote server. This kind of source traffic is asynchronous and bursty, and, in the DASH case, the flow adapts itself to the varying capacity offered by the network, after some delays due to the signaling and convergence of the algorithm. Therefore, the network is not as stressed as in the previous experiment, and in this case the advantage of IAB is more visible. Indeed, thanks to the better channel seen

on average by the user due to more numerous nodes compared to the only donor case, and thanks to the asynchronous and independent nature of the traffic at each user, which provides greater multiplexing gains, the performance of the IAB network is not far from that of the network with all wired access points. In particular, Fig. 5.8a reports the average duration of a rebuffering event for a DASH stream, for all the users in a target base station. The rebuffering happens when the DASH framework does not adapt fast enough to the network conditions, or if the network capacity is not sufficient to sustain even the minimum video quality available in the DASH remote server. As can be seen, the only donors setup has the worst performance, with a 5 and 2 times higher rebuffering than the all wired configuration, for $p_w = 0.3$ and 0.5, respectively. The IAB deployment, instead, degrades the performance of the all wired only by 1.4 and 1.3 times, for $p_w = 0.3$ and 0.5, respectively. Likewise, Fig. 6.11 shows the average time it takes to completely download a web page, from the first client HTTP request to the reception of the last object: the trend is similar to that of the DASH rebuffering. Finally, for this kind of traffic, the improvement introduced by densification of IAB-donors (i.e., increasing p_w from 0.3 to 0.5) is less marked than with the constant bit-rate traffic shown in Fig. 5.7.

5.6 Conclusions and Design Guidelines

IAB networks present both benefits and limitations with respect to deployments where the radio resources are not multiplexed between the access and the backhaul. First, the IAB solution may present lower deployment costs and complexity with respect to the all wired setup, but, at the same time, splitting the available resources between access and backhaul traffic makes the overall network performance worse than in the all wired case under heavily loaded network scenarios. However, for bursty traffic the performance of the IAB solution approaches that of the all wired case. This shows that when evaluating the performance of IAB networks it is important to consider the specific use case and end-to-end applications that run on top of the network. Moreover, the results suggest that the main advantages of an IAB deployment, when compared to the only donors setup, come from an improvement in channel quality for cell edge users, on average, which consequently improves the area spectral efficiency. On the other hand, the deployment of an IAB network presents challenges related to the design and interactions at different layers of the protocol stack. An important issue is related to the enforcement of QoS guarantees in single and multi hop scenarios, so that mixed IAB traffic flows for end-to-end applications can safely coexist. Similarly, during the setup phase, in which the IAB-nodes join the network by performing initial access to their IAB parents, it is important to consider the attachment strategies to avoid overloading some IAB-donors, or excessively increasing the number of hops. In particular, how to design path selection strategies which are robust to network topology changes and end terminals' mobility is still an open research challenge which deserves further investigation.

Most of these system-level challenges are related to the design of ad hoc scheduling procedures at the MAC layer, able to efficiently split the resources between the access and the backhaul and provide interference management. Another important challenge is related to cross-layer effects emerging from retransmissions at multiple layers, and the configuration of RLC and transport layer timers may need to account for the additional delays related to the retransmissions over multiple hops and the reordering of packets at the receiver. At the PHY layer, it is interesting to evaluate the gain of the spatial multiplexing of the access and the backhaul, by using digital or hybrid beamforming, which could avoid the time/frequency multiplexing typically needed when using analog beamforming.

Overall, these challenges represent promising research directions to enable self-configuring, easy-to-deploy and high-performance IAB networks, which could represent a cost-effective solution for an initial ultra-dense NR deployment at mmWave frequencies.

Part II

Millimeter Waves for Future 5G Vehicular Networks

Chapter 6

Enabling Technologies for Future Vehicular Networks

6.1 Introduction

As introduced in Chapter 1, in recent years there has been a significant interest in the context of Connected and Autonomous Vehicles (CAVs), which have rapidly emerged as a means to guarantee a safer travel experience and to support multimedia applications [189]. The potential of CAVs can be fully unleashed through direct wireless communications among autonomous vehicles and to/from roadside infrastructures, a concept that is usually referred to as Vehicle-to-Vehicle (V2V) and Vehicle-to-Network (V2N) connectivity, respectively, cumulatively indicated as Vehicle-to-Everything (V2X). Today, basic V2V services are supported by the IEEE 802.11p standard [25], which offers data exchange at a nominal rate from 6 to 27 Mbps within a range of a few hundreds of meters, while V2N services are enabled by the C-V2X standard – based on the LTE technology [190] –, which guarantees a data rate of around 100 Mbps in high mobility scenarios. These radio technologies, however, were primarily designed to provide coverage and may not be able to satisfy the unprecedentedly high throughput demands (i.e., in the order of terabytes per driving hour [191], according to some estimates) of envisioned vehicular services [29], which will range from generation and maintenance of high-resolution road maps to cloud-assisted intelligent driving [192].

In this context, the automotive industry has devoted efforts to specifying new communication solutions, e.g., operating in the mmWave bands above 10 GHz (see Sec. 2.1), that allow vehicles to use very large bandwidths to communicate, thus guaranteeing very high transmission speeds [191]. Although this new band has gathered great attention for V2X, the mmWave paradigm comes with its own set of challenges [193], including severe path loss and susceptibility to blockage, thereby preventing long-lived communications, a critical prerequisite for safety operations. In order to fully understand the potential of this technology in a vehicular scenario, it is therefore important to carry out simulations that validate the feasibility of designing mmWave-aware solutions in view of the strict requirements of future transportation systems, a research challenge that is still largely unexplored.

This chapter is based on contributions presented in **[J4, J9, C1, C3, C5, C6, C9, C12,C14, C16, B2]**.

6.1.1 Motivations and Chapter Structure

Along these lines, in this chapter we provide simulation results that compare the performance of different RATs in realistic V2X environments. In Sec. 6.3 we overview the characteristics of the LTE, IEEE 802.11p and mmWave technologies as enablers for vehicular communications. In particular, we investigate the limits that prevent the direct employment of legacy communication protocols on high-frequency links, and highlight possible solutions at the different layers of the protocol stack to enable mmWave network operations.

In Sec. 6.4 we provide an end-to-end performance evaluation of the mmWave technology in a V2N scenario, considering different propagation models, application requirements, and traffic densities. ns-3 [151] is used as a simulator to consider end-to-end scenarios with a complete 3GPP-like TCP/IP protocol stack. Unlike most existing studies, we investigate both static and dynamic environments, thus exemplifying how high-speed mobility impacts the overall communication performance of vehicles. We show that mobility typically leads to very high levels of queuing and buffering at the MAC layer, dramatically increasing latency and packet loss. Frequent handovers in dense base station deployments is another potential drawback of mmWave systems, in particular when the vehicles operate in NLOS. We consider traditional UDP and TCP connections at the transport layer, which are of general interest for vehicular services. In addition, unlike typical performance analyses, we consider HTTP applications, which simulate web browsing traffic, and DASH flows, which simulate video streaming over the Internet, to consider more common use cases.

In Sec. 6.5 we focus on V2V networking. Unlike for the V2N case, it is not possible to simulate end-to-end scenarios due to the absence of appropriate simulation softwares.[1] We therefore present link-level simulations that exemplify, in terms of packet reception ratio, data rate capacity and outage probability, the impact of several automotive-specific parameters, e.g., the antenna array size and the vehicular traffic density, on the overall V2V network performance. We implement the channel model that the 3GPP has proposed for NR-V2X systems [31], which (i) supports deployment scenarios for urban/highway propagation, and (ii) incorporates the effects of path loss, shadowing, line of sight probability, and static/dynamic blockage attenuation. We also consider an extension of the 3GPP model, proposed in [195], that characterizes density-dependent propagation.

In Sec. 6.6 we demonstrate the potential of mmWaves to design fair and efficient attachment policies in a V2N scenario. In particular, we propose a QoS-aware scheme that integrates the mmWave and LTE radios to bias the cell selection as a function of the vehicular service requirements, preventing the overload of transmission links. Our simulations demonstrate that the proposed strategy improves the percentage of vehicles satisfying the application requirements, and delivers high-throughput association compared to state-of-the-art schemes based on received signal quality criteria.

Finally, in Sec. 6.7 we conclude the chapter by identifying guidelines for the design of V2X strategies at mmWaves. In particular, we make the case that, although the mmWave technology emerges as an enabler of the boldest data rate requirements of future automotive applications, still the characteristics of the legacy spectrum are able to deliver a good compromise between low end-to-end latency and high connection stability at long range, thereby making the LTE and IEE 802.11p standards promising access solutions to maintain robust communications.

[1] ns-3 features a WAVE module [194] for the performance of vehicular networks in an end-to-end environment in the IEEE 802.11p band. We are currently developing an extension of such module for the simulation of V2V communications in the mmWave bands too, which is fully compliant with the 3GPP standardization efforts in this domain and which follows the evaluation methodology agreed in [31] (whose main features have been summarized in Sec. 1.4.1). The source code will be made publicly available as soon as the implementation will be completed [27].

Table 6.1: Literature on mmWave vehicular networks.

Topic	Relevant References
V2X Operations	[196, 197] Use of mmWaves in a vehicular context. [198–200] IEEE 802.11ad V2X operations.
V2N Connectivity	[54, 201, 202] Stochastic geometry evaluations. [203–207] PHY/MAC evaluations. [208–210] End-to-end evaluation at sub-6 GHz frequencies. [211, 212] End-to-end evaluation at mmWave frequencies.
V2V Connectivity	[213] Analytical evaluation. [214–218] Multi-RAT integration. [219–221] Characterization of the mmWave channel.
V2N Attachment	[222–224] Attachment optimization. [225–227] Game theory approach. [87, 228] Combinatorial optimization approach. [201, 229–231] Stochastic geometry approach. [176] LTE/mmWave integration for cellular networks. [232–234] LTE/mmWave integration for vehicular networks.

6.2 Related Work

The application of the mmWave technology in a vehicular context is not new. Current automotive radars already operate in the 77 GHz spectrum [196]. For vehicular communications, mmWave was tested more than a decade ago [197], and dual-functional systems integrating the radar function to enable V2X communications based on the IEEE 802.11ad WLAN protocol have recently been investigated in the automotive world [198–200]. The potential of mmWaves to support intelligent transportation systems has been first recognized in [191, 191], while a list of challenges that need to be addressed to deploy mmWave automotive scenarios was provided in [193]. A mmWave architecture for future vehicular applications is also being investigated by the European Commission [235].

Vehicle-to-Network Connectivity Vehicular network performance in V2N has been traditionally evaluated through stochastic geometry [201]. In the mmWave context, papers [54, 202] modeled, through a stochastic approach, a V2N deployment operating at 28 GHz to characterize coverage and connectivity performance. Given the simplicity of their topology and their high level of automation, highway scenarios have been considered. Some other related work, e.g., [203–206], focused on PHY or MAC layer performance metrics (e.g., achievable transmission range or packet transmission probability) to motivate the use of mmWaves to facilitate Gbps communication for V2N scenarios. In our previous contribution [207], we showed that, even though mmWave systems are likely to offer high peak capacity, they may not be able to satisfy the stringent requirements of some automotive services in terms of stability and outage probability, though performance can be improved by considering very directional transmissions and frequent re-alignment operations. Such results, however, have not investigated the impact of the upper layers on the network behavior, thereby calling for more accurate system-level analyses.

System-level simulators, such as ns-3, have therefore been employed to simulate complex networks and enable full-stack simulation capabilities in realistic deployments [151]. ns-3 has been used, e.g., in [208–210], to simulate vehicular scenarios in end-to-end environments over LTE networks, even though propagation characteristics of legacy systems are completely different from those of mmWave channels. At the time of writing, our previous works [211, 212] are the only available con-

tributions providing an end-to-end evaluation of the performance of the mmWave technology in a V2N scenario.

Vehicle-to-Vehicle Connectivity While the research on V2N systems operating at mmWaves has recently been quite widespread, the literature on V2V networking is still very scarce. In [213] the authors shed light on the operational limits of mmWave bands as a viable technology for future high-rate V2V transmissions. Integration among different RATs has also been investigated as a means to support high-throughput and robust connectivity. In [214], for example, a multi-access edge computing framework integrating licensed sub-6 GHz band and mmWaves for inter-vehicle information distribution was proposed. In [215,216], link-layer simulations have been conducted to compare the performance of IEEE 802.11p and the mmWave technology to support V2V networking, aiming at providing insights on how both technologies can complement each other to meet the requirements of future automotive services. Petrov *et al.*, in [217], characterized the integration between mmWave and low terahertz bands and modeled the interference from the side lanes in two typical V2V deployments including highway and urban road environments. Coll-Perales *et al.*, in [218], introduced the first performance evaluation of the IEEE 802.11ad MAC and beamforming mechanism for mmWave V2V communications, highlighting opportunities and shortcomings that should guide the development of V2V connectivity in the above-6 GHz spectrum.

Some other papers have provided analytical characterizations of the mmWave channel in V2V scenarios. For example, in [219] an experimental characterization of the 38 GHz and 60 GHz radio channel was presented. In [220] we validated the channel model that the 3GPP has proposed for NR-V2X systems and exemplified the impact of several automotive-specific parameters on the overall network performance considering realistic simulation assumptions for typical vehicular scenarios. The impact of the positions of a mmWave antenna on a vehicle for typical V2V scenarios was finally studied in [221], and path loss measurements at 28 GHz were conducted to investigate how to minimize the effect of the surrounding objects.

Vehicle-to-Network Attachment Policies Different studies have tried to design the optimal vehicle association technique that avoids system overload and best distributes the available resources among the users. For example, Chen *et al.*, in [222], proposed joint optimization of channel selection, user association and power control to maximize spectrum and energy utilization efficiencies. In [223], Corroy *et al.* presented a new theoretical framework to study cell association for the downlink of multi-cell networks and developed a dynamic method that associates users to macro or pico nodes while maximizing the sum rate of all network users. Cell range expansion theory has also been used in [224] to perform user association based on the biased measured signal, i.e., balancing the load among high- and low-power eNBs.

Lately, researchers have tried to solve the user association problem using advanced mathematical tools, in particular game theory [236] and combinatorial optimization [237]. For instance, in [225], the authors have proposed load balancing methods for multi-tier networks with massive MIMO gNBs and demonstrated that the load-based association scheme terminates in a Nash equilibrium. Similarly, Xu *et al.*, in [226], presented a centralized user association algorithm that targets rate maximization, proportional fairness, and joint user association and resource allocation in a MIMO scenario. In [227], game theory was used to model user association in heterogeneous networks to guarantee QoS to human-initiated traffic while providing fair resource allocation for machine-to-machine services. In [228], Liu *et al.* formulated the user association issue as a nonlinear combinatorial problem and proposed a centralized scheme which guarantees fair and energy efficient attachment through Lagrange multipliers. In [87], the authors formulated a logarithmic utility maximization

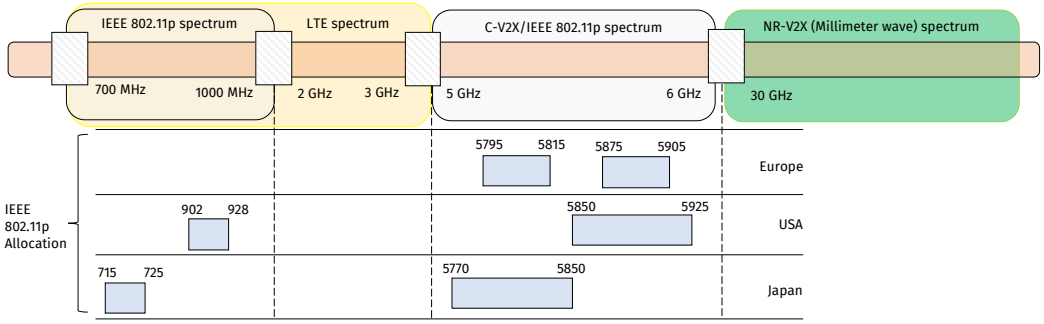


Fig. 6.1: Spectrum allocation for the main V2X radio technologies.

problem for single-user association, and showed that equal resource allocation is actually optimal, over a sufficiently large time window. However, most popular mathematical optimizations only apply to scenarios where the traffic flow generated by endnodes is approximately static. However, in the real world, traffic is not stable nor accurately predictable, thereby making traditional model assumptions invalid.

Stochastic geometry [201] has also emerged as a computationally tractable approach to model and analyze the performance of multi-tier heterogeneous networks [229]. In this regard, Dhillon *et al.*, in [230], exploited stochastic geometry to evaluate the performance of user association, based on received signal quality criteria, in a multi-tier cellular system. In a similar way, the authors in [231] formulated a throughput maximization problem subject to QoS constraints, and provided insights into the optimal spectrum allocation technique.

Most prior work on network association applies to LTE-only scenarios. LTE and mmWave heterogeneous networking, on the other hand, is much more sensitive to the cell association policy because of the significant propagation disparities of the two radios, and calls for innovative solutions that depend on the radio technology characteristics. In [176], Singh *et al.* made the case that, although mmWaves generally represent the preferred access technology, offloading users to more reliable radio interfaces may dramatically improve the rate of cell-edge users in case of sudden channel degradation. The aforementioned association policies were proposed for cellular networks, which might not be fully representative of a vehicular system due to the more challenging propagation and traffic characteristics of highly mobile vehicular nodes. Although some recent works in the literature have tried to provide preliminary insights into user association also in the context of vehicular networks [232], e.g., leveraging reinforcement learning [233] or information on the vehicular service requirements [234], there remain many open problems which call for innovative modeling and design solutions.

6.3 Vehicular Radio Technologies: an Overview

In this section we overview the characteristics of candidate RATs currently being considered to support V2X communications, and discuss their possible shortcomings in relation with future application requirements. Spectrum allocation of those RATs is reported in Fig. 6.1.

6.3.1 Long Term Evolution

Since its inception, the LTE cellular technology, operating in the sub-6 GHz spectrum, has represented an ideal candidate to support V2N operations [238]. First, LTE relies on a capillary deploy-

ment of eNBs offering wide area coverage and long-lived connectivity. Second, resource allocation is centrally managed by an eNB at every transmission opportunity, thereby satisfying service quality constraints while managing priorities in case of V2N applications competing for resources [23]. Third, LTE operates through omnidirectional transmissions and supports broadcast data distribution [190]. Fourth, the LTE interface may guarantee transfer latencies in the radio access theoretically lower than 100 ms, which is particularly beneficial for delay-sensitive vehicular applications.

Nevertheless, LTE was originally designed for mobile broadband traffic and its capability to support V2N communications is still an open question. The main concern comes from LTE's network architecture, that is configured to keep non-active terminals in idle mode: transitions to connected mode may require several seconds [239], which is intolerable for vehicular services. The access and transmission latency also increases with the number of users in the cell, thus raising issues. Moreover, despite the almost ubiquitous coverage of LTE, still the connection may not be always available (e.g., in underground areas). Finally, LTE offers limited downlink capacity (i.e., around 100 Mbps in Release 8, though much lower rates are expected in mobile scenarios), which might not be enough to satisfy the requirements of some V2N applications.

6.3.2 IEEE 802.11p

The IEEE 802.11p standard supports the PHY and MAC layers of the Dedicated Short Range Communication (DSRC) transmission service and offers V2V data exchange at a nominal rate from 6 to 27 Mbps within a range of a few hundreds of meters [25].[2] In the US, DSRC can count on a total spectrum of 75 MHz in the 5.9 GHz frequency band, divided into seven 10-MHz channels, with 5 MHz of guard band at the lower end of the spectrum, with an additional 14 MHz allocation in the 915 MHz bands. This standard embeds certain desirable features for V2V communications. Endpoints can operate without a network infrastructure, removing the need for prior exchange of control information and thus bringing a significant advantage in terms of latency with respect to regular Wi-Fi or legacy cellular operations. Moreover, IEEE 802.11p implements the carrier sensing multiple access with collision avoidance (CSMA/CA) mechanism at the MAC layer, thereby guaranteeing a fully distributed and uncoordinated access to the wireless channel, with no need for a resource allocation procedure.

Nevertheless, the IEEE 802.11p standard presents some inherent limitations. First, the throughput and delay performance degrades as the network load increases, even though there are ways of mitigating congestion by adjusting the message rate in the application layer [241]. Second, the channel access mechanism is prone to the *hidden node* problem, which may result in packet collisions. Third, the limited bandwidth of IEEE 802.11p systems results in limited data rates which may not satisfy the requirements of some categories of future V2V applications, e.g., extended sensors.

6.3.3 Millimeter Waves

Recently, the mmWave band has been investigated as a means to enhance automated driving and address the stringent throughput and latency demands of emerging vehicular applications. As described in detail in Chapter 2, these frequencies, combined with high-order modulation and MIMO techniques, offer orders of magnitude higher bit-rates than legacy vehicular technologies [242]. Moreover, unlike in LTE/IEEE 802.11p, mmWave systems operate through highly directional communications which tend to isolate the users and deliver reduced interference. Security and privacy are also inherently improved because of the short-range transmissions which are typically established [47].

[2]The IEEE 802.11p standard has also been considered as an option to serve V2N traffic [240]. However, it suffers from scalability issues, long delays, and short-range short-lived connectivity, thereby making LTE a more appropriate ratio technology for V2N connectivity, as explained in Sec. 6.3.1.

Although mmWave-assisted V2X operations are very attractive from the throughput perspective, they still pose significant challenges [51, 193]. Signals propagating in the mmWave spectrum suffer from severe path loss and susceptibility to shadowing, thereby preventing long-range transmissions (assuming isotropic propagation). Furthermore, directionality requires precise alignment of the transmitter and the receiver and implies increased control overhead. Additionally, mmWave links are highly sensitive to blockage and have ever more stringent requirements on electronic components, size, and power consumption. Given that the challenging radio conditions caused by the mobility of vehicles are further exacerbated considering the dynamic topology of the vehicular networks, how to directly apply the mmWave technology to a V2X deployment is still not clear and has become a research focus in the area of intelligent automotive systems.

6.3.3.1 PHY/MAC Protocol Design

As mentioned in Sec. 2.2, mmWave systems require require precise alignment of the transmitter and receiver beams, In the automotive context, beam management can be designed by extracting information from radar signals [243]. Simulations confirm that radars can be a useful source of side information and can help configure the mmWave V2X links. Location-aided beamforming strategies have also been studied, e.g., in [244], to achieve ultrafast connectivity between vehicles. In particular, adaptive channel estimation based on location information allows the connection time to be substantially reduced. The predictability of the vehicular mobility can be exploited to design beam management techniques that are able to estimate the vehicle's trajectory and derive the optimal beam orientation accordingly [204].

6.3.3.2 Network Protocol Design

Little work exists regarding the communication performance of the network layer (especially routing) for future V2X networks, as introduced in Sec. 1.5. In the mmWave context, some works tried to design network layer protocols specifically tailored to multi-hop systems with directional antennas. In [245], the authors proposed an Optimal Geographic Routing Protocol (OGRP) that selects the appropriate multi-hop relays considering the specific features of mmWave propagation. Other solutions implement some sort of multipath routing that allows a vehicular node to establish multiple connections through different access technologies, besides using device-to-device (D2D) transmissions. In [246], a multi-hop concurrent transmission scheme is proposed and, by properly breaking one single-hop low-rate link into multiple shorter high-rate links and allowing non-interfering nodes to transmit concurrently, the network resources can be efficiently used to improve the network throughput.

6.3.3.3 Transport Protocol Design

A relevant issue in vehicular networks is the performance analysis of transport protocols, especially congestion control using the TCP, as introduced in Sec. 1.5. One possible way to design mmWave-aware transport layer protocols is to dynamically adapt the TCP flow according to the instantaneous channel propagation conditions of the surrounding nodes, thereby reducing the congestion window size in case the path between the endpoints is obstructed [247]. The hybrid and joint use of V2V and V2N communications can also ensure better QoS and transmission efficiency, especially when delivering large data contents [248]. The message may indeed be divided into several segments and delivered to multiple vehicles (i.e., to remove possible points of failure on the propagation paths), or shared among multiple infrastructure nodes, which are less affected by forwarding

constraints. Finally, multipath-TCP [249], a standard that makes it possible to multiplex a TCP connection over multiple end-to-end paths, is another promising approach to improve the reliability of high-capacity networks. However, there are several issues with the traditional congestion control algorithms, in particular when coupling mmWave and LTE links. Nevertheless, the definition of innovative transport-layer schemes for V2X systems is just in its infancy and therefore represents a wide-open research area.

6.4 Millimeter Waves in V2N: Performance Evaluation

In this section we conduct extensive simulations to compare the performance of the mmWave and LTE RATs in realistic V2N environments. ns-3 is used as a simulator to consider end-to-end metrics. Unlike traditional performance analyses, which rely on PHY or MAC layer quality metrics, we investigates the impact of the upper layers on the network behavior, thereby guaranteeing more accurate system-level analyses. Moreover, unlike analytical evaluations, which typically adopt conservative assumptions on the signal propagation, we consider full-stack simulations, which allow to estimate the system performance accounting for detailed protocol implementations. In detail, Sec. 6.4.1 describes the ns-3 modules architecture and introduces our system-level parameters and overall simulation objectives. Sec. 6.4.2 discusses how the selection of the RAT affects the overall network performance, and Sec. 6.4.3 demonstrates the impact of mobility on the V2N scenarios.

6.4.1 Evaluation Methodology

6.4.1.1 The ns-3 Architecture

Our performance evaluation is conducted using ns-3, an open source software which allows the simulation of complex networks with a very high level of details. The ns-3 simulator features:

- An LTE module [250], which provides (i) a basic implementation of LTE devices, including propagation models, PHY and MAC layers, (ii) RRM of the data radio bearers, the MAC queues and the RLC instances, and (iii) full-stack support for both uplink and downlik packet scheduling. The path loss is based on pure geometric considerations which deterministically evaluate whether the V2N link is blocked by buildings or not, and is implemented in the `Lte3gppPropagationLossModel` class following the model in [251]. We consider a fast Rayleigh fading, which is modeled as a stochastic gain with unit power (in linear scale).
- A mmWave module, which builds upon the LTE module and implements a 3GPP-like cellular protocol stack including (i) a custom PHY/MAC layer implementation for both UE and eNB devices, (ii) support for directional transmissions through analog beamforming, and (iii) a complete TCP/IP protocol suite. The propagation is based on the 3GPP model for frequencies above 6 GHz [252], which characterizes the time correlation among the channel impulse responses to account for spatial consistency, and is implemented in the `MmWave3gppPropagation LossModel` class. Moreover, since the effects of high mobility at mmWaves result in rapidly time-varying channels, ns-3 implements a detailed fading model in the `MmWave3gppChannel` class. In particular, the model characterizes spatial clusters, subpaths, angular beamspreads and the Doppler shift, which is a function of the total angular dispersion, carrier frequency and mobile velocity.

Table 6.2: V2N Applications' design characteristics.

Parameter	Value
UDP packet size P_{UDP}	1400 B
UDP application rate R	$\{224, 28, 11, 1\}$ Mbps
UDP simulation time T_{UDP}	0.5 s
TCP version	`HighSpeed`
TCP min. RTO	200 ms
TCP segment size P_{TCP}	1400 B
TCP sender buffer size	52 KB
TCP receiver buffer size	52 KB
TCP simulation time T_{TCP}	10 s
DASH version	MPC
DASH interarrival time	$\text{Exp}(0.5)$ s
DASH segment history	5
DASH simulation time T_{DASH}	60 s
HTTP object size	$\log \mathcal{N}(100, 40)$ KB
HTTP reading time	Exp(30) s
HTTP simulation time T_{HTTP}	200 s

6.4.1.2 Application Characteristics

The behavior of different applications, whose design parameters are collected in Table 6.2, is investigated.

- *User Datagram Protocol (UDP)*. UDP-like vehicular applications generate packets of $P_{\text{UDP}} = 1400$ bytes at a constant interarrival rate ranging from $\tau_{\min} = 50$ μs to $\tau_{\max} = 10000$ μs, corresponding to application rates ranging from $R_{\max} \simeq 224$ Mbps to $R_{\min} \simeq 1$ Mbps, to test the performance of vehicles in relation with different service requirements [30]. In particular, high-rate UDP transmissions are compatible with V2N applications offering extended sensor sharing services, while lower source rates are typical of platooning systems having very stringent requirements in terms of communication delay and reliability but for which the size of the exchanged messages is reasonably small.
- *Transmission Control Protocol (TCP)*. TCP-like vehicular applications, unlike their UDP-like counterparts, implement congestion control algorithms at the transport layer to adapt and self-regulate the traffic flow to avoid congestion. TCP guarantees reliable, ordered, and error-checked packet delivery, although it may suffer from high end-to-end latency. In this study, `HighSpeed` TCP is used as the congestion control protocol, since it natively maintains large congestion windows and has been proven to efficiently support broadband connections.
- *Dynamic Adaptive Streaming over HTTP (DASH)*. DASH applications, which rely on TCP, simulate adaptive video streaming services in which, contrary to conventional HTTP-based mechanisms, the bit-rate level for future chunks is chosen according to the network conditions to deliver the highest possible quality of experience. In this study, *Model Predictive Control (MPC)* is implemented for bit-rate adaptation, and streaming data chunks are generated according to a PPP of parameter 0.5 s.
- *HyperText Transfer Protocol (HTTP)*. HTTP applications simulate web browsing traffic requests. Technically speaking, the client transmits request objects to demand a service from the server. Depending on the type of request received, the server transmits either a main object, i.e., the HTML file of the web page, or an embedded object, e.g., an image referenced by the HTML file. For these applications, the required throughput depends on the content type, while la-

Table 6.3: V2N system-level simulation parameters.

Parameter	Value	Description
LTE B	20 MHz	LTE total bandwidth
LTE f_c	2.3 GHz	LTE carrier frequency
mmWave B	1 GHz	mmWave total bandwidth
mmWave f_c	28 GHz	mmWave carrier frequency
P_{TX}	30 dBm	Transmission power
NF	5 dB	Noise figure
M_{eNB}	8×8	mmWave eNB antenna array size
M_V	4×4	mmWave vehicle antenna array size
B_{RLC}	10 MB	RLC buffer size

tency is reasonably tolerated because the download can be generally completed within some flexible time frame. In this study, the main object size is modeled as a lognormal random variable of mean 100 KB and standard deviation 40 KB. A major portion of the traffic pattern is reading time, which is modeled as an exponential random variable of mean 30 s during which traffic is not generated. Because of this, the simulation duration is set to $T_{\text{HTTP}} = 300$ s in order to inject significant traffic in the system.

6.4.1.3 Performance Metrics

The statistical results are derived through a Monte Carlo approach, where multiple independent simulations are repeated to obtain different quantities of interest. In particular, we analyze the following end-to-end performance metrics.

- *Average throughput*, the number of RX bytes per vehicle divided by the total simulation time.
- *Total throughput*, the sum of the throughput of all vehicles within the coverage of a given eNB.
- *5th percentile throughput*, the throughput that only 5% of the vehicles cannot achieve (it represents the performance of cell-edge nodes, the most resource-constrained network entities).
- *Average latency (per packet)*, from the time each packet is generated at the application layer to when it is successfully received assuming perfect beam alignment (it is therefore the latency of only the correctly received packets).
- *Jain's fairness index*, which is used to determine whether vehicles are receiving a fair share of the cell resources, and is defined as

$$J = \frac{\left(\sum_{i=1}^{\mathcal{N}_V} S_i\right)^2}{\mathcal{N}_V \sum_{i=1}^{\mathcal{N}_V} S_i^2}, \tag{6.1}$$

  where $\mathcal{N}_V$ is the number of users in the cell and S_i is the throughput experienced by the i-th vehicle. The result ranges from $1/\mathcal{N}_V$ (most unfair) to 1 (perfectly fair).
- *X2 traffic ratio*, the ratio between the average traffic load on X2 links connecting the different eNBs (which depends on the amount of traffic that is forwarded during handover events) and the average throughput.
- *Handover delay (per packet)*, the average time required to complete a handover, i.e., from the time the handover is initiated (the source eNB sends an *RRC Connection Reconfiguration* message) to the time the packet is forwarded to the target eNB.

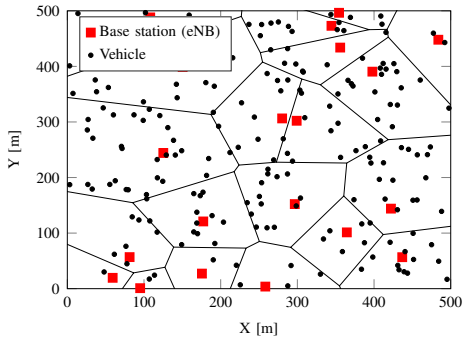


Fig. 6.2: Stochastic simulation scenario with $\lambda_b = 80$ eNB/km^2 and 10 vehicles per eNB. In the considered area (500×500 m), a total of 20 eNBs and 200 vehicles are deployed.

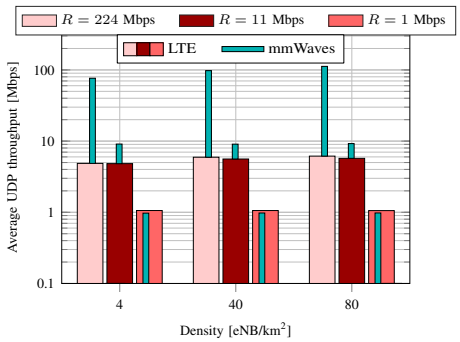


Fig. 6.3: Average UDP throughput for different values of the eNB density and UDP application rate. Narrow (wide) bars refer to a mmWave (LTE) system.

6.4.1.4 Simulation Parameters and Objectives

The simulation parameters are based on realistic system design considerations, and are detailed Table 6.3 as in [211,212]. At the PHY layer, LTE eNBs operate in the 2 GHz band, with 20 MHz of bandwidth and omnidirectional transmissions, while mmWave eNBs[3] operate at 28 GHz with 1 GHz of bandwidth and are equipped with UPAs of 8×8 elements to establish directional communications through beamforming. Vehicles are also equipped with 4×4 UPAs. For both LTE and mmWave systems, the transmission power and noise figure are set to $P_{\text{TX}} = 30$ dBm and NF $= 5$ dB, respectively. The MAC layer performs HARQ to enable fast retransmissions in case of corrupted receptions, and the RLC layer, whose buffer is $B_{\text{RLC}} = 10$ MB, uses AM to offer additional reliability. Our main objective is to investigate the impact of several automotive-specific parameters on the end-to-end communication performance of V2N nodes. In particular, our results are given as a function of (i) the RAT (LTE or mmWaves), (ii) the eNB density, (iii) the vehicle density, (iv) the vehicular application, (v) the signal propagation characteristics (urban or rural propagation, i.e., UMi-Street-Canyon or RMa according to the 3GPP terminology), (iv) the deployment scenario (urban or highway).

6.4.2 End-to-End Performance Evaluation: Impact of RAT Selection

In this section we provide numerical results to evaluate the end-to-end performance of the LTE and mmWave technologies in a V2N scenario considering different applications, eNB densities and channel propagation characteristics. To do so, we consider a stochastic, static deployment (as depicted in Fig. 6.2) in which mmWave and LTE eNBs are deployed over an area of 500×500 square meters according to a PPP of density λ_b varying from 4 to 80 eNB/km^2 (the trade-off involves signal coverage and deployment costs). We also deploy an average of 10 vehicles per eNB, as foreseen in [8] for a dense environment. This is to guarantee that results are not influenced by the interaction of the vehicles with the road and network elements nor by the dynamics of the mmWave channel.

[3]To be precise, Next Generation Node Base (gNB) is the 3GPP NR terminology for a base station, possibly operating at mmWaves, while eNodeB (eNB) identifies a legacy base station operating in the sub-6 GHz bands. However, for ease of notation, and with no ambiguity, in this chapter the two terms are used interchangeably.

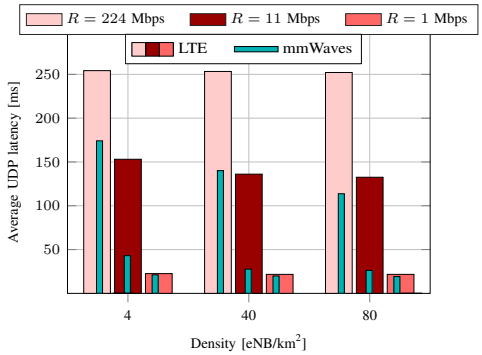


Fig. 6.4: Average UDP latency for different values of the eNB density and UDP application rate. Narrow (wide) bars refer to a mmWave (LTE) system.

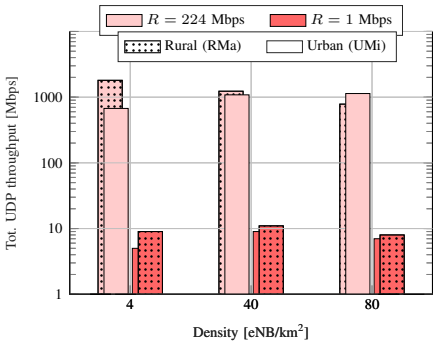


Fig. 6.5: Total UDP throughput for different values of the eNB density and UDP application rate. Dotted (straight) bars refer to a rural (urban) scenario.

6.4.2.1 UDP Application

Average Throughput Fig. 6.3 shows the average throughput that the communicating vehicles experience for different eNB densities. We observe that, for the low source rate scenario (i.e., $R = 1$ Mbps), both LTE and mmWave systems deliver comparable values of throughput, which is almost equal to the full UDP rate offered by the application layer. Conversely, higher-rate applications (i.e., $R = 11$ Mbps and $R = 224$ Mbps) are not well supported by LTE connections which are constrained by the limited capacity of the low-bandwidth physical channel. The mmWave spectrum, in turn, offers orders of magnitude higher data rates than lower frequencies even in case of congested channels (110 Mbps vs. 7 Mbps for $\lambda_b = 80$ eNB/km^2), thereby satisfying the requirements of most categories of vehicular services. Moreover, we see that the throughput generally increases with the eNB density, as a consequence of stronger channels. The effect of densification is particularly evident for mmWave networks (i.e., the throughput increases by more than 50% from 4 to 80 eNB/km^2 for $R = 224$ Mbps), since the endpoints are progressively closer, thus guaranteeing improved signal quality and higher received power. On the other hand, densification delivers negligible improvements for the LTE case due to the more serious impact of interference in case of omnidirectional communications. We finally highlight that, for low-rate applications, the UDP traffic injected in the system is sufficiently well handled by the buffer, with no overflow, also considering sparsely deployed networks.

Average Latency In Fig. 6.4 we measure the average communication latency as a function of λ_b for different UDP application rates. We observe that, for $R = 1$ Mbps, both the LTE and mmWave overlays guarantee very low latency (i.e., below 20 ms) since the MAC buffers are empty most of the time. For $R = 11$ Mbps, although the two technologies were proven to offer comparable average throughput (7 Mbps vs. 9 Mbps, respectively, for $\lambda_b = 40$ eNB/km^2), mmWave systems guarantee 5 times lower latency than legacy systems, which cannot ensure time critical message dissemination in case of highly saturated channels. For higher application rates, the end-to-end latency increases uncontrollably in all investigated configurations as a consequence of more populated MAC queues, although the overall average latency for the mmWave deployment (i.e., around 150 ms for $\lambda_b = 40$ eNB/km^2) is still more than 50% lower than its LTE counterpart. Additionally, Fig. 6.4 illustrates that increasing the eNB density in mmWave scenarios has beneficial effects in terms of latency reduction (as a results of densification from 4 to 80 eNB/km^2 for $R = 224$ Mbps) as compared to a reduction of only 1% in case of LTE connections.

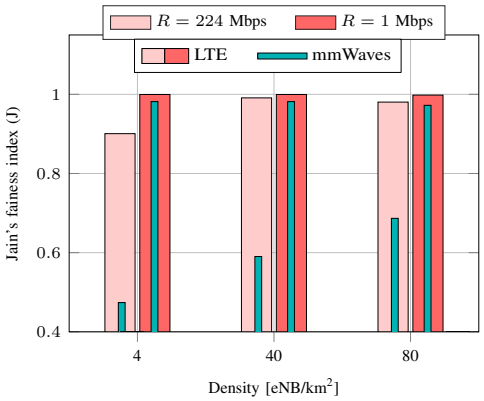


Fig. 6.6: Jain's index of the UDP throughput for different values of the eNB density and the UDP application rate.

Total Throughput (UMi vs. RMa) In Fig. 6.5 we plot the total UDP throughput as a function of the eNB density for both rural (RMa) and urban (UMi) scenarios and considering mmWave systems. Considering highly saturated channels (i.e., $R = 224$ Mbps), RMa generally guarantees throughput improvements with respect to UMi in case of sparse networks (i.e., $\lambda_b = 4$ eNB/km^2), since free-space propagation generally results in reduced outage probability. On the other hand, for dense and extremely dense deployments, the gain progressively reduces with λ_b, because of the increasing impact of the interference from the surrounding cells. In fact, while in the rural environment the propagating signals attenuate over distance following the square power law, i.e., Friis' law, the waveguide effect resulting from the more likely signal reflections and scattering in dense urban canyons generally result in reduced attenuation. Moreover, the presence of blockages in the UMi scenario may actually reduce the impact of the interference from neighboring eNBs when the obstructions block the interfering signals.Considering non-congested scenarios (i.e., $R = 1$ Mbps) instead, Fig. 6.5 proves that the experienced throughput becomes independent of the eNB density and the propagation environment as both UMi and RMa channels, regardless of their propagation characteristics, can support well the less stringent requirements of typical low source rate V2N applications.

Jain's Fairness Index In Fig. 6.6 we plot Jain's fairness index considering both LTE and mmWave scenarios,.Although fairness is not always required (e.g., some categories of applications, like those supporting time-critical operations, deserve prioritization), it still represents a major concern that should be taken into account to guarantee a minimum performance also to the cell-edge users (or, in general, to users experiencing bad channel conditions). We observe that, for LTE systems, Jain's index is very close to 1 for all density configurations, indicating that (i) cell-edge vehicles experience a throughput comparable to that of other vehicles in the cell regardless of the source application rate, and (ii) densification has a negligible impact on the overall network performance. Conversely, mmWave deployments are generally not compatible with fairness. In particular, the effect of a highly saturated network (i.e., $R = 224$ Mbps) makes Jain's index fall by an impressive 45% (for $\lambda_b = 40$ eNB/km^2) compared to LTE propagation, as a result of the increased variability of the mmWave channel due to scattering and reflection from nearby buildings, vehicles and terrain surfaces. However, such effect is partially mitigated considering denser deployments, i.e., as the probability of path loss outage decreases: in this case, the system is able to increase the coverage of cell-edge users, i.e., the most resource-constrained network entities, and consequently, provide more uniform quality of service throughout the network (for example, J increases by more than 30% when going from 4 to

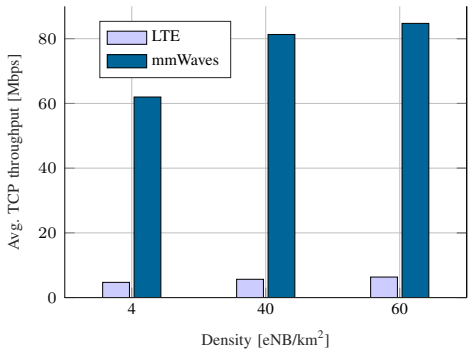


Fig. 6.7: Average TCP throughput for different values of the eNB density. The performance of the LTE and mmWave technologies is compared.

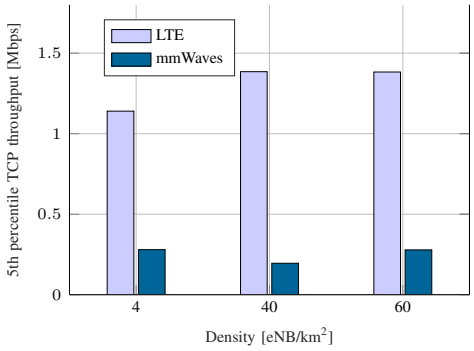


Fig. 6.8: The 5th percentile TCP throughput for different values of the eNB density. The performance of the LTE and mmWave technologies is compared.

80 eNB/km^2).

6.4.2.2 TCP Application

Average Throughput In Fig. 6.7 we plot the average TCP throughput that the communicating vehicles experience for different eNB densities and considering both LTE and mmWave technologies. To reduce the simulation complexity, we limit this analysis to the case in which $\lambda_b = 60$ eNB/km^2. We observe that, while UDP was proven unable to support high-rate applications (UDP connections delivered orders of magnitude lower throughput than the UDP source rates offered at the application layer), TCP's congestion control mechanisms regulate the source traffic and prevent network congestion, thereby guaranteeing very good throughput performance. Moreover, Fig. 6.7 shows that the average throughput is more than 15 times higher when mmWave communications are used. This is due to the fact that (i) mmWave systems leverage a much larger bandwidth than LTE systems, and (ii) the aggressive nature of the `HighSpeed` TCP congestion control protocol favors large bandwidth connections. Finally, we see that the throughput generally decreases for sparsely deployed networks (e.g., $\lambda_b = 4$ eNB/km^2) due to frequent NLOS propagation. In fact, NLOS regimes result in very high levels of queuing and buffering which dramatically increase the latency and lead to throughput degradation, a problem that is usually referred to as *bufferbloat* [56].

5th Percentile Throughput Fig. 6.8 represents the 5th percentile TCP throughput. First, we observe that LTE eNBs offer higher throughput to cell-edge vehicles than mmWave eNBs. In this region, most vehicles are in NLOS and, unlike sub-6 GHz propagation, the challenging communication characteristics of high-frequency channels might result in outage to the serving cell. Moreover, as edge vehicles are power-limited, they are unable to fully exploit the potential of the increased spectrum availability at mmWaves [10]. Densification also shows negligible effects in terms of cell-edge throughput improvement. Second, Fig. 6.8 shows that, for LTE deployments, the mutual interference from omnidirectional eNBs eventually has an impact on the cell-edge throughput, which decreases for increasing values of λ_b. Similarly, we see that, although the directional nature of mmWave systems guarantees reduced interference, there are some special cases where interference is not negligible, i.e., when $\lambda_b > 45$ eNB/km^2. Third, while for LTE the 5th percentile rate reported in Fig. 6.8 compares similarly to the average values measured in Fig. 6.7, mmWave systems alone cannot provide uniform capacity, with cell-edge users suffering significantly. In particular, the 5th percentile

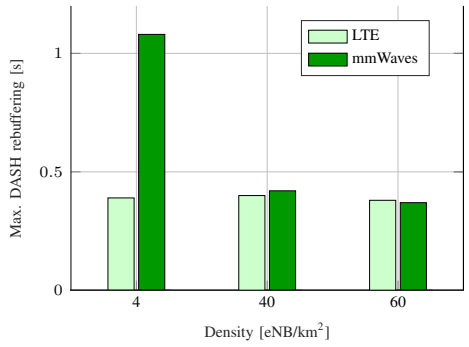


Fig. 6.9: Maximum DASH rebuffering for different values of the eNB density. The performance of the LTE and mmWave technologies is compared.

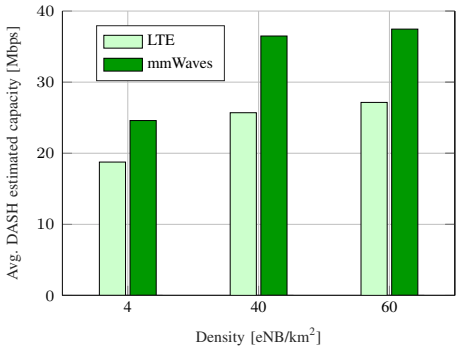


Fig. 6.10: Estimated DASH capacity for different values of the eNB density. The performance of the LTE and mmWave technologies is compared.

throughput experiences a dramatic 60 fold decrease (from around 80 Mbps to only 1.3 Mbps considering $\lambda_b = 40$ eNB/km^2) with respect to average conditions, demonstrating a significant limitation of mmWaves under NLOS propagation.

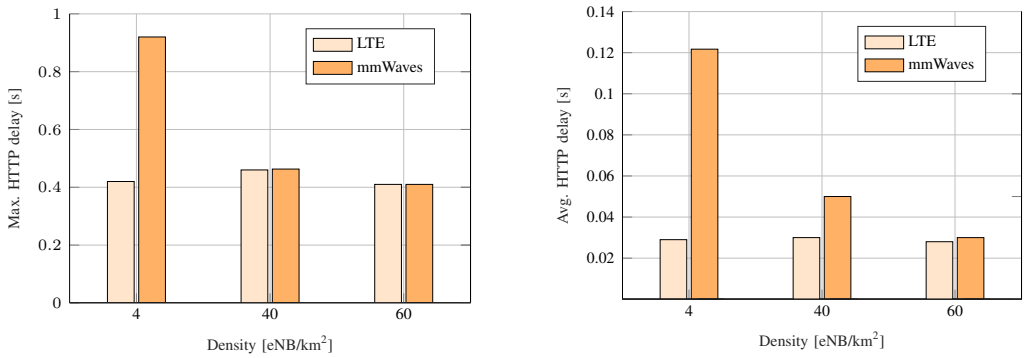
6.4.2.3 DASH Application

Max. Rebuffering Time. In Fig. 6.9 we show the maximum rebuffering time for different values of the eNB density and considering LTE and mmWave propagation. A buffering event is characterized as the video player stopping the video playback while waiting for the buffer to be sufficiently replenished. The poor streaming experience caused by long rebuffering time has become a major threat to the video service ecosystem, so that achieving optimal QoE of end viewers has been the central goal of modern video delivery services [253]. We observe that, for sparsely deployed networks, e.g., $\lambda_b < 40$ eNB/km^2, LTE guarantees about 2.5 times lower rebuffering time than operating at mmWaves. The reason is that large drops in rate, which are likely to be common in NLOS, can result in significantly populated queues, which dramatically increase the latency of the streaming service. Densification, in turn, increases the LOS probability and avoids the presence of coverage holes, thereby making the LTE and mmWave radio solutions roughly comparable in terms of rebuffering time (for $\lambda_b > 60$ eNB/km^2, mmWave video streaming is even preferable over LTE video streaming).

Estimated Capacity. Fig. 6.10 plots the average estimated capacity, i.e., the average bit-rate of all video segments requested by a user according to the DASH client predictions (it is therefore an indication of how well the DASH client is able to adapt to the network conditions and guarantee good QoE). As intuitively expected, the very large bandwidth available to the mmWave systems (50 times larger than in LTE) ensures much higher capacity than operating at sub-6 GHz frequencies (around 30% increase when $\lambda_b = 60$ eNB/km^2).

6.4.2.4 HTTP Application

HTTP Delay. In Fig. 6.11a and Fig. 6.11b we plot the maximum and average HTTP delay, respectively, i.e., the average delay experienced by the UEs to retrieve a complete webpage, which consists of multiple message exchanges between the HTTP server and the HTTP client. We observe that LTE systems guarantee improved delay performance with respect to their mmWave counterparts for all



(a) Maximum HTTP delay.

(b) Average HTTP delay.

Fig. 6.11: Maximum and average HTTP delay for different values of the eNB density. The performance of the LTE and mmWave technologies is compared.

the investigated configurations. In fact, HTTP-like applications do not generally congest the network and the HTTP traffic, which typically consists of objects of limited size, can be easily handled by low-capacity LTE connections. It hence becomes counterproductive to establish high-bandwidth mmWave communications which, unlike sub-6 GHz transmissions, suffer from very unstable propagation which may, in the long term, lead to performance degradation (especially when low values of λ_b are considered).

6.4.3 End-to-End Performance Evaluation: Impact of Mobility

In this section we evaluate the performance of the mmWave technology in a dynamic environment in which vehicles move at a speed that depends on the interaction with the road configuration (e.g., traffic lights or road intersections). UDP traffic is injected at the application layer, and we consider high- and low-traffic rates, i.e, $R = 224$ Mbps and $R = 28$ Mbps, respectively (we already showed in the previous section that lower traffic rates, i.e., $R = 11$ Mbps and $R = 1$ Mbps, are easily supported, as long as mmWave connections are established). Both urban and highway scenarios are simulated (to consider different mobility regimes), and we vary both the density of vehicles and the eNB inter-site distance (ISD_{eNB}, to switch from sparsely to densely deployed networks. In particular, we are interested in characterizing the impact of handover requests as well as of the mmWave channel variability on the end-to-end communication performance.

For the urban case, the map is composed of a number of horizontal and vertical streets (each of which has two lanes per direction of width $w_u = 3.75$ m), to represent a typical Manhattan scenario (as the one depicted in Fig. 6.12). eNBs are deployed on top of buildings (the white rectangles in Fig. 6.12) with $\text{ISD}_{\text{eNB}} \in \{65, 120, 340\}$ m that depends on the size of the building. Vehicles are dropped according to a PPP of density λ_{V}, with $\lambda_{\text{V}} \in \{24, 333\}$ veh/km^2 to consider low- and high-traffic densities, respectively [195]. In order to consider realistic mobility models and representative speed traces, we simulate the mobility of vehicles using the `randomTrip` mobility model of Simulation of Urban MObility (SUMO) [254], a powerful, open-source traffic generator that supports the modeling of intermodal traffic systems including road vehicles and structures, public transports and pedestrians. Mobility traces are finally imported to ns-3 through the `ns2MobilityHelper` class.

For the highway case, we consider 3 parallel lanes per direction of width $w_h = 4$ m (as depicted in Fig. 6.13). eNBs are located along the upper side of the road with $\text{ISD}_{\text{eNB}} \in \{50, 100, 340, 1000\}$ m. Vehicles move at constant speed $v = 130$ km/h and are dropped according to a PPP of density λ_{V}, with $\lambda_{\text{V}} \in \{500, 3000\}$ veh/h/dir (i.e., $\lambda_{\text{V}} \in \{4, 24\}$ veh/km/dir) to consider low- and high-traffic

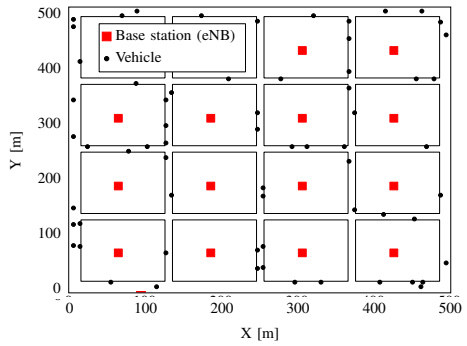


Fig. 6.12: Example of urban simulation scenario in which 16 non-overlapping buildings of 105 × 105 m are deployed. 16 eNBs are placed on top of buildings, leading to an ISD of 120 m, while $\lambda_V = 333$ veh/km^2 (i.e., 67 vehicles in the considered area of 450 × 450 m).

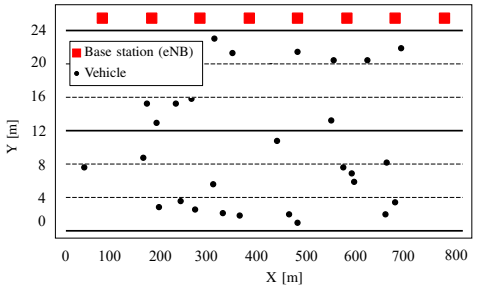


Fig. 6.13: Example of highway simulation scenario. 8 eNBs are deployed, leading to an ISD of 100 m, while $\lambda_V = 3000$ veh/h/dir (i.e., 18 veh/dir in the considered road length of 800 m).

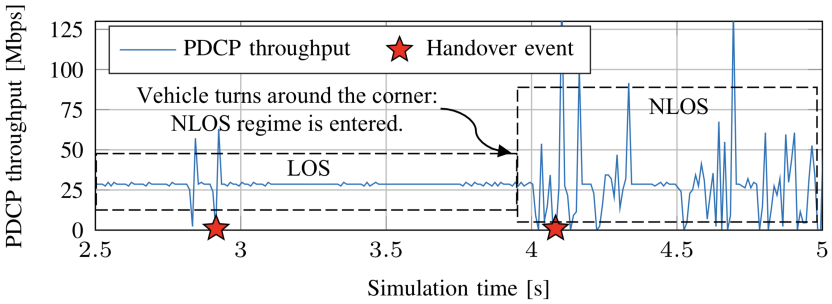


Fig. 6.14: Evolution, for a specific simulation, of the PDCP throughput in a mmWave urban scenario, with $\text{ISD}_{\text{eNB}} = 65$ m, $\lambda_{\text{V}} = 333$ veh/km^2, $R = 28$ Mbps.

densities, respectively [195].

6.4.3.1 Urban Scenario

In this subsection we evaluate the impact of mobility in an urban scenario. As an example, in Fig. 6.14 we plot the evolution of the PDCP throughput over time, with $\text{ISD}_{\text{eNB}} = 65$ m and $\lambda_{\text{V}} = 333$ veh/km^2. In the first 4 seconds, the vehicle is in LOS and experiences the maximum achievable rate offered at the application layer (i.e., $R = 28$ Mbps). In particular, the handover that is triggered at time $t = 3$ s helps maintain connectivity when the vehicle transitions from the coverage of one eNB to the coverage of another eNB and requests reassociation [123]. At time $t = 4$ s, the throughput starts decreasing and fluctuating as the vehicles enter a NLOS regime.[4] In these circumstances, not even a handover (that is triggered at time $t = 4.1$ s) can help stabilize the rate.

Average Throughput and Latency. Figs. 6.15 and 6.16 show the average throughput and latency, respectively, that the vehicles experience for different eNB ISDs and vehicular traffic densities. We observe that more crowded scenarios (i.e., $\lambda_{\text{V}} = 333$ veh/km^2) severely deteriorate the

[4]Rapid fluctuations of the throughput in NLOS are due to variations of the large/small scale fading parameters of the channel and mainly to the Doppler effect experienced by the moving vehicle.

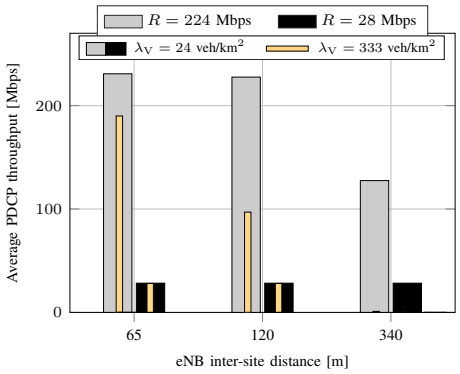


Fig. 6.15: Average PDCP throughput for different values of the eNB ISD and the UDP application rate in a mmWave urban scenario. Narrow (wide) bars refer to different values of λ_V.

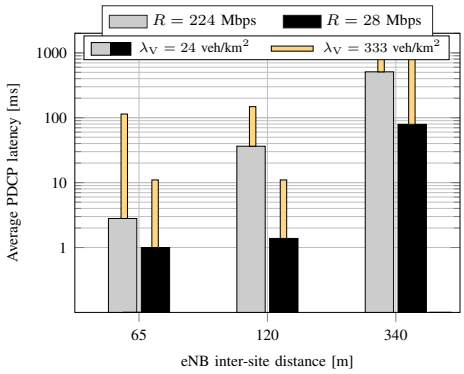


Fig. 6.16: Average PDCP latency for different values of the eNB ISD and the UDP application rate in a mmWave urban scenario. Narrow (wide) bars refer to different values of λ_V.

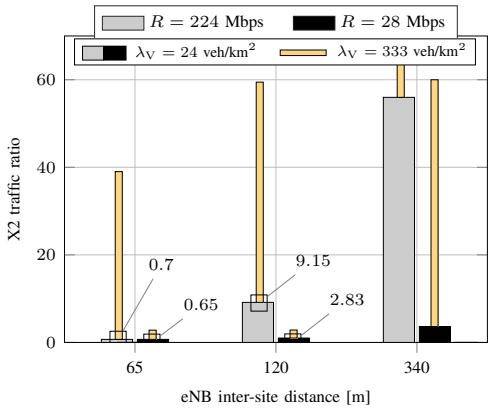


Fig. 6.17: X2 traffic ratio for different values of the eNB ISD and the UDP application rate in a mmWave urban scenario. Narrow (wide) bars refer to different values of λ_V.

throughput (-60% with $\text{ISD}_{\text{eNB}} = 120$ m and $R = 224$ Mbps) compared to less congested scenarios (i.e., $\lambda_V = 24$ veh/km^2) since each eNB serves more vehicles and can handle traffic requests less efficiently. Such negative effect can be partially alleviated by densifying the network and/or reducing the UDP traffic injected at the application layer. The potential of densification is particularly exemplified by the fact that, for $\lambda_V = 333$ veh/km^2, the throughput drops to almost zero (and the latency increases above 900 ms) when eNBs are too far (i.e., $\text{ISD}_{\text{eNB}} > 300$ m), as a consequence of lower received power. In general, the throughput and latency performance for a highly mobile urban scenario (Figs. 6.15 and 6.16) is significantly worse than for a static deployment (Figs. 6.3 and 6.4). In fact, large drops in rate, which are likely to be common at high speeds in urban environments due to frequent LOS-NLOS transitions, can result in very high levels of queuing and buffering, dramatically increasing the latency and the packet loss. Moreover, frequent handover events additionally put a strain on already congested links, as we will describe later.

X2 traffic ratio Fig. 6.17 plots the ratio between the average traffic load on X2 links connecting the different eNBs and the average throughput, varying the ISD and the vehicle density. It is an indication of the amount of traffic that is forwarded between two eNBs during handovers, given the rate at which data packets are sent over the radio access network. We see that the load on the X2

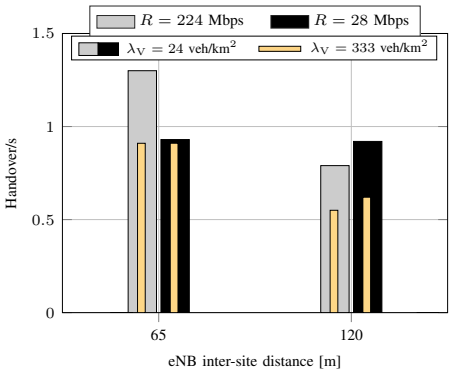


Fig. 6.18: Average number of handover events per second for different values of the eNB ISD and the UDP application rate in a mmWave urban scenario. Narrow (wide) bars refer to different values of λ_V.

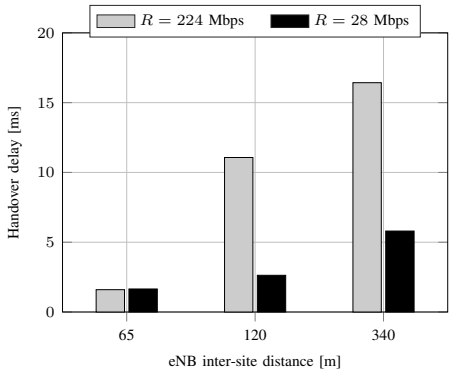


Fig. 6.19: Average handover delay for different values of the eNB ISD and the UDP application rate in a mmWave urban scenario. $\lambda_V = 24$ veh/km^2.

links increases with the ISD, thus considering more sparsely deployed networks, and with λ_V, since more populated scenarios may overload the available resources and introduce congestion in the X2 interface [110]. In particular, we observe that, when $\text{ISD}_{\text{eNB}} > 100$ m, the X2 traffic ratio becomes greater than 1, i.e., the X2 links must be dimensioned to offer much more than the target throughput for each mmWave eNB, thereby calling for the design of denser network deployments to satisfy handover requests more efficiently. For the $\text{ISD}_{\text{eNB}} = 65$ m case, for example, the ratio is around 0.7 for both $\lambda_V = 24$ and 333 veh/km^2, meaning that the X2 links need to offer just 70% of the target application rate. Finally, as expected, for all investigated configurations, the ratio is higher for higher UDP traffic rates, since there are more packets buffered at the RLC layer that must be forwarded.

Impact of handover Fig. 6.18 illustrates the average number of handover events per second vs. the ISD and λ_V. We see that this number is much higher when considering more densely deployed networks, thus reducing the eNB ISD, since the moving vehicle has more chances to change its current cell. Moreover, the number of handovers decreases with the increase of the vehicular traffic density (-30% for $\text{ISD}_{\text{eNB}} = 65$ m and $R = 224$ Mbps) since vehicles generally move at a lower speed in crowded scenarios and therefore have fewer opportunities to update the serving cell per unit of time. Despite the higher average number of handovers in dense networks, the average time required to complete each handover, i.e., from the time it is initiated to the time the packet is forwarded to the target eNB, decreases when reducing the ISD, as reported in Fig. 6.19. This is due to the fact that X2 links are less congested (as it appears also from Fig. 6.17) and the traffic can be forwarded to the target eNB as soon as the handover is requested, with limited queuing time in the RLC buffers. Moreover, according to Fig. 6.19, the average handover time decreases when considering $R = 28$ Mbps. In fact, with this setup, the RLC buffer is empty most of the time and, when a handover is triggered, very few UDP packets need to be forwarded to the destination mmWave eNB, thus limiting the impact of latency.

6.4.3.2 Highway Scenario

In this subsection we evaluate the impact of mobility in a highway scenario. As an example, in Fig. 6.20 we plot the evolution of the PDCP throughput over time, with $\text{ISD}_{\text{eNB}} = 100$ m and $\lambda_V = 500$ veh/h/dir. In this scenario, the endpoints always communicate in LOS and the throughput

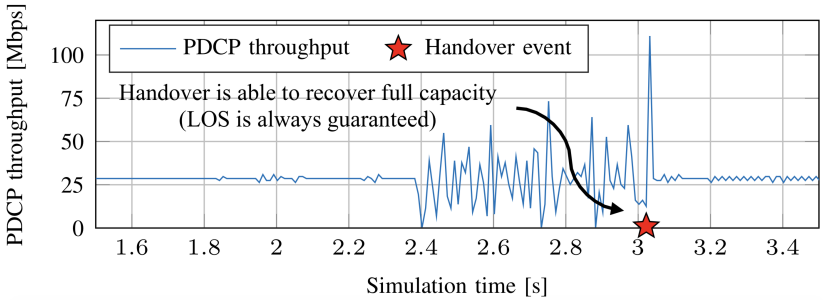


Fig. 6.20: Evolution, for a specific simulation, of the PDCP throughput in a mmWave highway scenario, with $\text{ISD}_{\text{eNB}} = 100$ m, $\lambda_{\text{V}} = 500$ veh/h/dir, $R = 28$ Mbps.

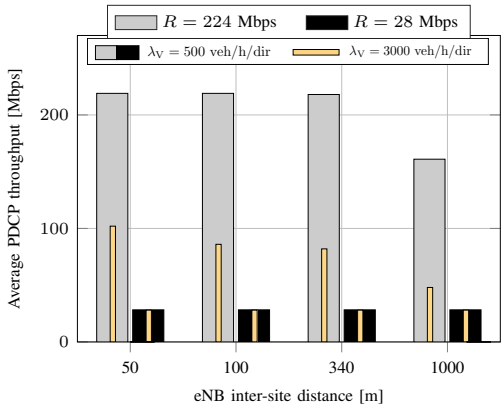


Fig. 6.21: Average PDCP throughput for different values of the eNB ISD and the UDP application rate in a mmWave highway scenario. Narrow (wide) bars refer to different values of λ_{V}.

is equal to the rate offered at the application layer (i.e., $R = 28$ Mbps). At time $t = 2.4$ s, the throughput starts decreasing as a consequence of the vehicle moving away from its serving eNB even though, at time $t = 3$, the vehicle can successfully hand over to a new eNB providing better communication performance and recover full capacity. Notice that the throughput spike observed when the handover is triggered is due to the RLC buffer, which accumulates a number of packets during the reduced rate regime, and transmits the entire burst when the high-capacity connection is restored [56].

In Fig. 6.21 we plot the average PDCP throughput experienced by the moving vehicles as a function of the ISD, the application rate R and the vehicular traffic density λ_{V}. In general, the effect of eNB densification is not very significant for the highway scenario (compared to the urban results in Sec. 6.4.3.1) since LOS propagation results in low outage probability even in sparsely deployed networks. The throughput loss between the $\text{ISD}_{\text{eNB}} = 50$ m and the $\text{ISD}_{\text{eNB}} = 340$ m configurations is as low as 0.4%, and still it is limited to 30% when considering an ISD of 1000 m. In turn, the combined effect of a highly populated network and highly saturated channel (i.e., $\lambda_{\text{V}} = 3000$ veh/h/dir and $R = 224$ Mbps, respectively) makes the number of received packets decline severely, thereby leading to a throughput decrease of more than 50% with respect to $\lambda_{\text{V}} = 500$ veh/h/dir, even when $\text{ISD}_{\text{eNB}} = 50$ m. For the low source rate scenario (i.e., $R = 28$ Mbps), instead, all investigated configurations deliver comparable values of the throughput, which is limited by the full UDP rate offered at the application layer: this confirms the conclusions we drew in Sec. 6.4.2 for the static deployment.

6.5 Millimeter Waves in V2V: Performance Evaluation

In this section we conduct extensive simulations to compare the performance of the mmWave and IEEE 802.11p RATs in realistic V2V environments. In detail, Sec. 6.5.1 describes our system model and simulation parameters, Sec. 6.5.2 present our main findings and simulation results, and Sec. 6.5.3 presents the limitations of existing channel models for V2V propagation and discusses how to extend existing simulators, e.g., ns-3, to conduct end-to-end simulation analyses.

6.5.1 Evaluation Methodology

The 3GPP has recently developed some functionalities to provide enhancements of the traditional communication standards for V2Vcommunications, starting with the 3GPP Release 16 SI on NR-V2X [255], as introduced in Sec. 1.4.1. These developments include sidelink design and enhancements of Release 15 NR UL/DL for advanced V2V services in both sub-6 GHz and mmWave bands. From a PHY-layer perspective, the channel has been designed to be optimized for uses under 52.6 GHz and with the potential to be used for above 52.6 GHz (including therefore mmWave transmissions) [256]. The 3GPP also specifies how to model the V2V channel at mmWaves [31]. In particular, distinctions between environmental and vehicular blockages, as well as between urban and highway propagation scenarios, have been proposed. However, before the model can be adopted as a feasible solution to simulate V2V propagation, it is fundamental to evaluate its performance in realistic vehicular scenarios, a research issue that, to date, has not yet been thoroughly addressed.

Unlike the V2N case, there currently are no open source simulators capable of modeling the mmWave channel in V2V scenarios as well as TCP/IP protocol stack and mobility. Existing ns-3 mmWave modules [156], for instance, are based on the NR cellular standard, which however is infrastructure-based and do not characterize infrastructure-less (i.e., ad hoc) vehicular scenarios. Alternatively, Veins [257], CARLA [258], or VSimRTI [259], among the most common open source simulators for vehicular network that support development, training, and validation of autonomous urban driving systems, do not implement mmWave communications and cannot support future vehicular applications. For this reason, it is currently not possible to conduct end-to-end simulation studies in a V2V scenario (even though we will see in Sec. 6.5.3 how ns-3 can be extended to simulate mmWaves V2V operations in a end-to-end fashion following the evaluation methodology agreed in [31]). On the contrary, in this study we evaluate, using MATLAB, the performance of the network at a link level, thereby investigating the impact of PHY- and MAC-layer parameters, e.g., the channel characterization, the antenna array size, and the vehicular traffic density, in terms of Shannon capacity, received power and outage probability.

6.5.1.1 Channel Characterization

The first step towards proper vehicular protocol design is a deep understanding of the propagation model. In V2V systems, due to the relatively low elevation of the vehicle antennas, we reasonably expect that other vehicles as well as environmental objects will act as obstacles to the propagation of the signals. The path loss is therefore modeled according to the following three states [31]:

1. *Line of Sight (LOS)*, i.e., the propagation path is not blocked by vehicles/environmental objects.
2. *Vehicle Non Line of Sight (NLOSv)*, i.e., the LOS is blocked by dynamic blockages (vehicles).
3. *Non Line of Sight (NLOS)*, i.e., the LOS is blocked by environmental blockages (buildings).

Table 6.4: V2V link-level simulation parameters.

Parameter	Value	Description
W_ℓ	$\{3.5, 4\}$ m	{Urban, highway} lane width
N_ℓ	$\{2, 3\}$	{Urban, highway} number of lanes
l_v	{5, 13} m	{Type 2, Type 3} vehicle length
w_v	{2, 2.6} m	{Type 2, Type 3} vehicle width
h_v	{1.6, 3} m	{Type 2, Type 3} vehicle height
h_a	{1.6, 3} m	{Type 2, Type 3} antenna height
IEEE 802.11p ℓ_a	10 cm	IEEE 802.11p antenna length
IEEE 802.11p B	75 MHz	IEEE 802.11p total bandwidth
IEEE 802.11p f_c	5.9 GHz	IEEE 802.11p carrier frequency
mmWave B	1 GHz	mmWave total bandwidth
mmWave f_c	63 GHz	mmWave carrier frequency
P_{TX}	21 dBm	Transmission power
NF	13 dB	Noise figure
M_V	$\{1, 4, 32, 64\}$	Vehicle antenna array size
Γ_{th}	-5 dB	SNR threshold

The LOS and NLOSv probabilities, i.e., P_{LOS} and P_{NLOSv}, are defined in [31, Table 6.2-1] and reported in [220]. Although the model distinguishes between urban and highway scenarios (respectively denoted with superscripts u and h throughout this section), it does not differentiate between different densities of vehicular traffic. Moreover, the determination of the NLOS state is deterministic, i.e., it is based on purely geometric considerations which evaluate whether the V2V link is blocked or not by buildings, therefore a closed-form expression for the NLOS probability, i.e., P_{NLOS}, is not currently provided. As we will numerically demonstrate in Sec. 6.5.2, such assumptions reduce the accuracy of the analysis and might result in misleading conclusions. In this study, we therefore consider an extension of the path loss probability equations proposed in [31] based on [195], as summarized in [220]. The model in [195] (i) characterizes low, medium, and high densities of the vehicular traffic in both urban and highway scenarios, and (ii) introduces a probabilistic model for the NLOS probability as a function of the inter-vehicle distance (the longer the link, the more likely to intersect one or more blockages). In order to have realistic mobility traces for the vehicles in the considered environments, the authors in [195] have used SUMO [254], an open-source road traffic simulator designed to handle and model the traffic of large road networks. The LOS/NLOSv/NLOS classification is finally provided by GEMV2, a freely available vehicular propagation modeling software which performs geometry-based blockage analyses based on the outlines of buildings and vehicles.

As soon as the communication states have been identified, the path loss follows a *dual-slope piecewise-linear model*, which is deemed suitable to represent the real propagation in a V2V scenario.

- For mmWave systems, the LOS/NLOSv/NLOS path loss is modeled as defined in [31, Table 6.2.1-1]. In case of NLOSv, in particular, the model provides the attenuation factor $\mathcal{A}_{\text{NLOSv}}$ to be summed to the LOS path loss. $\mathcal{A}_{\text{NLOSv}}$ is modeled according to a lognormal random variable with mean μ_a and standard deviation σ_a, whose values depend on the vehicle size.

- For IEEE 802.11p systems, the path loss follows the model in [260]. In case of NLOS, in particular, the path loss is increased by a factor $\mathcal{A}_{\text{NLOS}}$ according to a *knife-edge* attenuation model. For the tractability of the simulation, we assume that $\mathcal{A}$ follows a *single knife-edge* model, which considers one single vehicle obstructing the LOS (the extension of the single knife-edge obstacle case to a multiple knife-edge is not immediate and is left for future investigations).

6.5.1.2 Simulation Parameters

Evaluation Scenarios The channel characteristics are heavily influenced by the proprieties of the environment in which the vehicles are deployed, i.e., urban or highway. The parameters regarding the road configuration for both scenarios are taken from [31, Table A-1]: for the urban (highway) case, $N_\ell = 2$ ($N_\ell = 3$) lanes per directions are assumed, where each lane has width $W_\ell = 3.5$ m ($W_\ell = 4$ m).

Vehicle Characteristics Three types of vehicles are defined according [31, Sec. 6.1.2]: Type 1 and Type 2 vehicles are passenger cars with bumper and rooftop antenna position, respectively, while Type 3 vehicles are large trucks or buses. The vehicle type may potentially affect the path loss equation (since the antenna height is determined according to the type of vehicle), the loss caused by the blocking vehicle, and the radiation pattern. We consider both Type 2 and Type 3 deployments.

Vehicle Dropping Models Vehicles are dropped according to a random process, so that the distance between the rear bumper of a vehicle and the front bumper of the following vehicle in the same lane is equal to $\max\{2\text{ m}, \text{Exp}(\xi)\}$, where $\xi = v \cdot 2$ s. and v is the average speed [31, Sec. 6.1.2].

Link-Level Parameters The simulation parameters are based on the system-design considerations specified in [31, Sec. 6.1.1] and are summarized in Table 6.4. IEEE 802.11p systems operate in the legacy band, i.e., at 5.9 GHz, with a total bandwidth of 75 MHz. Antennas are supposed to be located on top of vehicles, in the middle of the roof (which was experimentally shown to be the overall optimum placement when considering omnidirectional transmissions [261]), and to be of length $\ell_a = 10$ cm. For mmWave links, the central frequency is set to 63 GHz while the total bandwidth is set to 1 GHz. The vehicles' noise figure is set to 13 dB and the transmit power is set to 21 dBm. In order to establish directional transmissions, vehicles are equipped with UPAs of M_V elements. For above 6 GHz propagation, the maximum value of M_V ranges from 4 to 64 [31, Tab. 6.1.4-12]. For completeness, in our study we also consider omnidirectional mmWave transmissions, i.e., $M_V = 1$. For the beam alignment, we assume that measurement reports are periodically exchanged (i.e., at the beginning of every slot of duration T) among the vehicles so that they can periodically identify the optimal directions for their respective beams (see Chapters 3 and 4). Such configuration is kept fixed for the whole slot, during which nodes may lose the alignment due to mobility. In case the connectivity is lost during a slot, it can only be recovered at the beginning of the subsequent slot, when the re-alignment procedure is performed again. In this respect, geographical information, e.g., obtained from Global Positioning System (GPS) signals, can be used to geometrically point the beam towards the intended receiver at any given time: the inaccuracy of such data is modeled according to a Gamma distribution with parameters $\alpha = 3.14733$ and $\beta = 0.462432$ [262].

6.5.1.3 Performance Metrics

The results are derived through a Monte Carlo approach as a function of the inter-vehicle distance d, with d varying from 2 m to 500 m. In particular, we analyze the following performance metrics.

- *Path loss probabilities* and *path loss*.
- *Shannon capacity*, the maximum data rate offered for each pair of vehicles as a function of the available bandwidth and the instantaneous channel quality. Such metric represents an *upper bound* for the performance of the vehicular nodes, as we it does not incorporate the effect of interference nor make any medium access control consideration.

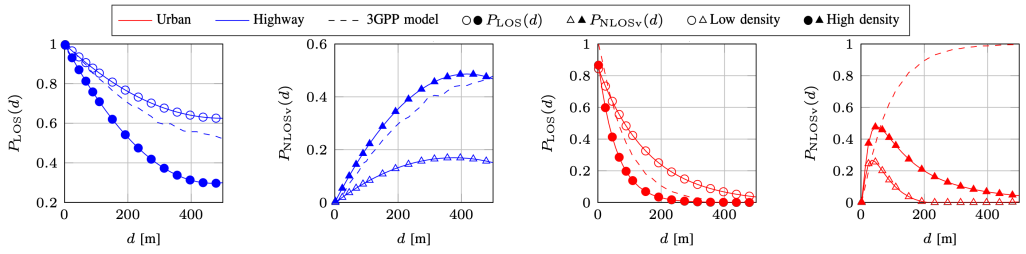
- *Outage probability*, which measures the percentage of vehicles experiencing unsuccessful packet reception. Low values of outage ensure more reliable V2V communications, a critical prerequisite for safety services requiring ubiquitous and continuous connectivity. In this study, unsuccessful reception is achieved if the SNR experienced between the transmitting and receiving vehicles is below a predefined threshold, taken to be $\Gamma_{\text{th}} = -5$ dB in our simulations.
- *Robustness*, which measures (in terms of received power) the effect of misalignment between the transmitter and the receiver. Misalignment, which can occur during lane change operations or as a consequence of the dissemination of inaccurate vehicle position information, may have in fact a very detrimental impact on the performance of some V2V applications, as sensors may provide skewed or delayed readings and vehicles may lose connectivity. In our scenario, the degree δ_m of misalignment is distance-dependent and is given by $\delta_m = \arctan(W_\ell N_\ell / d)$, where $W_\ell N_\ell$ is the width of one carriageway lane and d is the inter-vehicle distance.

6.5.2 Link-Level Performance Evaluation

In this section we provide some numerical results to compare the performance of IEEE 802.11p and mmWave technologies in V2V scenarios. The impact of the propagation scenario, the vehicle type, the vehicle density, the antenna array configuration, the inter-vehicular distance, and the misalignment probability is evaluated.

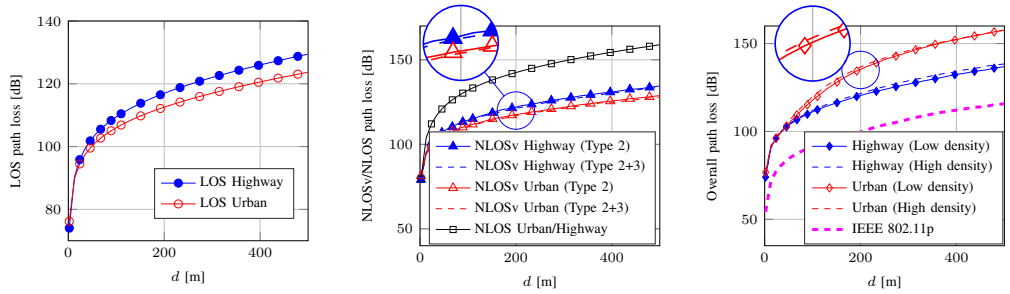
6.5.2.1 Path Loss Results

In Fig. 6.22 we plot the mmWave LOS and NLOSv probabilities vs. d considering both the 3GPP model [31] (which does not distinguish between low- and high-traffic densities) and the extended model [195], according to the description in Sec. 6.5.1.1. As foreseen, the LOS probability is significantly higher in case of highway deployments than in urban scenarios since the signal usually propagates in free space. Moreover, it is clear that the impact of different density regimes is not negligible (the gap is particularly evident for large distances). While, for the NLOSv case, the 3GPP model behaves as in a high-density scenario, for the LOS case it operates as in a low-density scenario, thereby setting a lower bound to the path loss. For the urban case, P_{NLOSv} peaks at around $d = 50$ m and then starts decreasing for larger values of d when the model in [195] is considered. In fact, although the probability of both dynamic and static blockages potentially obstructing the propagation path between the endpoints increases with d, [195] assumes that the channel condition is categorized as NLOS when the line of sight is blocked by both vehicles and buildings. The 3GPP model [31], instead, does not make this distinction and shows a monotonically increasing trend.



(a) P_{LOS} for highway scenario. (b) P_{NLOSv} for highway scenario. (c) P_{LOS} for urban scenario. (d) P_{NLOSv} for urban scenario.

Fig. 6.22: P_{LOS} and P_{NLOSv} vs. d for urban and highway scenarios for different traffic densities. The 3GPP [31] and the extended [195] models are compared.



(a) LOS path loss. Type 2 vehicles are deployed.

(b) NLOSv and NLOS path loss.

(c) Overall path loss. Type 2 vehicles are deployed.

Fig. 6.23: Path loss vs. d for urban and highway scenarios for different traffic densities, deployment options and RAT. The extended model in [195] is considered for mmWave transmissions.

The following results are derived considering the extended model in [195]. The 3GPP model, in fact, does not define a closed-form expression for the NLOS probability and prevents a complete stochastic analysis for the path loss, which has to be based on geometric simulations instead. In Fig. 6.23 we plot the mmWave V2V path loss as a function of d. Different densities of vehicular traffic and vehicle deployment options are considered. We see that, for the LOS case (Fig. 6.23a), we register better propagation (i.e., lower path loss) in urban rather than highway scenario (i.e., around 5 dB at 200 m). In fact, while in the highway environment the propagating signals attenuate over distance following Friis' law, in the urban environment the observed path loss is significantly lower, indicating a waveguide effect resulting from the more likely reflections from walls of static blockages in street canyons. From Fig. 6.23b, we observe that, for the NLOSv case, the path loss slightly decreases when deploying both Type 2 and Type 3 vehicles, i.e., when tall vehicles (e.g., trucks) are deployed. In fact, although Type 3 blockage implies higher attenuation, larger vehicle heights may guarantee higher LOS probability when the obstacle is small. Moreover, we see that the NLOS path loss is generally more than 20 dB higher than its NLOSv counterpart, demonstrating the much stronger impact of static/environmental blockages like buildings or vegetation, compared to dynamic obstructions like pedestrians and cars, on the received signal strength. Finally, Fig. 6.23c measures the overall path loss as a function of d. We observe that the urban path loss is significantly higher than its highway counterpart (although the waveguide effect caused by the more likely signal reflections and scattering in street canyons generally results in reduced attenuation) due to the much higher probability of blockage intersection in contrast to free-space propagation. Furthermore, the higher the vehicle density, the more probable the NLOSv state and, therefore, the larger the overall sidelink path loss. Fig. 6.23c also demonstrates that the mmWave path loss is significantly higher than for IEEE 802.11p transmissions in all investigated scenarios. The reason is that, unlike sub-6 GHz frequencies, mmWaves have increased reflectivity, poor diffraction and penetration capabilities in NLOS situations, and therefore are affected by significant attenuation. However, the effect of such properties is small for short distances (i.e., up to a few tens of meters), which therefore represent a suitable range for mmWave links in vehicular scenarios.

6.5.2.2 Shannon Capacity Results

In Fig. 6.24, we compare the average Shannon data rate of both the IEEE 802.11p and the mmWave technologies, for different antenna configurations (including omnidirectional mmWave transmissions, i.e., $M_V = 1$). We observe that the very large bandwidth available to the mmWave systems (5 times larger than in IEEE 802.11p) ensures much higher throughput than operating at legacy frequencies (up to two orders of magnitude more in short range). This performance can be further magnified by configuring very directional transmissions. In fact, there exists a strong correlation

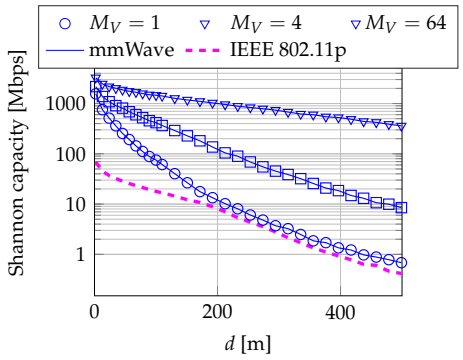


Fig. 6.24: Average Shannon capacity *vs* d for IEEE 802.11p (dashed line) and mmWave communications (solid lines), with different antenna array configurations (markers).

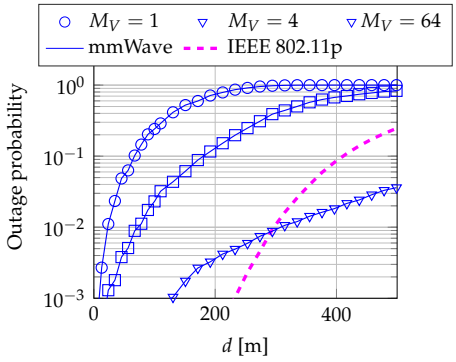


Fig. 6.25: Outage probability *vs* d for IEEE 802.11p (dashed line) and mmWave communications (solid lines), with different antenna array configurations (markers).

among beamwidth, number of antenna elements and beamforming gain: the more antenna elements in the system, the narrower the beams, the more directional the transmission, the higher the gain that can be achieved by beamforming. It should also be noted that, even implementing omnidirectional strategies at mmWaves, the connection still guarantees acceptable average bit-rate, provided that the endpoints are sufficiently close (to increase the LOS probability).

6.5.2.3 Outage Probability Results

In general, Fig. 6.25 shows that lower outage probability can be achieved when considering short-range communications and, in case of directional transmissions, when using large arrays. In the first case, the endpoints are progressively closer, thus ensuring better signal quality and stronger received power. In this region, the channel conditions are sufficiently good to ensure satisfactory signal quality (and, consequently, acceptable outage) even when considering small antenna factors or omnidirectional transmissions. In the second case, narrower beams are needed to guarantee higher gains, produced by beamforming. Moreover, we observe that IEEE 802.11p systems usually provide more reliable communications than mmWave links since they present a lower outage probability. Nevertheless, mmWave transmissions also achieve sufficient detection performance for spatially close vehicles (i.e., $d < 110$ m) employing very narrow beams (e.g., $M_V = 64$). Finally, for very large distances (i.e., $d > 300$ m), all the investigated configurations achieve unacceptable reliability values. However, mmWave communications with sharp beams (e.g., $M_V > 64$) have the potential to support unreliable long-distance inter-vehicle communications for which the IEEE 802.11p signal is basically undetectable.

6.5.2.4 Robustness Results

In Fig. 6.26, we observe that the impact of the misalignment on the communication performance depends on several factors, including d and the beamwidth. In case of very directional mmWave transmissions (e.g., $M_V = 64$), the quality of the received signal significantly decreases as a result of misaligned nodes (i.e., more than 50 dB for short distances), mainly due to the non-continuous beamtracking mechanism: after the alignment is lost, vehicles need to wait for a new tracking op-

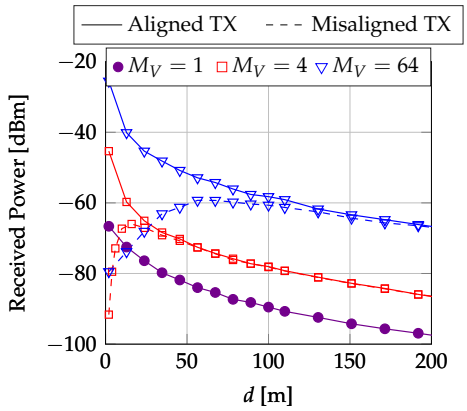


Fig. 6.26: Average received power *vs* d for aligned (solid lines) and misaligned (dashed lines) mmWave communications, with different antenna array configurations (markers).

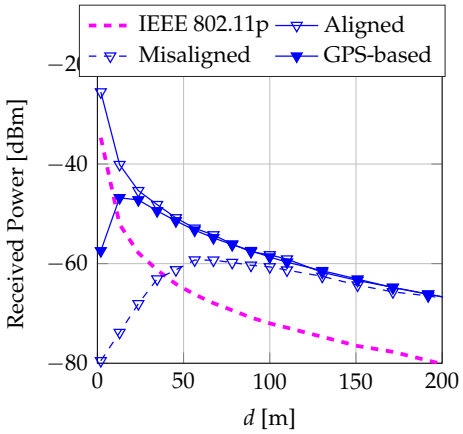


Fig. 6.27: Average received power *vs* d for IEEE 802.11p (dashed line) and directional mmWave communications (solid lines) w/o GPS information, with $M_V = 64$.

eration to be performed to recover their optimal beam configuration (see discussion in Chapter 4). Conversely, more robust alignment can be achieved when considering smaller array factors since wider beams enlarge the area in which the vehicles are within coverage. Omnidirectional strategies are not affected by misalignment (in Fig. 6.26, the solid and dashed curves for the $M_V = 1$ case overlap perfectly). In this approach, however, the well-known robustness versus data rate trade-off is exposed [54]: wider beams guarantee more robust and continuous connectivity but generally yield lower received power and transmission rates, as shown in Fig. 6.24. Moreover, from Fig. 6.26 we see that the impact of misalignment is dominant at short ranges. Indeed, the received power initially increases with d since, for larger distances, the projection of the beam's shape onto the road surface is geometrically larger, thereby increasing the maximum distance that the vehicles can cover before leaving their respective communication ranges. However, beyond a certain value of d (i.e., $d > 50$ m for $M_V = 64$ and $d > 10$ m for $M_V = 4$), beams are already sufficiently large to allow for loose alignment and the received power decreases just because of the path loss.

If sensory information (e.g., GPS coordinates) is available, it can be used to aid the configuration of the mmWave communication link and to remove the need for periodical beam tracking operations. In this regard, Fig. 6.27 reports the effect of misalignment due to inaccurate data, which makes the nodes point their beams towards improper directions. Nevertheless, such inaccuracy compromises only very short-range transmissions (i.e., $d < 20$ m for $M_V = 64$).[5] Finally, Fig. 6.27 exemplifies how the omnidirectional transmissions of IEEE 802.11p systems offer more robust and, in some circumstances (e.g., $d < 40$), more efficient V2V communications than their mmWave counterparts.

6.5.3 Towards End-to-End Performance Analysis

As introduced in Sec. 6.5.1, it is currently not possible to conduct end-to-end V2V simulations at mmWaves due to the absence of full-stack simulators capable of modeling the mmWave channel in V2V scenarios as well as TCP/IP protocol stack and mobility. However, when it comes to mmWave

[5] The accuracy of the position information may be improved by the adoption of *data fusion* strategies which combine several localization techniques, e.g., dead reckoning, cellular localization, and camera image processing, into a single solution that is more robust and precise than any individual approach [263].

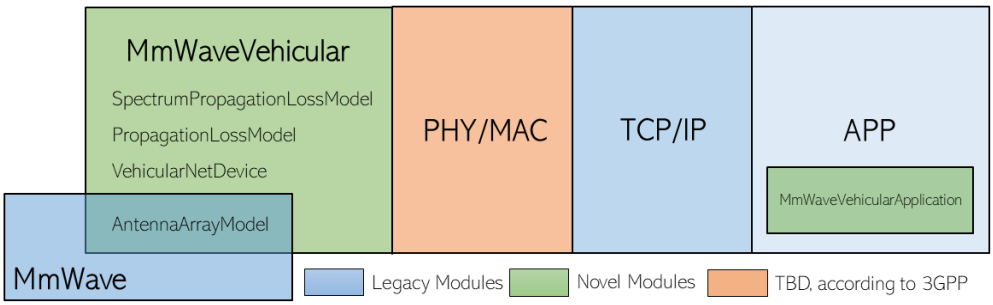


Fig. 6.28: Class diagram of the ns-3 `MmWaveVehicular` module, which is being extended to support V2V operations at mmWave frequencies as agreed in [31].

communications, it is important to consider the performance across the different layers of the protocol stack, given that unexpected interactions may arise between the characteristics of the channel and the legacy design of the higher layers of the protocol stack. To address this challenge, we recently started to develop an ns-3 module, i.e., `MmwaveVehicular` [27], for the simulation of V2V communications in the mmWave bands which is fully compliant with the 3GPP standardization efforts in this domain and which follows the evaluation methodology agreed in [31] and described in Sec. 1.4.1. ns-3 helps investigate the impact of the upper layers on the network behavior, thereby guaranteeing more accurate system-level analyses. Moreover, it enables the study of end-to-end metrics, including per user throughput, latency, application packet reception ratio. Furthermore, it can be integrated with other open source softwares, e.g., SUMO, to simulate realistic vehicle mobility on real-world topologies.

Our ns-3 module `MmwaveVehicular`, will feature, as depicted in Fig. 6.28:

- a `MmWaveVehicularNetDevice`, which models the behavior of the vehicular device and manages the operations of the different entities of the protocol stack.
- a `MmWaveVehicularSpectrumPropagationLossModel` class and a `MmWaveVehicularPropagationLossModel` class which implement the channel between two vehicular devices as specified in [31]. The channel model distinguishes between environmental and vehicular blockages, as well as between urban and highway propagation scenarios, and provides a complete fading model which characterizes spatial clusters, subpaths, angular beamspreads and the Doppler shift, which is a function of the total angular dispersion, carrier frequency and mobile velocity.
- a multi-panel multi-sector `AntennaArrayModel` class which applies beamforming on top of the channel to accurately model directional transmissions. Antenna parameters are compliant with the 3GPP design in [31].
- a `MmWaveVehicularApplication` class that provides custom models for next-generation vehicular scenarios.

Our module will allow us to run simulations to provide the first numerical validation of V2V operations at mmWave frequencies, and to understand which are the design trade-offs in this domain. Besides the performance assessment, the ns-3 module will be further developed according to future 3GPP standardization agreements for the protocol stack, e.g., following the discussion in [32].

6.6 Heterogeneous Networking: the Special Case of Attachment

The results in Sec. 6.4 and Sec. 6.5 demonstrate that the synergistic orchestration among the different radios, i.e., *heterogeneous networking*, makes it possible to complement the limitations of each type of network and deliver more flexible and resilient transmissions. For instance, we showed that sub-6 GHz technologies (i.e., LTE and IEEE 802.11p) offer low-rate connectivity but guarantee very stable, reliable and robust transmissions at short/medium distances (i.e., up to a few hundreds of meters) thanks to the intrinsic stability of the low-frequency channels and the omnidirectional transmissions. Conversely, mmWave systems support very high-throughput connections but exhibit high instability due to the severe signal propagation characteristics and the need to maintain beam alignment.

Despite some encouraging features, however, heterogeneous networking deployments lack at least one important design aspect, which is how to efficiently and fairly associate vehicles to the network. Maximum downlink received power based association, for example, typically leads to a limited number of nodes actually getting served by mmWave cells due to their much more unstable propagation characteristics compared to LTE cells. Maximum rate based association, on the other hand, tends to prioritize mmWave eNBs over legacy ones due to the much larger bandwidth available to high-frequency systems. This load disparity inevitably leads to suboptimal resource allocation, with a large number of vehicles experiencing poor date rates in overloaded cells while the resources in other lightly loaded cells can be underutilized. Following this rationale, in this section we address the issue of balancing network association requests between LTE and mmWave eNBs, avoiding the overload of transmission links. To do so, we design a novel QoS aware attachment strategy that identifies the most appropriate destination cell as a function of the vehicle's individual requirements and traffic demands.

6.6.1 Attachment Policies in Vehicular Networks

When a vehicle enters a vehicular network for the first time, it needs to establish an initial physical link connection with a cell, a procedure that is usually referred to as network attachment [52] (see Chapter 3). Traditional attachment procedures monitor the quality of the received signals, which is typically expressed in terms of SNR, and select, as a target cell, the eNB from which the maximum SNR was experienced. This procedure is described in Sec. 6.6.1.1 and represents the benchmark solution of our analysis. In this section we target tight integration of classic physical-layer performance metrics with additional network information in the upper layers. In particular, a maximum rate attachment policy, which takes data rate estimates into account, and a requirement-aware attachment policy, which biases cell selection as a function of the vehicle's traffic requirements, are proposed in Secs. 6.6.1.2 and 6.6.1.3, respectively. In the following sections, $\mathcal{N}$ is the set of vehicles (which we will refer to as Vehicular Nodes (VNs)) and $\mathcal{M}$ the set of eNBs. In particular, $\mathcal{M}_m \subseteq \mathcal{M}$ is the set of eNBs operating in the mmWave band, and $\mathcal{M}_L \subseteq \mathcal{M}$ is the set of eNBs operating in the legacy band.

6.6.1.1 Maximum SNR (MS) Policy

The Maximum SNR (MS) policy represents one of the most common techniques for performing user association: $\text{VN}_i \in \mathcal{N}$ always connects to $\text{eNB}_{j^*_{\text{MS}}(i)} \in \mathcal{M}$ (either LTE or mmWave) that provides the maximum downlink average SNR, i.e.,

$$j^*_{\text{MS}}(i) = \underset{j \in \{1,\ldots,|\mathcal{M}|\}}{\arg\max} \{\Gamma_{ij}\}, \forall i \in \{1,\ldots,|\mathcal{N}|\} \tag{6.2}$$

where Γ_{ij} is the SNR between eNB$_j$ and VN$_i$. Notice that, in an urban heterogeneous scenario, the MS policy does not guarantee that the eNB with the maximum SNR coincides with the closest one. First, LTE eNBs are generally preferred over mmWave ones due to the very low path loss experienced at below-6 GHz frequencies even at long distances. Second, the mmWave signal is much more sensitive to penetration loss than LTE links and, therefore, if the geographically closest mmWave eNB is obstructed, a further eNB in line of sight can potentially offer a better service (experiments performed for NLOS situations resulted in SNR degradation of more than 20 dB compared to LOS propagation [264]). We make the case that, although MS maximizes the SNR of vehicles, it does not properly reflect the achievable end-to-end throughput of users, thereby leading to suboptimal association decisions. This is because, even with a lower SNR, mmWave cells may potentially deliver higher data rates (due to the much larger bandwidth) compared to LTE cells. Moreover, downlink-based received signal quality criteria do not characterize well uplink scenarios where vehicles have strict battery limitations on their transmit power.

6.6.1.2 Maximum Rate (MR) Policy

MS attachment schemes can be improved by biasing cell selection with side information, e.g., network load. A Maximum Rate (MR) approach is therefore proposed: VN$_i \in \mathcal{N}$ connects to eNB$_{j^*_{\mathrm{MR}}(i)} \in \mathcal{M}$ (either LTE or mmWave) that provides the maximum achievable data rate $\mathcal{R}$, i.e.,

$$j^*_{\mathrm{MR}}(i) = \underset{j \in \{1,\ldots,|\mathcal{M}|\}}{\arg\max} \{\mathcal{R}_{ij}\}, \forall i \in \{1,\ldots,|\mathcal{N}|\} \tag{6.3}$$

In this study, the achievable data rate $\mathcal{R}_{ij}$ between eNB$_j$, $j \in \{1,\ldots,|\mathcal{M}|\}$ and VN$_i$, $i \in \{1,\ldots,|\mathcal{N}|\}$ is an indication of the cell's maximum capacity and is computed from Shannon's formula as a function of the SNR, i.e.,

$$\mathcal{R}_{ij} = \frac{B}{\mathcal{N}_j} \log_2(1 + \Gamma_{ij}) \tag{6.4}$$

where B is the available bandwidth, $\mathcal{N}_j$ is the number of vehicles connected to eNB$_j$, and Γ_{ij} is the SNR between eNB$_j$ and VN$_i$. Our results therefore represent an upper bound for the throughput of the VNs, as we do not investigate the effect of medium access control mechanisms nor that of higher-layer retransmissions. We also assume that, if the measured SNR is below a predefined threshold Γ_{th}, the data rate is equal to 0. The MR strategy generally guarantees higher average throughput compared to the MS approach [123]. However, it is recognized that maximizing the data rate of all vehicles may result in an unfair data rate allocation [223]. In particular, the huge bandwidth available to mmWave systems would make the load of mmWave cells much heavier than that of LTE ones, hence resulting in mmWave cells that are congested.

6.6.1.3 Requirement-Aware (RA) Policy

To cope with MS and MR limitations, we propose a Requirement-Aware (RA) attachment policy which simultaneously maintains fairness and balances the traffic load among the cells. The association decision is therefore made as a function of the vehicle's individual QoS requirements and the availability of radio resources. The RA policy tries therefore to associate vehicles with strict reliability constraints (e.g., for advanced safety applications enabling semi- or fully-automated driving, the required data rate is relatively low, although very high levels of reliability are expected due to the sensitive nature of the exchanged information) to LTE cells since the propagation characteristics of the legacy spectrum generally deliver a good compromise between low end-to-end latency and high connection stability at long range. On the contrary, mmWave cells are selected to support those

Table 6.5: Simulation parameters for V2N attachment policies.

Parameter	Value	Description
A	1 km^2	Simulation area
h_{eNB}	30 m	Height of eNB
h_{VN}	2 m	Height of VN
M_{eNB}	8×8	mmWave eNB antenna array
M_{VN}	4×4	mmWave VN antenna array
N_{sim}	2000	Simulation runs
Γ_{th}	-5 dB	SNR threshold
$f_{c,\text{L}}$	2.4 GHz	LTE carrier frequency
$f_{c,\text{m}}$	28 GHz	mmWave carrier frequency
$P_{\text{TX,L}}$	46 dBm	LTE transmissions power
$P_{\text{TX,m}}$	27 dBm	mmWave transmissions power
B_{L}	20 MHz	LTE bandwidth
B_{m}	1 GHz	mmWave bandwidth
λ_L	4 eNB/km^2	LTE eNB density
λ_m	$\{4, \dots 80\}$ eNB/km^2	mmWave eNB density

categories of applications with the boldest per user data rate requirements (e.g., extended sensor applications, which enhance a vehicle's perception range through dissemination of sensor observations) but with looser reliability constraints. VNs may therefore be able to exploit the whole available mmWave bandwidth since less demanding VNs are associated to LTE cells. Formalizing, $\text{VN}_i \in \mathcal{N}$ connects to $\text{eNB}_{j^*_{\text{RA}}(i)} \in \mathcal{M}$ (either LTE or mmWave) that satisfies the following conditions:

$$j^*_{\text{RA}}(i) = \begin{cases} j^*_{\text{MR|L}}(i) & \text{if } \mathcal{R}_{ij^*_{\text{MR|L}}(i)} > \bar{\mathcal{R}}_i, \\ j^*_{\text{MR}}(i) & \text{otherwise,} \end{cases} \qquad \forall i \in \{1, \dots, |\mathcal{N}|\} \tag{6.5}$$

where $j^*_{\text{MR|L}}(i) = \arg\max_{j \in \{1,\dots,|\mathcal{M}_L|\}} \{\mathcal{R}_{ij}\}$ and $j^*_{\text{MR}}(i)$ is as in Eq. (6.3). In particular, the LTE eNB offering maximum data rate is chosen if the offered data rate $\mathcal{R}_{ij^*_{\text{MR|L}}}$ is above the data rate $\bar{\mathcal{R}}_i$ required by VN_i, otherwise the MR policy is applied.

6.6.2 Performance Evaluation of Attachment Policies

In this subsection we compare the performance of the proposed V2N attachment mechanisms. eNBs are deployed according to a PPP of density $\lambda_L = 4$ eNB/km^2 for LTE eNBs and λ_m spanning from 4 to 80 eNB/km^2 for mmWave eNBs, over an area A of 1 km^2. VNs are also uniformly deployed over A but, to avoid boundary effects, we collect statistics of just the VNs in a subset of the simulation area. We consider a heavily loaded scenario in which an average of 10 vehicles per mmWave eNB are deployed (so that the actual number of VNs in the network is a function of λ_m), as foreseen in [8]. To evaluate the steady-state behavior of the network, VNs' deployment consists of two steps, following the approach used in [265]. In the first step each $\text{VN}_i \in \mathcal{N}$ is attached to $\text{eNB}^*_j \in \mathcal{M}$ according to either of the algorithms described in Sec. 6.6.1. Once all VNs are attached to the network, in the second step we iteratively update the cell association by randomly picking one VN at a time. We repeat this procedure by re-allocating a random VN at each step for a fairly large number of iterations, until convergence to the long-term VN distribution among the eNBs is achieved. We also consider two different categories of vehicular traffic, to reflect the heterogeneity of future V2N applications' characteristics and requirements (a VN is assigned to one of these classes with equal probability): *class 1 applications*, where very high levels of communication stability are required although data rates

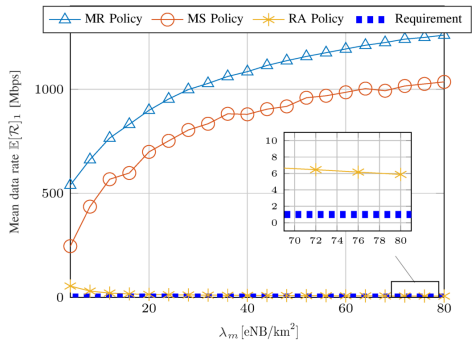


Fig. 6.29: Mean data rate $\mathbb{E}[\mathcal{R}]_1$ for VNs of class 1 vs. λ_m, for different attachment policies.

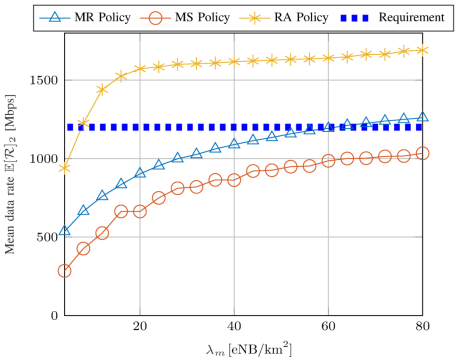


Fig. 6.30: Mean data rate $\mathbb{E}[\mathcal{R}]_2$ for VNs of class 2 vs. λ_m, for different attachment policies.

are typically below 1 Mbps, and *class 2 applications*, where data rate demands are likely to exceed 1000 Mbps while some latency is tolerated. The remaining simulation parameters are summarized in Table 6.5.

Data Rate In Figs. 6.29 and 6.30 we plot the average data rate ($\mathbb{E}[\mathcal{R}]_1$ and $\mathbb{E}[\mathcal{R}]_2$) that class 1 and class 2 VNs experience, respectively, when implementing either of the attachment policies presented in Sec. 6.6.1. At first glance we observe that, while all investigated attachment schemes satisfy class 1 data rate requirements, i.e., 1 Mbps (Fig. 6.29), MS and MR are generally not able to sustain class 2 requests, e.g., 1200 Mbps for 4K resolution cameras (Fig. 6.30), thereby making the proposed RA solution the only viable approach to maximize the communication performance for all categories of vehicular services. MR eventually meets class 2 requirements, though only for very high values of λ_m (i.e., $\lambda_m > 70$ eNB/km^2); such ultra-dense deployment, however, could be costly for network operators, in terms of capital and management expenditures, and should therefore be avoided. With the MS approach, VNs connect to eNBs showing the instantaneous highest signal strengths and avoid instead nodes that provide lower SNR values (but possibly higher rates, due to their low traffic loads). With the MR approach, VNs connect to mmWave eNBs and share the same amount of radio resources regardless of the individual traffic requirements, with class 1 VNs experiencing much higher date rate than requested, at the expense of class 2 VNs experiencing poor date rate in overloaded cells. On the other hand, the RA strategy, which biases association decisions with side information about vehicle requirements, tends to associate class 1 VNs to LTE cells (which, despite the limited capacity of the physical channel, can easily support class 1's rate requests) and saves network bandwidth for those categories of VNs with the most stringent connectivity demands. Numerically, for class 2 VNs, RA delivers up to 1.5 times higher throughput compared to MR and a 2 fold throughput increase compared to MS.

Fairness Although fairness is not necessarily a pre-requisite for V2N systems (e.g., safety-critical operations shall deserve prioritization), it still represents a major concern that should be taken into account to guarantee a minimum level of performance to the cell-edge users (or, in general, to users experiencing bad channel conditions). In Fig. 6.31 we plot Jain's fairness index J_1 for class 1 VNs, as defined in Eq. (6.1), which gives an indication on whether network resources are shared fairly among the VNs. We demonstrate that the RA solution, which associates VNs with low data rate requirements to LTE cells, guarantees more fair cell association compared to traditional attachment schemes. On one side, MS strategies homogeneously attach VNs to LTE when a few eNBs are de-

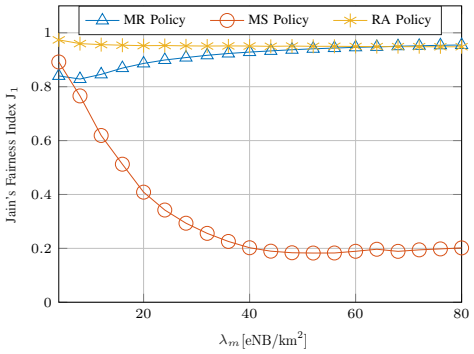


Fig. 6.31: Jain's index J_1 for VNs of class 1 vs. λ_m, for different attachment policies.

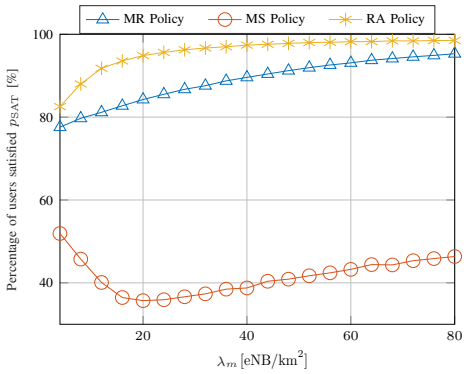


Fig. 6.32: Percentage of VNs satisfied vs. λ_m, for different attachment policies.

ployed but, as λ_m gets higher, start associating some of the VNs to mmWave eNBs too, thereby offering completely different access channels. On the other side, MR strategies attach VNs to mmWave cells which are generally not compatible with fairness as a result of the increased variability of the above-6 GHz channel. However, for high values of λ_m, i.e., when pushing the network into LOS regimes, MR's fairness performance is deemed comparable to that of RA.

Percentage of VNs Satisfied. Finally, it is interesting to compare the three attachment algorithms in terms of percentage (p_{sat})of VNs which satisfy application demands. We see that the MS approach, which tries to associate vehicles to LTE cells, is penalized by class 2 VNs whose very rigid data rate requirements cannot be sustained by low-bandwidth LTE connections. The performance particularly degrades when $\lambda_m \geq 5$ eNB/km^2, i.e., when the number of VNs in the network starts increasing as a result of denser mmWave deployments, and then slightly increases when $\lambda_m \geq 20$ eNB/km^2, i.e., when VNs that connect to mmWaves find LOS eNB. On the other hand, we observe that, although the MR approach guarantees a good level of satisfaction among the vehicles, i.e., $p_{\text{sat}} > 85\%$ for highly dense networks, with the RA scheme more than 95% of VNs are able to meet QoS demands even in low-density deployments. This is because RA discriminates association requests as a function of QoS requirements and balances VNs between LTE and mmWave eNB avoiding the overload of transmission links. Based on the above discussion, we therefore make the case that the proposed framework represents the most appropriate attachment strategy to maximize the communication performance.

6.7 Conclusions and Design Guidelines

In this chapter we compared the performance of the LTE, the IEEE 802.11p and the mmWave technologies in a V2X scenario considering both static and dynamic deployments. From the results we presented in Secs. 6.4, 6.5 and 6.6, the following conclusions can be drawn.

Impact of RAT. Given the intrinsic stability of the sub-6 GHz propagation channel, LTE and IEEE 802.11p radios deliver a good compromise between fairness and reliability for some applications characterized by less stringent requirements. They also guarantee service continuity to cell-edge users, i.e., the most resource-constrained network nodes, which are typically out of mmWave coverage. Conversely, the combination of massive bandwidth and spatial degrees of freedom has the

potential for mmWave systems to meet some of the boldest requirements of next-generation transportation systems, including high peak per user data rate and very low latency, that sub-6 GHz networks alone cannot sustain. Heterogeneous networking has thus been considered to improve network capacity by combining a reliable sub-6 GHz link with a high-capacity mmWave connection. In addition, heterogeneous networking allows for fair, robust and efficient vehicle association in a V2N scenario in which both mmWave and LTE cellular infrastructures are deployed. We show that benchmark methods which bias cell selection decisions with received signal quality or network load information cannot support those categories of vehicular traffic with the boldest connectivity requirements, and can therefore lead to sub-optimal association. On the contrary, we demonstrate that the proposed approach, which makes attachment decisions as a function of the vehicular service requirements, prevents the overload of transmission links and represents the most appropriate strategy to meet QoS demands even considering low-density deployments.

Impact of Densification. Densification of mmWave infrastructures in urban scenarios is beneficial to decrease the outage probability and deliver uniform service quality throughout the cell. Nevertheless, it must be mentioned that an ultra-dense deployment involves high capital and operational expenditures for network operators [57], and requires the provision of high-capacity backhaul connections. Our analysis also demonstrates that densification delivers negligible improvements for LTE networks due to the impact of interference in case of omnidirectional communications.

Impact of Mobility. Traditional performance analyses are conducted in static deployments, which significantly limits the level of realism and detail that can be achieved in simulations. In turn, our investigation considers vehicular mobility and demonstrates that frequent LOS/NLOS transitions, which are quite common at high speeds, can significantly degrade the performance of the network. In particular, the combined effect of a highly populated network, highly saturated channel, and high-speed mobility, can lead to values of latency even higher than 1 second, which is clearly unacceptable in the context of CAV deployments. Handover has also been identified as one of the main drivers for latency increase in congested scenarios, due to overloading of the X2 interfaces.

Impact of V2X Application. Considering UDP applications, the flow of each end-to-end connection does not self-regulate to the actual network conditions, thus congestion arises. In case of TCP connections, instead, congestion control mechanisms are used to adapt the source traffic, thereby preventing network congestion. Finally, for DASH and HTTP applications, the source traffic is asynchronous and bursty, and the network is not as stressed as in the TCP and UDP experiments. As a consequence, sub-6 GHz connections can generally guarantee improved delay performance compared to mmWave ones (up to -60% in low-density scenarios) with limited throughput degradation.

Impact of Propagation Scenario and Vehicular Density. Highway/rural scenarios generally deliver very high throughput, even in sparsely deployed networks, because the signal propagates in LOS. In turn, densification is beneficial in urban deployments to sustain acceptable quality of service: the waveguide effect resulting from the signal propagating in urban canyons can result in data rates close to the theoretical bound. Moreover, even though in crowded scenarios vehicles move at a lower speed, thus resulting in fewer handovers per second, throughput is significantly deteriorated compared to less congested scenarios due to very high levels of queuing and buffering (e.g., in highways, throughput can decrease by more than 50% in highly populated networks).

Chapter 7

Value of Information in Future Vehicular Networks

7.1 Introduction

As the automotive industry evolves towards CAVs to offer the support of more safety-critical applications [29], research efforts have been made for the design of novel architectures that guarantee timely and accurate positioning of vehicles [266]. Positioning is typically provided by on-board GPS receivers, which however may not always be accurate enough [267]. For this reason, *data fusion* techniques have also been considered by combining several positioning strategies (including, but not limited to, dead reckoning, map matching, and camera image processing) into a single solution that is more robust and precise than any individual approach [268]. In this perspective, the unique characteristics of CAVs might cause rapid dynamics and unpredictable changes in the network topology [269], thereby requiring regular position updates to be disseminated as timely as possible, i.e., ideally at the very same instant they are generated. At the same time, there is a constraint on the amount of information that can be successfully broadcast over bandwidth-limited communication channels to avoid packet collisions and overload of transmission links [270]. New communication radios operating in the mmWave bands have thus been investigated as a means to increase network capacity thanks to the large bandwidth available at high frequencies [191], as explained in Chapter 6. However, we argue that even a significant increase in the channel capacity may not be sufficient to satisfy the boldest QoS requirements of future automotive applications, in particular in scenarios with multiple active services requiring different degrees of automation. In this context, it is thus fundamental to set a bound on the amount of information that is distributed over bandwidth-constrained communication channels.

A traditional approach is to monitor the *Age of Information (AoI)* [271], i.e., the obsolescence of the data, so that vehicles broadcast sensory messages that are not too old. However, the complex dynamics of vehicular networks affect the rate of decay of the information, making it difficult to set a fixed threshold for the AoI to discriminate between useful and obsolete pieces of data. Another approach is to discriminate the *Value of Information (VoI)* [272–274] in order to use the (limited) transmission resources in a way that maximizes the utility for the target applications. The value assessment process should be computationally efficient, so that it can be executed even with the limited on-board com-

This chapter is based on the contributions presented in **[J10, C10, C11, C13, C15, P1]**.

putational resources of mid-range and budget car models, and be completed in real-time. However, despite these desirable characteristics, it is still unclear how VoI theory can be applied in vehicular deployments to improve broadcasting of positioning information.

7.1.1 Motivations and Chapter Structure

Traditionally, VoI has been investigated in underwater and sensor networks to decide how much information to transmit in resource-constrained channels. In this chapter, we investigate for the first time the concept of VoI in vehicular networks. In particular, Sec. 7.3 provides a taxonomy of information sources in vehicular networks and describes the attribute categories that should be accounted for in the VoI assessment process. We then shed lights on two fundamental research questions.

- *When to Broadcast Information?* Vehicles typically broadcast periodic updates with positioning information. The intrinsically variable topology of CAVs, however, might make periodic broadcasting strategies inefficient: long inter-transmission intervals may prevent the timely dissemination of positioning information in safety-critical situations, while very frequent broadcasting may overload the transmission links with useless data. Congestion avoidance mechanisms have thus been proposed in the literature to regulate information distribution as a function of the network load [275]. However, these techniques dynamically adapt the CAVs transmission parameters, e.g., by controlling the number of neighboring vehicles [276] or assigning different priorities to vehicles with different operating conditions [277], regardless of the level of positioning accuracy that is achieved from the information that is successfully delivered. In Sec. 7.4 we propose to use VoI as the decision factor for broadcasting and design a *threshold-based* broadcasting algorithm which *(i)* estimates the positioning error of the vehicle and of its neighbors within communication range, based on purely predictive Unscented Kalman Filter (UKF) tracking operations, and *(ii)* makes vehicles distribute state information messages in case the estimated error is above a predefined threshold. The performance of this approach is compared with a baseline *periodic broadcasting* solution that instructs vehicles to broadcast state information updates at regular intervals. Our results show that the proposed broadcasting algorithm, in spite of its simplicity, can reduce the average position estimation error by more than 10%, and its 95th percentile by more than 20%, compared to the literature approach.
- *What Information to Broadcast?* Sending information about all of on-board sensor observations may cause high network load, especially in a dense urban environment with a large number of vehicles. The high load on the radio channel would likely lead to packet loss and large communication latency, potentially degrading the accuracy and timeliness of vehicles' perception of road objects. It is therefore important to intelligently select the more valuable information to be sent at each transmission opportunity. In Sec. 7.5 we propose a method to assess VoI and rank scheduling options as a function of the characteristics of the network in which the nodes are deployed. Our innovative framework exploits Analytic Hierarchy Process (AHP) to quantify the expected VoI based on time, space and quality interdependencies. We validate the technical accuracy of this approach in realistic scenarios and show how VoI evolves as a function of the scenario (i.e., urban or highway), the sensor resolution, the type of observation, the communication distance and the age of information. Thanks to its generality and computational simplicity, our method guarantees timely and efficient value assessment operations.

Finally, in Sec. 7.6 we conclude the chapter by identifying guidelines and open challenges for the design of VoI strategies in vehicular networks.

Table 7.1: Literature on value of information and broadcasting strategies in vehicular networks.

Topic	Relevant References
Tracking Strategies	[278–287] Traditional tracking [268, 288–294] Long-term tracking
Broadcasting Strategies	[295–297]
Congestion Control Strategies	[241, 298, 299]
Value of Information Strategies	[300–304] Literature on VoI theory [305] Heuristic VoI strategies [306, 307] Adaptive VoI strategies [308, 309] Machine learning VoI strategies [273, 310, 311] Analytic VoI strategies

7.2 Related Work

In V2X , vehicles are equipped with on-board sensors, which are used to gather data about the surrounding environment. These pieces of data are then broadcast within the network through wireless radios (e.g., LTE, IEEE 802.11p and mmWaves, as seen in Chapter 6), and are used to implement a tracking system [266] whose target state is the set of positions of all surrounding neighbors. Dissemination of sensory information may however exceed the capabilities of existing communication technologies, thereby making it fundamental to prioritize the transmissions that have greatest value for target applications. In this section we present a selection of related work for each of such topics.

Vehicular Tracking Strategies An example of a tracking framework for CAVs is given in [278] and most of research in this area is based on similar architectures, but with different vehicle mobility and data processing schemes. An analysis of the main mobility models used in vehicle tracking was given in [279]. For what concerns data processing, the most common choice is adopting a Bayesian Filtering (BF) approach, typically based on the Kalman Filter (KF) [280], the UKF [281] or the PF [282]. A tracking framework based on the UKF and the Constant Turn Rate and Acceleration (CTRA) motion model was presented in [312]. In [283], route information and digital map data were jointly processed by a PF algorithm. In [284], position forecasting was achieved by using a system based on a Hidden Markov Model (HMM) [285] and the Viterbi algorithm [286]. We highlight that, in all the BF-based architectures, the performance greatly depends on the filter settings, e.g., the process and estimation noise covariances, which must be known *a priori*.

Conventional tracking approaches mainly focus on the real-time estimation of the target state. However, most advanced CAV applications also require a prediction of vehicles' future trajectories. Long-term forecasting can be achieved by simply applying the predictive step of a BF filter to the last available state estimate. However, this solution is very sensitive to imperfections of the motion model: to overcome this issue, more sophisticated approaches have been proposed in the literature. In [288], the output of a KF was used to perform a parametric interpolation of the future path of the target vehicle. In [289], Dead Reckoning (DR) was used to improve the performance of packet forwarding in a highway scenario. Another possibility consists of describing vehicle position prediction as a time series forecasting problem [268]. Hence, Machine Learning (ML) techniques can improve target state estimation over a large time horizon: in [290], Support Vector Machines (SVMs) were used to forecast vehicle trajectories, allowing the estimation of target positions when the GPS signal is not available. In [291], a Neural Network (NN) system was trained with historical traffic data and then used to predict vehicles' speeds. Although the ML approach generally guarantees

high performance, it requires a large amount of data for the initial training and suffers from significant computational complexity. ML techniques are often combined with BF algorithms: in [292], the authors presented a system that makes use of an HMM module to estimate vehicle maneuvers and an SVM module to predict future vehicle trajectories. In [293], a Radial Basis Function (RBF) classifier was used to compute the inner parameters of a PF, which estimated the long-term motion of the target. In [294], the results of a maneuver recognition system were combined together with the output of a tracking system based on the CTRA motion model.

Broadcasting Strategies Regardless of the complexity of the tracking framework, the overall system performance degrades if on-board sensor measurements are not sufficiently accurate. Users can share local information to compensate the low quality of the input data. For such purpose, CAV nodes periodically broadcast their own system state by using the IEEE 802.11 technology [25], that, however, is based on a random channel access scheme. A solution for reducing the channel occupancy and collision probability is to select the optimal transmission strategy as a function of the instantaneous positioning error of nearby vehicles. The authors in [295], for example, proposed for the first time a broadcasting strategy in which the ego vehicle triggers new transmissions whenever the estimates of the neighbors' errors are above a predetermined threshold. However, such analysis was provided only for specific case scenarios and the framework that predicts future vehicles' states was quite obsolete with respect to current vehicular tracking techniques. Similarly, in [296], the transmission rate by which new information is disseminated within the network is regulated according to both the positioning error and the estimated number of packet collisions. Nevertheless, the authors assumed that vehicles are always aware of the number of packets lost during each timeslot. This may not always be true in vehicular networks where most packet collisions are caused by the hidden terminal problem and, thereby, cannot be directly sensed by other nodes. In [297], the authors analyzed the inter-vehicle communication dynamics that cause the hidden terminal problem. In the same work, the limitations of the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol were addressed by varying the vehicle communication range according to the channel occupancy. However, the validity of the approach was proven only in a highway scenario and cannot be generalized to more complex and unpredictable environments.

Congestion Control Strategies The high number of broadcast communications and the variability of network topology are very critical issues for CAV applications. In scenarios with a large vehicular density, the wireless channel may become congested and, consequently, the transmitted information may be lost because of packet collisions. Defining novel congestion control schemes, which suit the characteristics of modern vehicular networks, is a problem of considerable interest. Over the years, many researchers have proposed different MAC strategies that adapt inter-vehicle communications to channel conditions. In [241], the authors presented a rate-adaption strategy that ensures channel stability through vehicular networking. The convergence of the proposed algorithm was theoretically proven and guidance for the choice of the algorithm parameters is provided. In [298], the hidden terminal problem was avoided by adopting a time-slotted structure and a TDMA scheme. In particular, each vehicle is assigned a dedicated timeslot for each frame, during which it can alert its neighbors about its future transmissions. In [299], the authors focused on improving congestion control in road intersections by using a locally-distributed strategy based on ML. Dedicated road infrastructures have the task of deleting redundant communications and assigning specific CSMA/CA parameter settings to different clusters of transmission requests.

Value of Information Strategies Traditionally, VoI has been studied under an economic perspective to support business-level processes ranging from data management to efficient decision making [300]. However, such traditional methods implicitly assume that information is consumed by humans with limited automation capabilities and do not account for cases where the information sources are not directly under the users' control (e.g., for data automatically generated by vehicle's on-board sensors). The VoI theory has also been applied to underwater systems to decide when and how much information to transmit through resource-constrained acoustic networks [301]. Such techniques, however, present many limitations in a V2X context, due to the completely different propagation characteristics of highly mobile nodes and the intrinsic variability of the vehicular channel. Relevant VoI determination approaches have also been investigated in the military context to prioritize the information to be disseminated to or gathered from soldiers in a battlefield environment [302]. These strategies, however, are too restrictive in view of the emerging vehicular application requirements. Recently, VoI-aware methods have been proposed for sensor network applications, e.g., environmental sensing, surveillance and health monitoring, to provide data that fulfills the needs of users under resource, monetary and latency constraints [303]. Sensor network optimization, however, has been mostly associated to traditional QoS paradigms, e.g., to minimize power consumption [304]. In the vehicular context, instead, sensory information should be valued based on the utility provided to the final receiver in a specific scenario. Moreover, VoI is expected to change much more significantly than considering static sensors deployed, e.g., in smart cities.

VoI has been typically assess through different approaches. *Heuristic strategies* (e.g., greedy methods) are well-known solutions for performing excellent VoI assessment when good empirical functions are available, and can therefore be used as a benchmark against other solutions [305]. However, they may fail when constrained by runtime limitations or insufficient network resources, and may suffer from significant power consumption and non-negligible delays.

Adaptive strategies, e.g., [306, 307], hierarchically refine value assessment operations by relying on feedback messages which describe how helpful the received information was in relation with the requirements of target applications. Both distributed and edge-assisted approaches are being discussed, and the trade-off involves signal latency, power consumption, system overhead, and cost. In the first case, the endpoints exchange feedback messages through point-to-point transmissions while, in the second case, the feedback is relayed through edge/cloud computing systems. While incurring communication overhead for both data collection and model distribution, edge-assisted solutions leverage a much larger data set for model training than in a distributed strategy, thereby resulting in a more accurate VoI estimation. In turn, distributed methods can process the feedback in real time, thereby yielding more responsive vehicular operations.

Machine Learning (ML) can also develop models that link VoI to actions (e.g., settings of link parameters) and effects (e.g., the corresponding performance metrics), in such a way that actions are optimized to the specific operational scenario. *Generative Deep Neural Networks*, for instance, can be used to measure the mutual information of different combinations of the sensory readings and dynamically assign them value scores which depend on the degree of correlation [308]. Similarly, autoencoders can be trained in an unsupervised manner to extract features from input vectors and predict the a posteriori probability of a sequence given its entire history [309]. Despite some encouraging features, ML approaches require a large amount of sensory observations for training, which are often not publicly available. Furthermore, ML operations can hardly be completed in low latency, especially considering the limited on-board computational resources of budget car models.

Analytic approaches achieve VoI estimation through mathematical models. Stochastic methods, i.e., continuous-time Markov chains, recreate network scenarios and estimate the value that the availability of various sensor information might bring to the receiver at different times. Information the-

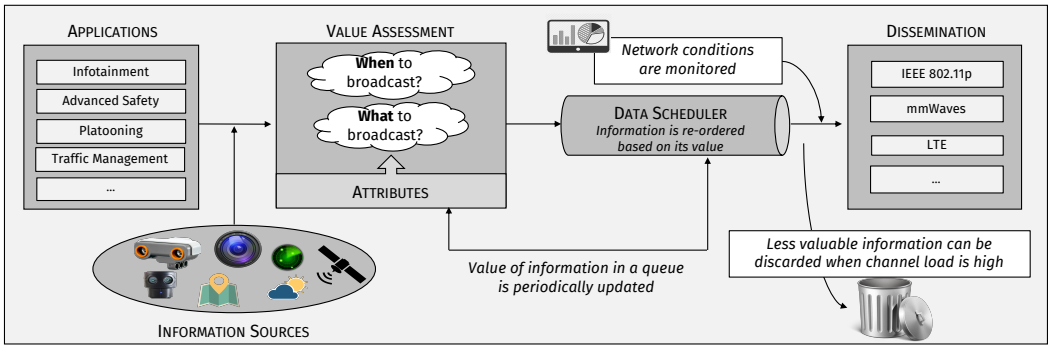


Fig. 7.1: Block diagram of the VoI assessment process.

ory can then be applied to quantify the expected VoI based on its novelty for potential receivers [310]. These approaches provide fine-grained analysis of VoI tuned to the specifics of the modeled scenario but are not suitable for real-time VoI. *Analytic Hierarchy Process (AHP)* methods [311] can be employed to value information in various application domains based on pairwise comparisons of specific criteria and to ultimately score the different data dissemination alternatives. However, the AHP methodology is not an absolute decision making technique, since relative importance levels are empirically determined based on subjective comparative [273].

7.3 Characterization of VoI in Vehicular Networks

VoI is extremely context-dependent, i.e., different outcomes are possible depending on the characteristics (and requirements) of the application and the sources of information that are being considered. This section provides an overview of the requirements of some automotive applications (Sec.7.3.1), the types of information sources they can use (Sec. 7.3.2), the attributes to be considered to assess the VoI (Sec. 7.3.3), and the ways the information can be broadcast (Sec. 7.3.5), as represented in Fig. 7.1.

7.3.1 Application Taxonomy

We focus on four broad application domains that, for their generality, complementarity and significance, we believe are good representatives of future automotive services. Although their requirements are not yet fully specified, some qualifying characteristics have already been outlined by the 3GPP in [30] (we refer to Sec. 1.4 for a detailed description of future vehicular network applications).

- *Infotainment* refers to services that deliver a combination of information and entertainment. Data rate demands are in the order of hundreds of Mbps, while latency is reasonably tolerated.
- *Advanced safety* enables semi- or fully-automated driving through persistent dissemination of perception data. The required throughput is relatively low (i.e., less than 50 Mbps), while latency must be very small (i.e., less than 100 ms for high degree of automation) to ensure prompt reactions to unpredictable events.
- *Platooning* refers to the set of services that enable the vehicles to cooperatively travel in close proximity to one another at highway speeds. Data rate up to 65 Mbps is required, while the latency ranges from 10 ms to 500 ms depending on the distance.
- *Traffic Management* enables traffic control and coordination through the creation of a Local Dynamic Map (LDM) [313], which integrates sensor data streamed by vehicles in a geographical

area. These operations may require relatively high data rate compared to safety applications, while some latency can be tolerated (depending on the degree of automation).

7.3.2 Information Sources Taxonomy

In the following, we list some typical examples of types of information that can be collected and exchanged by connected cars to optimize the performance of the vehicular communication networks and the supported services.

- *Surrounding Information:* The awareness of the surrounding environment is fundamental for a wide set of services. Such information can be provided by different on-board sensors. Radars, currently operating in the mmWave spectrum between 76 and 81 GHz, are used for applications like adaptive cruise control, cross traffic alerts, and assisted lane change. Although they enable accurate detection and localization of the surrounding objects, they are relatively less suitable for object recognition and classification purposes. For spatially close nodes (i.e., less than 60 meters), sonar sensors provide accurate echolocation through acoustic waves. Cameras use visible light or infrared and have been used for road signs recognition, enhanced blind spot detection and lane departure alert, but generate high bit-rate data flows (from 100 Mbps to 700 Mbps, depending on the image quality). Light Detection and Ranging (LIDAR) sensors make use of laser beams to generate high-resolution 3D depth images for accurate detection, localization and recognition of the surrounding objects, and the required data rate is comparable to that of automotive cameras.

- *Position Information:* GPS offers global time synchronization and absolute (though not always accurate) positioning, although other localization techniques, e.g., based on dead reckoning, cellular localization, or image processing, may be required to improve positioning accuracy. For spatially close nodes (i.e., less than 60 meters), echolocation can also be provided by sonar sensors, which measure distance using sound waves. Through Basic Safety Messages (BSMs) and Cooperative Awareness Messages (CAMs), vehicles can also determine their mutual speeds and accelerations.

- *Traffic Conditions and Prediction:* Real-time traffic information can be obtained collecting vehicles' localization measurements and re-broadcasted by network infrastructures. Traffic information may also be complemented with side information, e.g., the presence of sensitive locations (schools, hospitals) or temporary events (city marathons, social/political events), which may have an impact on the traffic in the area.

- *Environmental Conditions:* Weather conditions, including rain, snow, fog, dust, ice and black ice, can be provided by national weather-alerting systems and made accessible through network infrastructures. The corresponding information flow is light, with loose reliability and latency constraints [314].

- *Historical Data:* Past observations can be turned into experience. As a vehicle is able to recognize a specific profile (e.g., a driver, a place, a road), it can access its saved historical data and exploit this information, e.g., to adapt its driving decisions. Such historical data may not be available from the vehicles currently on the road, but can be stored in infrastructure or cloud servers and downloaded when required. Data traffic flow is expected to be low, with low latency requirements.

7.3.3 Attributes Taxonomy

VoI typically decays over time at a rate that is application dependent. Value determination should indeed account for specific attributes, categorized as follows.[1]

- *Timeliness:* Typically, the VoI in vehicular networks is affected by obsolescence, so that the value decreases with the relative age of information, i.e., the time between the generation and reception of the information, normalized to its lifetime, i.e., the temporal horizon over which that piece of information is considered valuable. The temporal horizon over which the data content embeds valuable information is highly application-dependent. For example, safety messages are extremely delay sensitive, while infotainment data is more time-resilient, since buffering and error concealment techniques may maintain good service quality even in the presence of communication gaps.
- *Proximity:* Similarly, we can define a spatial horizon over which an information is valuable for the potential receiver(s). For example, for collision warning applications, the dissemination of sensory data between close-by vehicles is more valuable than those coming from farther nodes, while the situation can be the opposite for some perception applications.
- *Information Quality:* The VoI may depend on the quality of the information source, which may be assessed in terms of accuracy (e.g., for GPS coordinates), resolution (which is a measure of the image width/height and the frame rate, e.g., for video), and variance. The importance of quality attributes depends on the target application, e.g., high-definition LIDAR images may be of little value for infotainment services, but very precious for safety services instead.
- *Urgency:* This attribute discriminates the different pieces of information based on the level of urgency of their target applications, e.g., data used for safety applications have higher urgency than those required for infotainment.
- *Generalizability:* This attribute captures the extent of the interest of the information to multiple applications, e.g., GPS can be exploited by applications ranging from infotainment to safety. Such pieces of information should be awarded higher value than those of interest for a narrower set of services.
- *Novelty:* The novelty attribute captures the relative importance of a certain piece of data with respect to the standard flow generated by a source, e.g., a piece of information that can be easily predicted by the receiver, based on the available knowledge and past observations, will have low novelty.
- *Provenance (Information Path):* This attribute refers to the entire end-to-end source-to-destination path that the data has followed. If the message has been relayed through multiple hops (e.g., platooned nodes) the carried information may have been more likely accessed or corrupted by malicious attackers, thereby limiting its integrity and, consequently, its value. The VoI assignment may also be based on the *trust* of the destination towards the source providing the information, which may result from past interactions between the endpoints.

7.3.4 Value Assessment

Let V denote the set of vehicles in the scenario which are equipped with on-board sensors (e.g., cameras) and a hardware unit enabling V2V communications. At time t, a vehicle $v_i \in V$ perceives a road object through its on-board sensors, and generates a perception record $o_i(t)$, i.e., the vector of

[1] We do not preclude in the future to extend the attribute classification proposed in Sec. 7.3.3, although we expect such extension to be highly correlated to the proposed structure.

sensors' measurements, for each of the detected objects. Our key research question in this study is whether vehicle $v_i \in V$ should broadcast perception record $o_i(t)$ to potential receivers and, if so, when it should be broadcast to avoid channel congestion and overload of the transmission links. For example, $o_i(t)$ may not be so valuable if destination nodes are not able to receive the perception message in a timely way. To do so, automotive applications annotate each piece of information they generate with *metadata* to facilitate the value assessment operations. Metadata can include (i) type of information, (ii) timestamp of when the information is generated, (iii) network-level requirements to be satisfied, (iv) importance level of data, (v) accuracy and resolution of data, (vi) source of information. We assess that vehicles should distribute the information about each perceived observation whenever the positioning error estimated from the exchanged information passes a reference value, as we will describe in Sec. 7.4. Moreover, observations should be broadcast only if their value is above a pre-defined threshold for at least one vehicle within the communication range, as we will describe in Sec. 7.5.

7.3.5 Data Scheduler and Dissemination

Once VoI has been assessed, the data scheduler sorts the pieces of information in a descending order of value and sequentially forwards them to the surrounding receivers. Timely and accurate data dissemination must be guaranteed through any type of wireless interface for V2X communications, including, but not limited to, IEEE 802.11p, LTE, Wi-Fi and the mmWave technology (see Chapter 6). Generally, IEEE 802.11p and LTE systems offer relatively low-rate connectivity but guarantee very stable and robust transmissions at short/medium distances (thanks to the intrinsic stability of the low-frequency channels and the omnidirectional transmissions). Conversely, mmWave systems support very high-throughput connections but exhibit high instability due to the severe signal propagation characteristics and the need to maintain beam alignment. In this context, dissemination operations can be improved by using multiple radios in parallel (i.e., heterogeneous networking) to complement the limitations of each type of network. The data scheduler has the responsibility to continuously select the most suitable technology to transmit each data packet, a research problem that, to date, has not yet been throughly investigated. When it becomes impractical to deliver valuable information through inter-vehicular transmissions (e.g., in case of congested networks or considering limited-capacity channels), the data can be offloaded to more powerful nodes (e.g., road side units, cellular infrastructures, network elements offering edge-cloud services) which act as coordination units for relaying the most valuable data packets through robust connections. Those schemes enable energy-efficient information distribution but suffer from significant delays (the sensor-based measurements may be forwarded through possibly multiple wireless links before being actually accessible to the destination node).

7.4 When to Broadcast?

The broadcast decision must depend on the type of service that the acquired measurements intend to provide. In particular, a distinction between non-critical and critical services must be made. In the first case, the different pieces of data can be aggregated or modified at the application layer before being actually exchanged among the nodes, thereby promoting resource consumption minimization and preventing the circulation of duplicate or superfluous data. In the second case, instead, more reactive dissemination actions are needed, e.g., urgent pieces of data may be continuously broadcast to neighboring nodes until the emergency signals have been successfully distributed. Markov-based techniques might also be implemented to estimate the correlation and interdependencies among

multiple signals, thus making it possible to statistically compress the number of observations to be transmitted [306]. Moreover, the broadcast operations should account for the instantaneous channel characteristics (e.g., the channel delivery ratio) especially when considering ultra-dense vehicular scenarios that would further strain the network bandwidth and reliability. In particular, if the network load exceeds a predefined threshold, the data scheduler should either reduce the data transmission speed or discard the least valuable data packets.

In this context, among all possible sources of information, the dissemination of vehicle position data all over the network is fundamental in V2X operations, as safety-critical applications often need to know the position of other vehicles over a large area. Therefore, in this study, investigate ***when*** to broadcast positioning information in vehicular networks in such a way to ensuring accurate position estimation and preventing the congestion of the communication channel. To do so, we design an innovative threshold-based broadcasting algorithm that forces vehicles to distribute state information if the estimated positioning error is above a certain error threshold.

7.4.1 System Model

In this subsection we present the system model that is considered in our study. First, we theoretically model a CAV network as a time-varying Euclidean graph, whose nodes and edges represent vehicles and their communication links, respectively. Then, we define a performance metric that takes into account both the tracking errors and the vehicle positions. Finally, we describe the tracking system implemented by each node in the network and the communication channel through which vehicles' state information are broadcast.

7.4.1.1 General Model

We represent a CAV network as a Euclidean graph $G = (V, E, r)$, i.e., an undirected graph whose vertices are points on a Euclidean plane [268]. V represents the set of nodes, E represents the set of edges and r is the node's communication range. We say that two vehicles $v_i, v_j \in V$, $i \neq j$, are connected by the edge $< v_i, v_j >$ if the distance $d_{i,j}$ between them is shorter than the communication range r, i.e., $E = \{< v_i, v_j >: i \neq j \wedge d_{i,j} < r\}$. Since the composition of the edge set depends on the positions of the vehicles, the topology of the network is time-varying, e.g., new edges can be activated or disabled according to how vehicles move. In our model, we assume that vehicles move in a two-dimensional space; while not always realistic, this hypothesis does not compromise the accuracy of our analysis. To highlight the time dependency of the network, we denote by $G(t) = (V(t), E(t), r)$ the network graph at time t. For simplicity, we assume that time is divided into discrete timeslots, so that the system evolves in steps. Hence, we define the neighbor set $N_i(t)$ of v_i during t as the set of all the vehicles connected to v_i by an edge in $E(t)$, i.e., $N_i(t) = \{v_j \in V(t) :< v_i, v_j >\in E(t)\}$.

The behavior of each vehicle v_i in the network at time t is represented by a 6-tuple $s(t) = \Big(x(t), y(t), h(t), u(t), a(t), \omega(t)\Big)$, which we call *vehicle state*. In particular, x and y are the Cartesian coordinates of the vehicle on the road topology, h is the vehicle's heading direction, u and a are the vehicle's tangent velocity and acceleration, respectively, and ω is the vehicle's angular velocity. The physical distance between the positions of vehicles v_i and v_j at time t is given by $d(s_1(t), s_2(t)) = \sqrt{(x_1(t) - x_2(t))^2 + (y_1(t) - y_2(t))^2}$.

7.4.1.2 Error Function

Consider a reference vehicle $v_i \in V(t)$, called the *ego vehicle*, which tracks a group of other vehicles that we call *target vehicles*. Hence, we denote by $\hat{N}_i(t)$ the subset of $N_i(t)$ containing the target vehi-

cles, and by $\hat{s}_{i,j}(t)$ the state estimate of $v_j \in \hat{N}_i(t)$ performed by v_i at time t. Under these hypotheses, the performance of the ego vehicle in terms of position estimation accuracy can be assessed by an *error function* $\mathcal{F}(v_i, t)$, which is a weighted average of the position estimation errors made by the ego vehicle with respect to itself and all its target vehicles. Formally, we have

$$\mathcal{F}(v_i, t) = \frac{1}{|\hat{N}_i(t)| + 1} \left(\lambda_{i,i}(t) d(\hat{s}_{i,i}(t), s_i(t)) + \sum_{v_j \in \hat{N}_i(t)} \lambda_{i,j}(t) d(\hat{s}_{i,j}(t), s_j(t)) \right). \tag{7.1}$$

In (7.1), $d(\hat{s}_{i,i}(t), s_i(t))$ and $d(\hat{s}_{i,j}(t), s_j(t))$ represent the error made by v_i in estimating its own state $s_i(t)$ and the neighbor state $s_j(t)$, respectively, $|\hat{N}_i(t)| + 1$ represents the total number of estimations carried out by v_i, and $\lambda_{i,k}(t)$ is the generalized logistic function defined as:

$$\lambda_{i,k}(t) = A_\lambda + \frac{E_\lambda - A_\lambda}{\left(C_\lambda + D_\lambda e^{-B_\lambda (d(s_i(t), s_k(t)) - d_0)} \right)^{1/\nu_\lambda}}. \tag{7.2}$$

The parameters in (7.2) characterize the logistic function's shape. In particular, A_λ is the function lower asymptote, E_λ is the carrying capacity, B_λ is the growing rate, ν_λ affects the function steepness, D_λ defines the value of $\lambda_{i,i}(0)$, C_λ determines the function upper asymptote while d_0 is the threshold beyond which $\lambda_{i,j}(t)$ starts decreasing. To evaluate the performance of the whole network, we define $\mathcal{F}(t)$ as the average of $\mathcal{F}(v_i, t)$ among all vehicles $v_i \in V(t)$:

$$\mathcal{F}(t) = \frac{1}{|V(t)|} \sum_{v_i \in V(t)} \mathcal{F}(v_i, t). \tag{7.3}$$

7.4.1.3 Tracking System

To minimize the positioning error defined in (7.1), the ego vehicle must estimate its state and the state of every other vehicle in the set $\hat{N}_i(t)$ in every timeslot. To do so, the ego vehicle exploits both the information gathered by its on-board sensors and the information received from its neighbors through inter-vehicle communications. To allow the estimation of $s_i(t)$, we assume that, in every timeslot, the ego vehicle's on-board sensors provide a new observation $o(t)$ of $s_i(t)$. Hence, the ego vehicle can model the evolution of its own state through a Bayesian approach, obtaining the system

$$\begin{cases} s(t+1) = f(s(t)) + \zeta(t), \\ o(t) = m(s(t)) + \eta(t). \end{cases} \tag{7.4}$$

In (7.4), the first equation describes the evolution of the vehicle state $s(t)$ over time, while the second equation describes the relation between $s(t)$ and the state observation $o(t)$. In particular, $f(\cdot)$ is a function describing the CTRA motion model given in [315], while $m(\cdot)$ is a function representing the vehicle's measurement system. Moreover, $\zeta(t)$ and $\eta(t)$ represent the process and measurement noises, respectively, and are modeled as independent Gaussian processes with zero mean and covariance matrices Q and R. Once all the parameters in (7.4) are defined, the ego vehicle can estimate its own state by using a BF algorithm. In our model, each vehicle implements a UKF algorithm exploiting the *sigma points* parameterization given in [316]. By exploiting the UKF and the system equations given in (7.4), the ego vehicle obtains the estimate $\hat{s}_{i,i}(t)$ of its own state $s(t)$ and the related covariance matrix $P_{i,i}(t)$, which represents the uncertainty of the state estimation, in each timeslot t.

To allow v_i to estimate the states of the other network nodes, each vehicle $v_j \in V(t)$ can transmit the estimate $\hat{s}_{j,j}(t)$ of their own state and the related uncertainty $\hat{P}_{i,i}$. The time frame by which new transmissions are initiated depends on the selected broadcasting strategy, as described in Sec. 7.4.2.

Each message transmitted by v_j is received by all the vehicles in $N_j(t)$ after a certain communication delay (provided that the transmission is not interfered, as we will explain later). Whenever the ego vehicle gets a message from another node that was not previously in its neighbor set, it initializes a new UKF having as initial state and uncertainty the received state and covariance matrix, respectively. The new filter propagates the initial state over time by evolving the model blindly, and is then updated when new information is received. If a vehicle v_i does not receive state updates from a neighbor v_j for a period longer than Δ_{track}, it stops to track the considered target, i.e., v_j is removed from the set N_i.

7.4.1.4 Channel Access Model

Inter-vehicle communications are modeled following the IEEE 802.11p standard (we refer to Sec. 6.3.2 for a more detailed description of the IEEE 802.11p technology). DSRC defines seven different channels at the PHY layer, each containing $n_{sc,tot} = 52$ subcarriers [317]. For simplicity of discussion, we assume that only a limited number of subcarriers $n_{sc} \leq n_{sc,tot}$ can be used for broadcasting state information messages, while the rest is reserved for other applications. DSRC implements the CSMA/CA scheme at the MAC layer, where nodes listen to the wireless channel before sending.

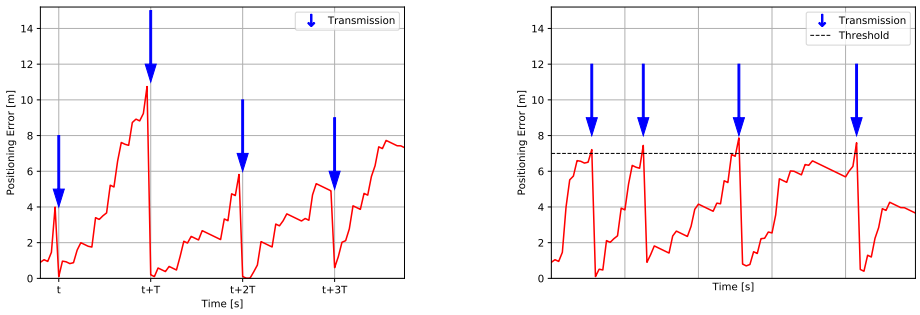
We consider an ideal 1-persistent CSMA/CA scheme, capable of successfully arbitrating the channel access among in-range vehicles in such a way that one single transmission per subcarrier and timeslot is enabled, even in case of multiple potential transmitters. However, we assume that collisions can still occur among out-of-range vehicles that transmit towards the same receiver, an issue known in the literature as hidden node problem. Therefore, the transmission from a vehicle v_a to a vehicle v_b will suffer from a hidden terminal collision if any of v_b's neighbors that are out of v_a's range start a transmission that overlaps in time and frequency with v_a's signal. We also design and implement a congestion control algorithm to reduce the channel collision probability. More details will be given in Sec. 7.4.3.

7.4.2 When to Broadcast? Proposed Strategies

In this subsection we describe the communication strategies that are used to regulate the broadcasting frequency in the model described in Sec. 7.4.1. In particular, two different strategies are considered. The first strategy is called Periodic Broadcasting (PB) and is already implemented by most CAV applications. The second strategy is called Error Threshold Broadcasting (ETB) and is our original proposal [318, 319].

7.4.2.1 Benchmark: Periodic Broadcasting (PB)

The PB strategy represents the benchmark solution of our analysis. In the PB scenario each vehicle chooses a constant inter-transmission period T_{period}, so that its communication process follows an almost regular time-frame. Reducing T_{period} would allow for a reduction of the misdetection probability, which is the probability that a neighbor v_j belongs to N_i but not to $\hat{N}_i$, at the expense of increasing the probability of channel access collisions. The false alarm probability, i.e., the probability that a neighbor v_j belongs to $\hat{N}_i$ but not to N_i, should follow the same trend. In particular, a new transmission is allowed each time a new neighbor is sensed and no transmissions were initiated in the previous two timeslots.



(a) PB broadcasting strategy.

(b) ETB broadcasting strategy.

Fig. 7.2: Position error evolution for the broadcasting strategies presented in Sec. 7.4.2.

7.4.2.2 New Proposal: Error Threshold Broadcasting (ETB)

In the ETB scenario each vehicle chooses an error threshold E_{thr} and regulates its communication behavior so that the overall position estimation error never exceeds E_{thr}. To reach this goal, the ego vehicle defines an additional UKF, which replicates the UKF operations of all the neighbor nodes that are tracking the ego vehicle itself. This filter propagates the ego vehicle's state by using only its predictive step with no sensor input, as done by the other vehicles. Each time the ego vehicle triggers a new communication, the filter state is updated mimicking the operation performed by neighbor vehicles upon reception of the packet. Hence, at each timeslot t, the ego vehicle knows both the a posteriori state estimate $\hat{s}_{i,i}(t)$, which is the output of its main filter, and the a priori state estimate $\hat{s}^p_{i,i}(t)$, which is the output of its purely predictive filter. $\hat{s}^p_{i,i}(t)$ therefore represents the state estimate of v_i made by its neighbor vehicles. At each timeslot, the two different estimates are compared: if $d(\hat{s}_{i,i}(t), \hat{s}^p_{i,i}(t))$ exceeds E_{thr}, a new transmission is initiated. We observe that, as before, the communication process can vary according to some specific events. A maximum inter-transmission period T_{max} is defined to mitigate the undetection of new neighbors, and additional transmissions are initiated in case new neighbors are detected.

An intuitive understanding of the PB and ETB dynamics is provided by Figs. 7.2a and 7.2b, which represent the evolution of the position error $d(\hat{s}_{i,i}(t), \hat{s}^p_{i,i}(t))$ according to the transmission process in the two cases. In the PB scenario, we can observe that new transmissions are initiated in a regular fashion, regardless of the value of $d(\hat{s}_{i,i}(t), \hat{s}^p_{i,i}(t))$. Instead, in the ETB scenario, new transmissions are initiated only when $d(\hat{s}_{i,i}(t), \hat{s}^p_{i,i}(t))$ is above a certain threshold.

In both the PB and ETB scenarios, the best strategy setting would require to determine the optimal value of T_{period} and E_{thr}, respectively. Such values can be computed through an exhaustive and computationally heavy approach that iterates on all the possible values of number of available subcarriers, the vehicular density, the characteristics of the road map, and other automotive-specific parameters, or with less resource-heavy congestion control techniques. In the next section, we will show how to adapt existing congestion control techniques to set both T_{period} and E_{thr} to minimize channel congestion.

7.4.3 When to Broadcast? Congestion Control

Considering the dynamic nature of vehicular networks, the potential of the broadcasting strategies described in Sec. 7.4.2 can be fully expressed when coupled with congestion control mechanisms that regulate information distribution as a function of the network load and minimize the packet

collision probability. In Sec. 7.4.3.1 we present a benchmark congestion control mechanism, which we call Channel Sensing Congestion Control (CSCC). The CSCC scheme is based on the LIMERIC protocol [241], which will be reviewed in the following, and, like most state-of-the-art approaches, relies on channel sensing. We observe that channel sensing based mechanisms such as CSCC present several limitations, especially in highly dynamic scenarios. To address these issues, in Sec.7.4.3.2 we design an alternative congestion control approach, which we call Neighbor Aware Congestion Control (NACC), exploiting the network topology information to reduce the packet collision probability. We remark that such congestion control algorithms can be used with either broadcast strategy, though NACC has been designed for ETB and, hence, may underperform when combined with the PB strategy.

7.4.3.1 Benchmark: Channel Sensing Congestion Control (CSCC)

In the CSCC scenario, each vehicle constantly listens to the wireless channel and estimates the amount of resources that it is allowed to use to avoid congestion. We consider a vehicle $v \in V$, which is assigned to subcarrier $c_v \in \{0, 1, ..., n_{sc} - 1\}$. In each timeslot, v senses the channel and determines if a new transmission has been initiated. The fraction of time during which the channel is sensed busy in last N_{cbr}^{avg} timeslots is called Local Channel Busy Ratio ($\text{CBR}_{\text{local}}$). Every N_{cbr}^{update} timeslots the value of $\text{CBR}_{\text{local}}$ is smoothed as

$$\text{CBR}_{\text{vehicle}} = 0.5 \cdot \text{CBR}_{\text{vehicle}} + 0.5 \cdot \text{CBR}_{\text{local}}. \tag{7.5}$$

The output of (7.5) is called Vehicle Channel Busy Ratio ($\text{CBR}_{\text{vehicle}}$) and represents the channel occupancy sensed by v over the subcarrier c_v. In the CSCC approach, each network node aims at keeping the value of $\text{CBR}_{\text{vehicle}}$ as close as possible to a target value, which is called Target Channel Busy Ratio ($\text{CBR}_{\text{target}}$). Practically, every N_{cbr}^{update} timeslots, v evaluates the difference between $\text{CBR}_{\text{vehicle}}$ and $\text{CBR}_{\text{target}}$ and updates accordingly the values of ρ, which is the fraction of time that v can exploit to transmit over the wireless channel. We assume that v is adopting the PB strategy. Since ρ represents the percentage of time in which v is allowed to transmit, the value of T_{period} is updated as $T_{period} = \frac{1}{\rho}$. In case v is adopting the ETB strategy, the value of ρ should be associated to a specific error threshold E_{thr}.

7.4.3.2 New Proposal: Neighbor Aware Congestion Control (NACC)

In the CSCC approach, each vehicle computes the value of ρ as a function of its knowledge about neighbors' positions. In particular, vehicles can increase or decrease the channel occupancy with the aim of minimizing the packet collision probability. We start by theoretically modeling the communication that takes place in a group of vehicles when CSMA/CA is implemented at the MAC layer. Then, we describe how a user can estimate the number of neighbors that may potentially result in packet collisions. Finally, we find a relation between the vehicular density sensed by a user and the packet collision probability itself.

CSMA/CA Analysis We saw in Sec. 7.4.1.4 that vehicles access the channel following a 1-persistent CSMA/CA protocol. To efficiently model this communication scheme, we need to make some assumptions. We consider a population of N vehicles that share the same subcarrier $c_v \in \{0, 1, ..., n_{sc} - 1\}$, which is supposed to be reserved to this set of vehicles. First, we assume that all the vehicles are always mutually in-range, i.e., at a distance lower than r. Then, we assume that all the vehicles access the channel with the same transmission probability and that the resulting process can

be modeled as a PPP of parameter ρ. Under these hypotheses, we can describe the communication dynamics with an $M/D/1/N/N/random$ queuing system, where each arrival corresponds to a new channel access request and the queue contains all the requests that have not yet been fulfilled. In particular, the arrival process is Markovian (M), the service time is deterministic (D), there is only one server (1), the queue capacity is equal to the population size (N), and the queuing discipline is random (*random*), i.e., the customer to be served is chosen randomly from the queue in each timeslot.

We now define x_t, the number of channel access requests held in the queue at the end of timeslot t, and a_t, the number of new arrivals during t. Since the system will serve one request per timeslot, provided that $x_{t-1} + a_t > 0$, we have always $x_t \leq N - 1$. Furthermore, we have always $x_t = x_{t-1} - 1 + a_t$ after a non-empty slot and $x_t = 0$ after an empty slot. Now, given $x_{t-1} = i$, ρ and N, the probability of having a arrivals during t is equal to:

$$P(a_t = a | x_{t-1} = i, \rho, N) = \begin{cases} \binom{N-i}{a} \rho^a (1-\rho)^{N-i-a} & 0 \leq a \leq N-i, \\ 0 & a > N-i. \end{cases} \tag{7.6}$$

We observe that, if ρ and N are fixed, the channel dynamics at the end of any timeslot t are completely characterized by the number of users that need to transmit, i.e., the queue size x_t. Hence, we can describe the overall system by a Markov Chain, whose states x_t are in the set $X = \{0, \ldots, N-1\}$ and whose transition probability matrix $\mathbf{T}(\rho, N)$ is given by:

$$T_{i,j}(\rho, N) = \begin{cases} 0 & j < i-1, \\ P(a_t = j - i + 1 | x_{t-1} = i, \rho, N) & 0 < i < N \wedge i - 1 \leq j < N, \\ P(a_t = j + 1 | x_{t-1} = i, \rho, N) & i = 0 \wedge 0 < j < N, \\ P(a_t \leq 1 | x_{t-1} = i, \rho, N) & i = 0 \wedge j = 0. \end{cases} \tag{7.7}$$

Given $\mathbf{T}(\rho, N)$, we compute the steady state vector $\overline{\Pi}(\rho, N) = [\Pi_0(\rho, N), \Pi_1(\rho, N), ..., \Pi_{N-1}(\rho, N)]$. Then, we can compute the probability of different transmission events. In particular the probability that during a generic timeslot t no transmissions are initiated is given by

$$P(x_{t-1} = 0, a_t = 0 | \rho, N) = \Pi_0(\rho, N) \cdot P(a_t = 0 | x_{t-1} = 0, \rho, N). \tag{7.8}$$

Vehicle Position Distribution We recall that our objective is to minimize the number of packet collisions, which in our model are caused only by the hidden terminal problem. To compute the collision probability in the described scenario, we should estimate how many neighbors of the target receiver v_b can interfere. We denote this value by N_{ht}. Assuming that the vehicular density in the communication area of v_b is constant, we can estimate N_{ht} as

$$\hat{N}_{ht} = \frac{N_b + 1}{n_{sc}} \frac{E[\mathcal{A}(d)]}{\pi r^2}. \tag{7.9}$$

In (7.9), d is the distance between v_a and v_b, $\frac{N_b+1}{n_{sc}}$ is the estimate of the number of vehicles contained in the communication area of v_b that are using the same subcarriers of v_a, and $\mathcal{A}(d)$ is the area within the coverage of v_b but not of v_a. In other words, $\mathcal{A}(d)$ is the area from which a transmission would be hidden from v_a, possibly causing a hidden node collision. Let us define by $\Phi(d)$ the intersection of the communication areas of v_a and v_b, so that $\Phi(d) = \pi r^2 - \mathcal{A}(d)$. A mathematical expression for $\Phi(d)$ is given by

$$\Phi(d) = 2r \left(r \arccos\left(\frac{d}{2r}\right) - \frac{d}{2} \sqrt{1 - \left(\frac{d}{2r}\right)^2} \right). \tag{7.10}$$

Assuming that the geographical distribution of the nodes can space can be modeled as as a PPP, we obtain that the probability distribution of d is equal to

$$f_d(d) = \frac{2d}{r^2}. \tag{7.11}$$

Given (7.11) and (7.10), the mean value of $\Phi(d)$ can be computed as

$$E\left[\Phi(d)\right] = \int \Phi(d) f_d(d) dd = r^2 \left(\pi - \frac{3\sqrt{3}}{4}\right). \tag{7.12}$$

Recalling that $\Phi = \pi r^2 - \mathcal{A}(d)$, we can write

$$E\left[\Phi(d)\right] = r^2 \left(\pi - \frac{3\sqrt{3}}{4}\right) = \pi r^2 - E\left[\mathcal{A}(d)\right], \tag{7.13}$$

so that $E\left[\mathcal{A}(d)\right] = \frac{3\sqrt{3}}{4} r^2$. Replacing $E\left[\mathcal{A}(d)\right]$ in (7.9) we finally obtain the expression

$$\hat{N}_{ht} = \frac{N_b + 1}{n_{sc}} \frac{3\sqrt{3}}{4\pi}. \tag{7.14}$$

Packet Collision Probability We consider a vehicle v that is tracking $\hat{N}$ neighbors (the value of $\hat{N}$ depends on the overall system dynamics and may differ from N, which instead represents the true number of neighbors that are in the communication area of v). Suppose that v starts a new transmission during a generic timeslot t. On average there are $\hat{N}_{ht} = \frac{\hat{N}+1}{n_{sc}} \left(1 - \frac{\overline{\Phi}}{\pi r^2}\right)$ vehicles which can interfere with the communication. Hence, according to our channel model, the probability that the transmission will not fail corresponds to the probability that none of those $\hat{N}_{ht}$ interfering nodes transmits during t. If we assume that the considered $\hat{N}_{ht}$ vehicles have the same transmission probability ρ and do not interact with other network nodes during t, the packet collision probability P_{coll} can be derived from (7.8), obtaining

$$P_{coll}(\rho, \hat{N}_{ht}) = 1 - \Pi_0(\rho, \hat{N}_{ht}) P(a_t = 0 | x_{t-1} = 0, \rho, \hat{N}_{ht}). \tag{7.15}$$

To reduce the number of collisions, the vehicle v with N neighbors should have a transmission probability ρ such that P_{coll} equals a predetermined threshold P_{thr}. In other words, v chooses ρ so that the difference between P_{coll} and P_{thr} is minimized, which means

$$\rho = \arg\min_{\rho} \left(\left| P_{coll}(\rho, \hat{N}_{ht}) - P_{thr} \right| \right). \tag{7.16}$$

Following the NACC protocol, each vehicle v changes the value of ρ according to the vehicular density in its surroundings. In particular, in case the vehicle is using the PB strategy, the value of T_{period} is updated as $T_{period} = \frac{1}{\rho}$. We highlight that, by adjusting the value of T_{period} in this way, we violate the assumption regarding the distribution of the packet inter-transmission time considered in the definition of the system Markov model. Indeed, with the PB strategy, the time between two subsequent transmissions is constant rather than geometrically distributed, while in the ETB strategy scenario it depends on the position error evolution. This approximation may impair the performance of our congestion control mechanism. In particular, we expect to observe a significant performance reduction in the case of the PB strategy.

7.4.3.3 Implementing Congestion Control for the ETB Strategy

Both the CSCC and the NACC approaches improve the efficiency of the broadcasting strategies described in Sec. 7.4.3 by adapting the inter-transmission period to the deployment scenario. As stated previously, to combine a congestion control scheme with the ETB strategy, we have to relate the inter-transmission period to the error threshold. Practically, we need to build a map $\mathcal{F}$ such that the transmission period $T_{period} = \mathcal{F}(E_{thr})$ yields an average map estimation error close to E_{thr}. Unfortunately, the relation between E_{thr} and T_{period} is subject to multiple factors and cannot be easily modeled. $\mathcal{F}$ depends on how the position estimation error of vehicles evolves in time, i.e., on both the road map and the users' behaviors.

To reach our goal, we hence resorted to a pragmatic approach. By simulating a purely predictive UKF in the considered scenario, we derive an empirical estimate of the statistical distribution $P(e_h \leq E_{thr})$ of the position estimation error e_h after h timeslots since the last update, for any $h \geq 0$. Denoting by $\mathcal{H}$ the number of timeslots at which the error e_h exceeds the threshold E_{thr}, we can set $T_{period} = E[\mathcal{H}]T_t$, where T_t is the timeslot duration. Now, pretending the e_h are independent random variables, the complementary cumulative distribution function of $\mathcal{H}$ can be expressed as

$$P(\mathcal{H} > H) = \prod_{h=1}^{H} P(e_h \leq E_{thr}) \tag{7.17}$$

from which we easily get

$$T_{period} = T_t \sum_{H=1}^{\infty} \prod_{h=1}^{H} P(e_h \leq E_{thr}) \tag{7.18}$$

Equation (7.18) hence provides the desired map $\mathcal{F}$ from the error threshold E_{thr} to the inter-transmission period T_{period}. Such a function can also be used to determine the value ρ of the broadcast policy ETB, which can be computed as follows:

$$\rho = \frac{1}{\mathcal{F}(E_{thr})}. \tag{7.19}$$

We highlight this approach requires that vehicles know the distribution of the position estimation error in the map. In a realistic scenario, such information can be provided to vehicles by the road infrastructure, or pre-programmed into the channel access algorithm (possibly with multiple choices, depending on the road conditions). The investigation of such aspects, however, is left to future work.

7.4.4 Performance Results

In this section we evaluate the performance of the proposed ETB strategy for broadcasting operations, compared to a traditional PB approach. Moreover, we exemplify how the proposed NACC mechanism can improve the performance of the broadcasting strategies by exploiting network topology information, with respect to the benchmark CSCC scheme that relies only on channel sensing. The results of our study are derived through a Monte Carlo approach, where multiple independent simulations of duration $T_{\text{sim}} = 100$ s are repeated to get different statistical quantities of interest. The simulation parameters listed in Table 7.2 are based on a realistic urban CAV scenario.

7.4.4.1 Simulation Parameters

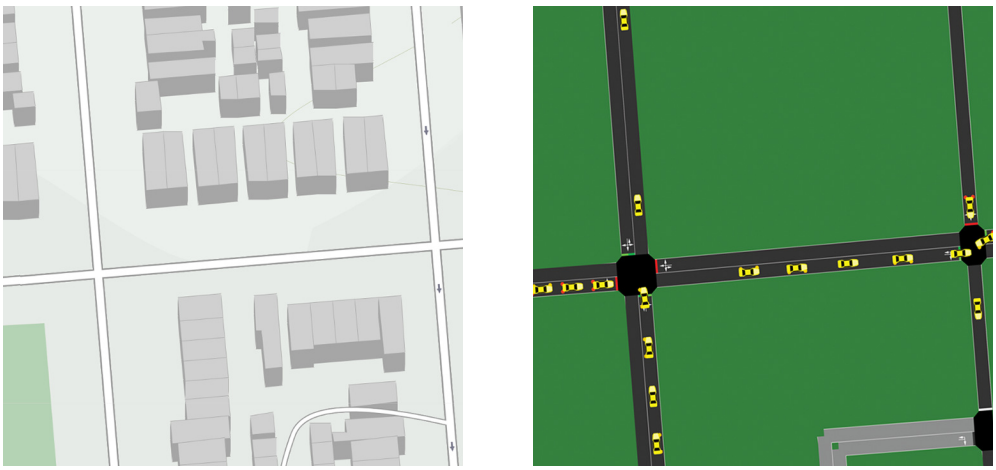
We use conservative IEEE 802.11p PHY and MAC layer parameter settings, which yield a maximum discoverable range of $r = 140$ m [320], while the communication delay is set to $T_d = 100$ ms, corresponding to one timeslot T_t. When not implementing a congestion control scheme, the settings of

Table 7.2: When to Broadcast? General parameters.

Parameter	Value	Description	Parameter	Value	Description
T_{sim}	100 s	Simulation duration	v_{max}	13.89 m/s	Max. speed
N_{sim}	20	Number of runs	d_0	42 m	Safety distance
T_t	100 ms	Timeslot duration	A_S	0.5168 km^2	Area size
T_d	100 ms	Communication delay	d_v	120 vehicles/km^2	Vehicular density
r	140 m	Communication range	$\|V\|$	62	Number of nodes
n_{sc}	{2, 4, 6, 8, 10}	Number of subcarriers	ρ_{max}	1	Upper bound of ρ
$n_{sc,tot}$	52	Max. number of subcarriers	Δ_{track}	10 s	Max. tracking duration
E_{thr}	$\{0, \dots 42\}$ m	Error threshold	ρ_{min}	0.0006	Lower bound of ρ
T_{period}	$\{0, \dots 10\}$ s	Inter-transmission period	δ_{max}	1	Upper bound of δ
q	1	Process noise parameter	δ_{min}	-1	Lower bound of δ
$R_{1,1}$	1.18535 m^2	Position accuracy along x	K	$\|V\|/n_{sc}$	Max. number of users
$R_{2,2}$	1.18535 m^2	Position accuracy along y	α	0.1	Speed parameter
$R_{3,3}$	0.5 (m/s)^2	Speed accuracy	β	$(2-\alpha)/K$	Convergence parameter
$R_{4,4}$	0.39 (m/s^2)^2	Acceleration accuracy	CBR_{target}	0.68	Target Channel Busy Ratio
$R_{5,5}$	0.09211 rad^2	Heading accuracy	P_{thr}	0.3	Collision probability threshold
$R_{6,6}$	0.01587 (rad/s) ^{2}	Turn rate accuracy	$\{A_\lambda, B_\lambda, C_\lambda, D_\lambda, E_\lambda, \nu\}$	{1, 0.05, 1, 1, 0, 0.2}	Logistic function params

both the PB and ETB strategies must be defined *a priori*. In our simulations we adopt an exhaustive approach and consider $N_{\text{set}} = 30$ different settings. In particular, we make the inter-transmission period T_{period} vary from 0 to 10 s while the error threshold E_{thr} ranges between 0 and 42 m. Each choice involves a different trade-off between estimation accuracy and broadcasting overhead.

For our simulations, we use real road map data imported from OpenStreetMap (OSM), an open-source software which combines wiki-like user generated data with publicly available information. In particular, we consider the OSM map of New York City, as represented in Fig. 7.3a, so that to characterize a dynamic urban environment. In order to consider realistic mobility routes that are representative of the behavior of vehicles in the road network, we simulate the mobility of cars using SUMO, as represented in Fig. 7.3b. The vehicles move through the street network according to a `randomTrip` mobility model, which generates trips with random origins and destinations, and speeds which depend on the realistic interaction of the vehicle with the road and network elements. The



(a) Openstreetmap scenario. **(b)** SUMO scenario.

Fig. 7.3: Representation of portion of the urban map considered for the performance evaluation.

maximum speed is set to $v_{max} = 13.89$ m/s, which is consistent with current speed limits. Given v_{max}, we set $d_0 = 42$ m, which corresponds to the distance traveled in 3 s by a vehicle running at the maximum speed. In this way, d_0 represents the maximum *safety distance* that should be held in an urban scenario. Following the work of [195], we consider a vehicular density of $d_v = 120$ vehicles/km^2 for medium traffic conditions. Given the total road map area of A=0.5168 km^2, the number of vehicles deployed in the considered scenario is $|V| = 62$.

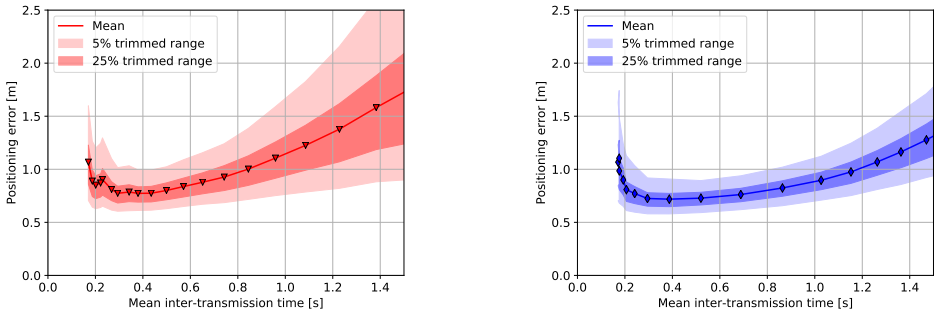
As we assessed in Sec. 7.4.1, the behavior of each vehicle can be fully represented by its state $s(t)$. Measurements of the components of $s(t)$ are affected by a non-negligible noise which is modeled as a Gaussian process with zero mean and covariance matrix R. The diagonal elements of R are given in Table 7.2 and are derived from the models in [262, 321, 322]. We define $Q = qI$, where q is the noise covariance parameter and I denotes the identity matrix. Table 7.2 also reports the parameters of the congestion control schemes from Sec. 7.4.3. For what concerns the CSCC approach, we use the parameters suggested in [241]. In particular, we set $\alpha = 0.1$, which ensures a sufficiently high convergence speed, and $\beta = (2 - \alpha)/K$, so that the algorithm convergence is guaranteed for any K. We observe that K represents the maximum number of users sharing the same communication channel that, in our scenario, is on average $|V|/n_{sc}$. Finally, we set $\text{CBR}_{\text{target}}$ to 0.68, so that vehicles aim at occupying the channel about 68% of the time. For what concerns our proposed NACC approach, we set the collision probability threshold to $P_{coll} = 0.3$.

To evaluate the performance of the proposed broadcasting strategies in the simulations, we take into account four main factors, namely:

- Average positioning error, i.e., the average error of the ego vehicle when estimating its own position and that of its neighbors, which is given by (7.1);
- 95th percentile of the positioning error, i.e., the positioning error that only 5% of the vehicles cannot achieve;
- Detection error, i.e., the sum of the misdetection (i.e., unknown vehicles in the ego vehicle communication area) and false detection (i.e., vehicles that are believed to be in the neighborhood but are actually beyond the communication range) event probabilities;
- Packet collision rate, i.e., the average number of packet collisions per vehicle and per second that occur because of the hidden terminal problem.

7.4.4.2 Simulation Results

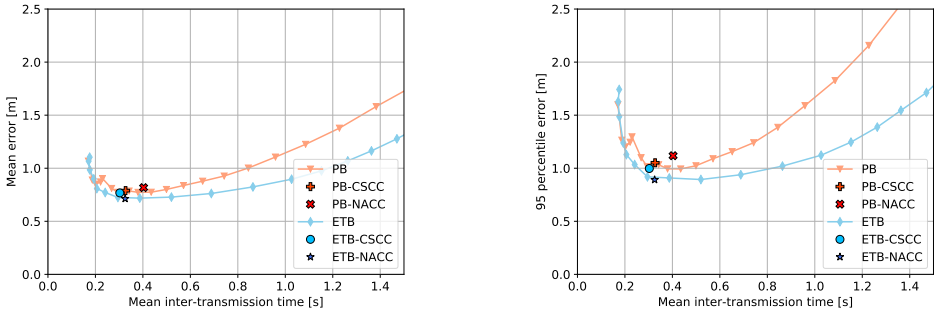
We now analyze the performance of the broadcasting strategies and the congestion control schemes that we described in Sec.7.4.3 and Sec.7.4.3, respectively. At first, we fix the number of the available subcarriers to $n_{sc} = 8$. Later, we will verify how different n_{sc} values may influence the simulation outcomes. As we already stated, if we do not implement a congestion control mechanism, we have to determine *a priori* the inner setting of the PB and the ETB strategies. To fairly compare the performance of the two techniques, we adopt an exhaustive approach, obtaining a different outcome for each choice of T_{period} and E_{thr}. In Fig.7.4, we analyze the statistics of the positioning error according to the mean inter-transmission time $\overline{T}_{tx}$, which is an indicator of the total channel occupancy. We highlight that $\overline{T}_{tx}$ does not coincide with the inter-transmission period used in the PB strategy. Indeed, while T_{period} is defined *a priori* and can assume all the values within the set $\{0.1\text{ s}, ..., 10\text{ s}\}$, $\overline{T}_{tx}$ is an outcome of the simulation. In particular, in a realistic scenario, $\overline{T}_{tx}$ never goes below the value of 0.2 s, i.e., two timeslots, because for the channel access contention. From Fig. 7.4, we can also observe how the limits of the CSMA/CA affect the positioning error: when the number of channel access requests is too high, i.e., $\overline{T}_{tx} < 0.3$ s, the channel gets congested and, consequently, the performance



(a) PB strategy.

(b) ETB strategy.

Fig. 7.4: Comparison of the positioning error statistics as a function of the average inter-transmission time and the broadcasting strategy, with $n_{sc} = 8$ and without channel access control.



(a) Mean error.

(b) 95th percentile of the error.

Fig. 7.5: Comparison of the positioning error statistics as a function of the average inter-transmission time, with $n_{sc} = 8$. The behavior of the PB and ETB strategies is compared, in combination with CSCC and NACC schemes for congestion control.

of the overall system degrades. Indeed, the positioning error can be described by a convex curve, with a minimum for $\overline{T}_{tx} \approx 0.3$ s; this value represents the level of channel occupancy that guarantees the best position estimation accuracy. By comparing Fig.7.4a with Fig.7.4b, we can observe that the ETB strategy outperforms the benchmark PB strategy. In particular, ETB proves to have a slightly lower average error and a significantly lower error variance.

By optimizing the channel access requests, the ETB strategy has a lower positioning error than the benchmark strategy. Fig.7.5 shows a direct comparison between the considered broadcasting strategies; in particular, Fig.7.5a reports the mean error while Fig.7.5b shows the 95th percentile of the error. In both cases, the ETB strategy ensures better position estimation accuracy for the same level of channel occupancy. The marks in Fig. 7.5 represent the performance of the congestion control schemes designed in Sec.7.4.3.[2] First, we observe that all the deployed solutions succeed in maintaining the channel occupancy close to the optimal working point, i.e., $\overline{T}_{tx} \approx 0.3$ s. Among all the possible solutions, the combination of the ETB strategy with the NACC approach ensures the best performance. In particular, this scheme outperforms the classical approach used in the literature, which is represented by the combination of the PB strategy with the CSCC approach, obtaining a

[2]Since congestion control can adapt the communication strategy to the scenario in real-time, we obtain a single outcome for each combination of broadcasting strategy and congestion control approach.

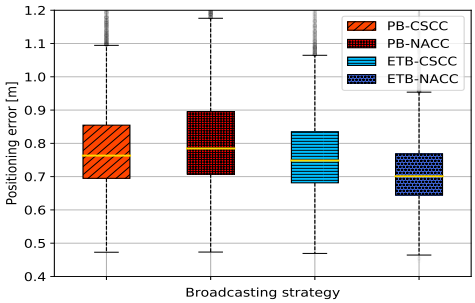
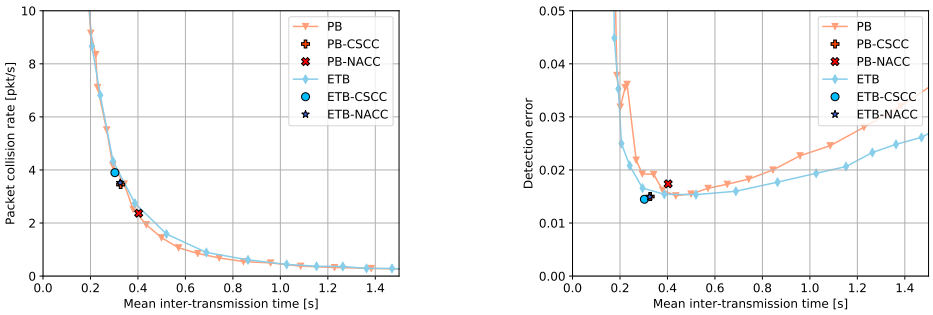


Fig. 7.6: Boxplot of the positioning error with $n_{sc} = 8$. The behavior of the PB and ETB strategies is compared, in combination with CSCC and NACC schemes for congestion control.



(a) Packet collision rate.

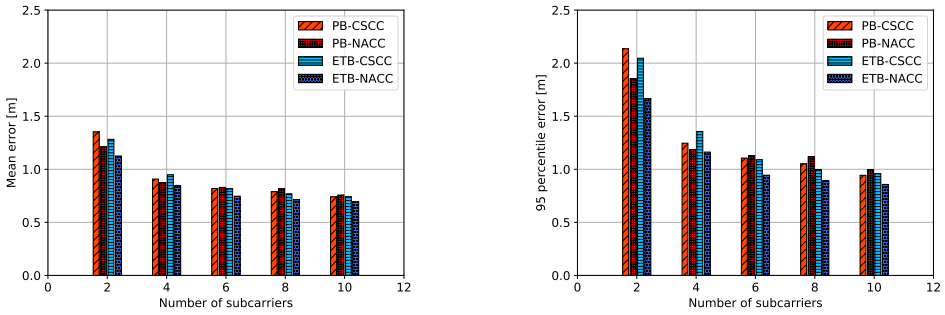
(b) Detection error probability.

Fig. 7.7: Collision and detection statistics as a function of the average inter-transmission time, with $n_{sc} = 8$. The behavior of the PB and ETB strategies is compared, in combination with CSCC and NACC schemes for congestion control.

10% gain when considering the mean error and 20% gain when considering the 95th percentile of the error.

The full positioning error statistics of the four congestion control solutions are shown as a boxplot in Fig.7.6. In the figure, each box is delimited by the first and the third quartiles of the error distribution. The box's center lines represent the median of the error, and the whiskers show the 95% confidence intervals. Outliers are represented as dots. We see that our solution is the only technique that ensures that the third quartile is below 0.8 m and that the confidence interval is below 1 m.

In Fig.7.7 we show the packet collision rate and the detection error probability. Taking into account the broadcasting strategies without the congestion control schemes, we observe that both techniques present almost identical trends. As we can observe from Fig.7.7a, the amount of information that gets lost in the channel significantly increases when $\overline{T}_{tx} < 0.5$ s, independently of the deployed strategy. This phenomenon explains the degradation of the positioning estimation accuracy that we observe in Fig.7.4. We highlight that the channel congestion does not affect only the positioning error but also the probability of misdetection or false detection of a neighbor vehicle. Indeed, by looking at Fig.7.7b, we can observe that the detection error probability increases exponentially as soon as $\overline{T}_{tx} < 0.25$ s. Since the strategies optimal working point is $\overline{T}_{tx} \approx 0.3$, we conclude that minimizing the positioning error does not necessarily imply an increase of the detection error probability. Considering the congestion control approaches, we observe that none of the obtained outcomes deviate from the curves defined by the exhaustive simulations. As already mentioned, the combination of the



(a) Average error.

(b) 95th percentile of the error.

Fig. 7.8: Positioning error as a function of the subcarrier number. The behavior of the PB and ETB strategies is compared, in combination with CSCC and NACC schemes for congestion control.

PB strategy with the NACC approach presents a slightly higher $\overline{T}_{tx}$ and, therefore, is characterized by a different packet collision rate and detection error probability.

In order to validate these results in a more general scenario, we analyze the performance of the four possible congestion control schemes with a different numbers of subcarriers n_{sc}. The results of this analysis are reported in Fig.7.8. We observe that the solution combining the ETB strategy and the NACC approach outperforms the other schemes for any value of n_{sc}, considering both the mean positioning error (Fig.7.8a) and the 95th percentile of the positioning error (Fig.7.8b). In particular, our solution outperforms state of the art solution by up to 20% mean error reduction and up to 30% 95th percentile error reduction. For what concerns the other techniques, we observe that the combination of ETB and CSCC approach performs poorly for $n_{sc} \leq 4$, while it leads to better results when the number of subcarriers is greater. Conversely, the combination of PB and NACC approach, performs well for $n_{sc} \leq 4$ but does not fully exploit the available resources when $n_{sc} \geq 6$.

7.5 What to Broadcast?

The large volume of data generated by automotive sensors will likely saturate the capacity of vehicular communication technologies, making it challenging to guarantee the required quality of service. The broadcast decision must therefore depend on the type of information that the vehicle has acquired, and the information already available at the receiver, so that to prioritize the transmissions that have the greatest importance for the target applications. Along these lines, in this study we investigate what information to broadcast in such a way to prevent the dissemination of redundant, duplicate and/or useless information. To do so, we propose and evaluate a framework that uses analytic hierarchy multicriteria decision processes to predict VoI based on space, time, and quality attributes and which depend on the receiver's context and application.

7.5.1 System Model

We consider two possible embodiments of VoI assessment: processed and non-processed approaches. The trade-off involves latency, energy consumption and VoI accuracy. In the first case, the perception record is analyzed by the sender to extract context information (e.g., estimated positions of objects in a captured image) before being broadcast. While incurring in some processing delays, this allows the sender to validate the integrity of the observation and determine whether it embeds valuable

characteristics for the potential receiver(s). In the second case, the perception record is broadcast immediately after it is generated, thereby yielding more responsive value assessment operation. The sender, however, needs to predict *probabilistically* the value of the observation to prevent circulation of duplicate or redundant data.

Moreover, among the applications, information sources and vehicular attributes we presented in Sec. 7.3, we focus on the following subsets of elements that, for their generality and complementarity, we believe are good representatives of future vehicular systems.

- **Applications** We focus on *advanced safety* and *traffic management* applications.
- **Information Sources** As CAVs evolve towards the support of safety-critical applications, it is fundamental to implement network architectures that guarantee timely and accurate positioning of vehicles. Positioning is typically provided by GPS (which also guarantees accurate time-synchronization among vehicles), although other localization techniques, e.g., based on image processing, can be useful to improve accuracy. We consider camera sensors as the principal information source to enable position estimation. The accuracy of the camera observations depends on (i) the resolution of the sensor, which is a measure of the image width/height and the frame rate, (ii) the field of view, and (iii) the operating distance.
- **Attributes** We focus on the following attribute categories: *(1) timeliness* (the VoI decreases with the relative age of information normalized to its lifetime), *(2) proximity* (the VoI is a function of (i) the distance between the information source and destination, i.e., sensory data generated by close-by vehicles are generally more valuable than data coming from farther nodes, and (ii) the scenario, i.e., urban/highway), and *(3) information quality* (the VoI depends on (i) the intrinsic quality of the information source, which may be assessed in terms of sensor resolution, and (ii) the distance between the source sensor and the observation – e.g., the depth measurement error increases proportionally with the distance).

7.5.2 What to Broadcast? Proposed Strategy

The proposed framework performs VoI assessment operations through three main phases, as illustrated in Fig. 7.9 and described next.

7.5.2.1 Phase 1: Attribute Priority Weights (via AHP)

First, the algorithm applies the AHP [311] to derive the relative degree of priority among the vehicular attributes, i.e., timeliness, proximity and quality, by populating a pairwise comparison matrix M (as illustrated in the left frame of Fig. 7.9) with comparison scores (i.e., α, β, γ). The comparison scores in M (ranging from 1/9 to 9) are assigned according to the Saaty comparison scale [323] and assess the importance of the attributes in the row relative to those in the column (e.g., the score 3 is assigned if the item on the row is "moderately more important" than the item on the column in the specified application domain). Note that M is reciprocal by construction, i.e., $M(j,k) = 1/M(k,j)$, $\forall j,k \in \{1,\ldots,n\}$, where n is the size of M, i.e., the number of attributes.

As soon as the comparison scores have been determined, priority weights w_a, $a = 1,\ldots,n$, are computed evaluating the normalized principal eigenvector $\vec{w} = \langle w_1,\ldots,w_n \rangle$ of M, i.e., the eigenvector that corresponds to the eigenvalue $\lambda_{\max}$ with the largest magnitude:

$$M\vec{w} = \lambda_{\max}\vec{w}. \tag{7.20}$$

The priority weights indicate how valuable each attribute is compared to the others. It appears clear that the AHP method determines relative (instead of absolute) priority weights, which are based on

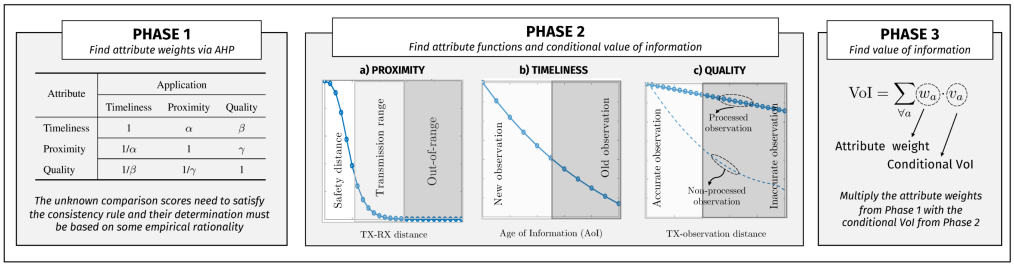


Fig. 7.9: Illustrative scheme of the VoI assessment framework proposed in Sec. 7.5.2.

empirical evaluation criteria, and consequently have a certain degree of arbitrariness. According to the AHP, in order for the weight vector $\vec{w}$ to be a good representative of the relative importance of the attributes, the matrix M should be consistent, i.e., such that $M(j,k) = M(h,k)/M(h,j)$, $\forall h,j,k \in \{1,\dots,n\}$. However, given the arbitrariness in the attribute selection, the matrix M is usually non consistent. A measure of the matrix consistency is given in [311] in terms of the so-called *consistency index* $C_I = (\lambda_{\max} - n)/(n-1)$. Based on [311], the weight vector $\vec{w}$ can be considered acceptable if M satisfies the following consistency rule:

$$C_R = \frac{C_I}{R_I(n)} = \frac{(\lambda_{\max} - n)/(n-1)}{R_I(n)} < 0.1, \tag{7.21}$$

where R_I is the average of the C_Is obtained by randomly generating reciprocal matrices of size n (for $n = 3$, we get $R_I = 0.58$).

7.5.2.2 Phase 2: Conditional VoI

The framework is now used to assess how the VoI evolves, conditioned to each attribute. Considering the $n = 3$ attributes presented in Sec. 7.3.3, different VoI functions are defined.

a) Proximity: For the proximity attribute, we propose to use a logistic function to model the VoI dependency on the distance d between the information source and the destination, as represented in the first box of the middle frame of Fig. 7.9. The logistic function is always monotonically decreasing in d, but tuning the function's parameters it is possible to move from a smooth and quasi-linear decrease of the VoI with d to a step-like behavior, where the VoI is almost constant within a certain range from the source and suddenly drops to zero beyond that range. The function is given by:

$$v_1 = A + \frac{K - A}{\left(C + Qe^{-B(d-d_s)}\right)^{1/\nu}}. \tag{7.22}$$

Parameters in Eq. (7.22) characterize the logistic function's shape and their values will be detailed in Sec. 7.5.3. In particular, A is the lower asymptote, K is the upper asymptote, B is the growth rate (which is selected so that $v_1 \to 0$ beyond the communication range), and d_s represents the safety distance that must be held and determines the threshold beyond which v_1 starts decreasing. d_s is therefore a function of the scenario, i.e., urban/highway, in which the vehicles are deployed.

b) Timeliness: For the timeliness attribute, we propose to use an exponential function [307], as represented in the second box of the middle frame of Fig. 7.9, i.e.,

$$v_2 = \exp(-P_{\text{td}} \cdot (t - t_0)), \tag{7.23}$$

where $(t - t_0)$ represents the AoI, and P_{td} is the temporal decay parameter. In particular, $P_{\text{td}} = 0$

indicates that the information is not delay sensitive, while a large P_{td} models a quick obsolescence of the information.

c) Information Quality: We propose that the VoI for the quality attribute evolves as a function of the distance d_o between the source sensor and the perceived observation, as represented in the third box of the middle frame of Fig. 7.9. We distinguish between processed (p) and non-processed (np) VoI assessment operations. In the processed case, the VoI depends only on the characteristics of the sensor [324]. With reference to the camera sensor, we get

$$v_3^p = 1 - \frac{d_o}{h \cdot f_d}, \tag{7.24}$$

where h is the height of the sensor, and f_d is the focal distance. In particular, f_d depends on the camera image resolution in the horizontal domain (r_h) and the field of view (f_w) in degrees, so that

$$f_d = \frac{r_h/2}{\tan(f_w/180 \cdot \pi/2)}. \tag{7.25}$$

In the non-processed case, the sender does not know the content of the perception record, therefore it has to identify a method to probabilistically predict whether such record embeds valuable information. We assume that a certain observation (e.g., object) can be detected if it is in LOS with respect to the sensor's field of view. The LOS probability is a function of d_o and the propagation scenario, and is modeled as in [31]:

$$P_{\text{LOS}}(d_o) = \begin{cases} \min\{1, 1.05e^{-0.0114 d_o}\} & \text{if urban} \\ \min\{1, (2.1 \cdot 10^{-6}) d_o^2 - 0.002 d_o + 1.02\} & \text{if highway} \end{cases} \tag{7.26}$$

The conditional VoI is finally computed as

$$v_3^{np} = \left(1 - \frac{d_o}{h \cdot f_d}\right) \cdot P_{\text{LOS}}(d_o). \tag{7.27}$$

7.5.2.3 Phase 3: Overall VoI

Finally, the framework assigns the value of information v by multiplying the attribute weights w_a, $a = 1, \ldots, n$, from Phase 1 with the conditional VoI v_a, $a = 1, \ldots, n$, from Phase 2:

$$v(d, t, d_o) = \sum_{a=1}^{n} w_a \cdot v_a. \tag{7.28}$$

A *data scheduler* finally sorts the information products in a descending order of values and sequentially forwards them to the surrounding receivers. The scheduler may also cancel transmissions of information whose value is below a pre-defined threshold θ_v.

7.5.3 Performance Results

In this section we validate the technical soundness of our proposed framework in target use cases. Our results can be used as a basis for evaluating the optimal data scheduling strategy that maximizes the utility of the transmitted information for the final receiver(s).

Table 7.3: What to Broadcast? General parameters.

Parameter	Value	
h	1.2 m	Sensor height
f_w	70 deg	Camera field of view
d	$\{1, \ldots, 500\}$ m	TX/RX distance
d_o	$d/2$	TX/obs distance
$t - t_0$	$\{0, \ldots, 5\}$ s	Age of information
d_s {urban, highway}	{24,72} m	Safety distance
P_{td} {static, variable, dynamic}	{0,1,10}	Temporal decay
{A,K,C,Q,B,V}	$\{1, 0, 1, 1, 0.03, 0.2\}$	Logistic function params
r_h {low, medium, high}	{640, 1280, 4096} px	Camera resolution

Table 7.4: Pairwise comparison matrices M_a and weights $\vec{w} = \langle w_1, w_2, w_3 \rangle$ of VoI attributes for safety and traffic management applications.

Attribute	Application: Safety			
	Timeliness	Proximity	Quality	Weight $\vec{w}$
Timeliness	1	1/7	1	0.1194
Proximity	7	1	5	0.7471
Quality	1	1/5	1	0.1336

Attribute	Application: Traffic Management			
	Timeliness	Proximity	Quality	Weight $\vec{w}$
Timeliness	1	9	3	0.6554
Proximity	1/9	1	1/7	0.0549
Quality	1/3	7	1	0.2897

7.5.3.1 Performance Parameters

The system parameters are based on realistic design considerations and are summarized in Table 7.3. For the proximity attribute, we calculate the safety distance d_s between vehicles as $d_s = 2 \cdot v_{\max}$ [325], where $v_{\max}$ is the speed limit, which we set to 12 m/s and 36 m/s in urban and highway scenarios, respectively. We also let the distance d between the sender and the receiver vary from 1 to 500 m. For the timeliness attribute, we consider static (e.g., fixed road construction), variable (e.g., temporary social/political events), and dynamic (e.g., pedestrian crossing the street) observations, so that $P_{\text{td}} =$ 0, 1, 10, respectively. We let the AoI parameter $t - t_0$ vary from 0 to 5 s. For the quality attribute, we consider low-, medium- and high-quality sensors, which are modeled as 640×480, 1280×720, and 4096×780 pixel cameras, respectively. The sensor is placed at a distance $h = 1.2$ m from the road surface, and the field of view is set to $f_w = 70$ degrees. We also assume that the target observation is placed at distance $d_o = d/2$ from the camera sensor. We recall that, in order to exemplify the approach, in this study we focus on the evaluation of the VoI for the position data provided by cameras.

7.5.3.2 Performance Results

Phase 1 results – *attribute weights.* In Table 7.4 we report the pairwise comparison matrices M which assess the interdependencies among the considered VoI attributes.[3] For instance, for

[3]Notice that the comparative scores we consider in Table 7.4 are chosen in such a way that the consistency rule defined in Eq. (7.21) is satisfied. Other combinations of scores can be also considered,

safety applications, we chose to set the proximity vs. timeliness score to 7 since we deem extremely important for vehicles to monitor space while broadcasting context information (e.g., vehicles need to know when the neighbors' distance falls below the safety-critical threshold to trigger collision avoidance transmissions and, at the same time, should defer or cancel transmissions relative to spatially far vehicles). For traffic management applications, we decided to set the proximity vs. timeliness score to 1/9 since the broadcasting decision does not have to be space-dependent (i.e., LDM updates should be addressed to both spatially close and far neighbors). Attribute weights $\vec{w} = \langle w_1, w_2, w_3 \rangle$ are determined from Eq. (7.20) and demonstrate that the dissemination of space-related information is very valuable to safety services ($\max_{\vec{w}} = w_2 = 0.7471$) while, for traffic management services, time-related information should be preferred ($\max_{\vec{w}} = w_1 = 0.6554$).

Phase 2 results – *conditional VoI.* In Fig. 7.10a we plot the conditional VoI for the proximity attribute as defined by the logistic function in Eq. (7.22). We can see that the value of position information is generally higher for a highway scenario than for an urban one. This is consistent with the fact that the larger safety distance between vehicles in highway scenarios requires the information to be disseminated over longer ranges, in order to reach the nearby vehicles. Furthermore, the highway scenario is usually characterized by better propagation conditions that increase the probability of successful reception at long distances and, consequently, the value of the packet transmission itself. Accordingly, the VoI drops to zero beyond 300 m, which thus represents a suitable communication range for vehicular networks. In Fig. 7.10b we plot the conditional VoI for the timeliness attribute as follows from the exponential function in Eq. (7.23), which is proportional to the AoI and the temporal characteristics of the perceived object, as defined in Sec. 7.4.1. In particular, the value is constant in case of static observations while, for dynamic observations, decreases at a pace that is a function of P_{td}. In Fig. 7.10c and Fig. 7.10d we plot the conditional VoI for the quality attribute, considering both processed and non-processed value assessment strategies, respectively. In the non-processed case the conditional VoI exhibits a significant difference between urban and highway scenarios. This gap is due to the higher probability that the line of sight is blocked in urban scenarios, in which case the image captured by the camera would be basically useless.

Phase 3 results – *overall VoI.* Our goal is to assign a value to different sources of information (in this study we consider camera observations) in such a way that the utility to potential receiver(s) is maximized. In Fig. 7.11 we evaluate the impact of the propagation scenario on the VoI. First, we observe that, for safety applications, the value of data transmission at short distances is high in all considered conditions, reflecting the importance of maintaining fresh and updated information

as long as they are selected in a way that guarantees that the assigned attribute interdependencies are fully representative of the characteristics of the application under consideration [273].

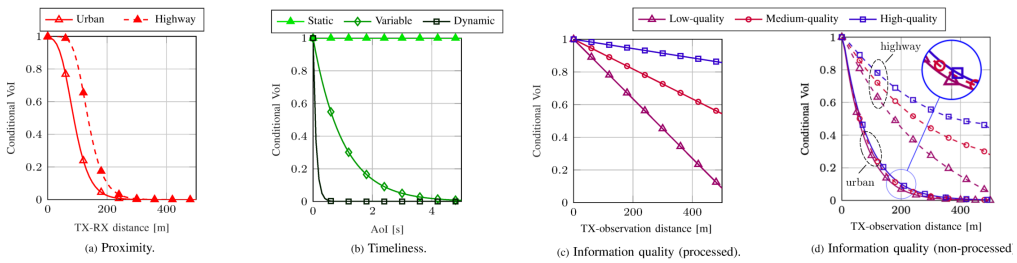
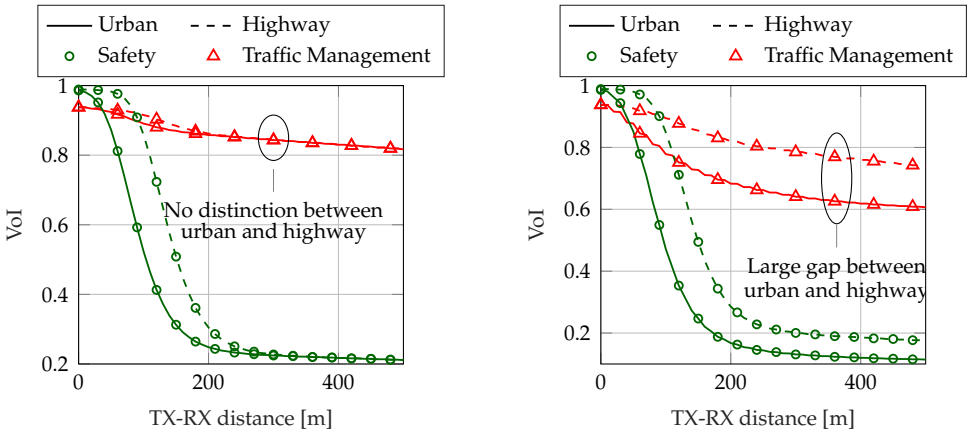


Fig. 7.10: Conditional VoI for space proximity, timeliness and information quality attributes. The plots show the impact of the TX-RX distance, the AoI, the TX-observation distance, the propagation scenario, the type of observation and the type of sensor.



(a) Processed case.

(b) Non-processed case.

Fig. 7.11: VoI vs. TX-RX distance for urban and highway scenarios for processed and non-processed strategies. $P_{\text{td}} = 1$, $r_h = 1080$ px, AoI $= 100$ ms.

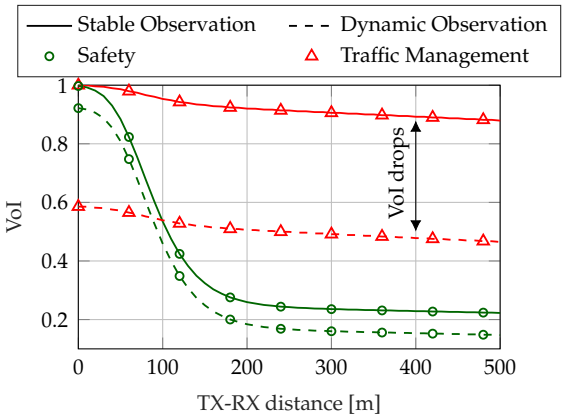
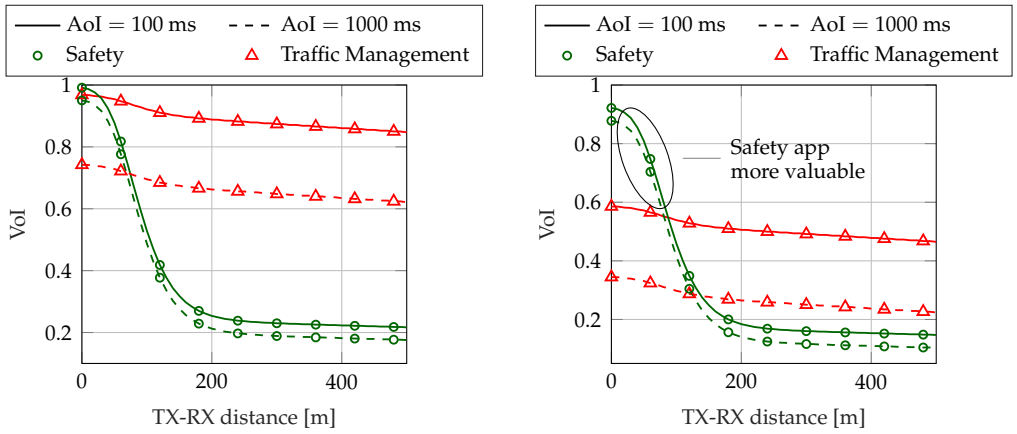


Fig. 7.12: VoI vs. TX-RX distance for different types of observations. Processed VoI operations, $r_h = 1080$ px, AoI=100 ms, urban scenarios are considered.

among close-by vehicles. Second, we see that, for traffic management services, the VoI is almost independent on the TX/RX distance (proximity weight $w_2 = 0.0549 \ll 1$), demonstrating the importance of sharing LDM updates even at large distance. Third, we notice that, for safety services, the VoI drops for values of d larger than 100 m, which is a rather safe inter-vehicle distance (in particular, in urban scenarios): transmitting data beyond this range would just increase the channel access contention without bringing much benefit in terms of safety. We can also observe that, in general, the VoI for the considered information is higher in highway than in urban scenarios, because of the higher probability of LOS between object and sensor (camera). In general, the non-processed case (Fig. 7.11b), although guaranteeing real-time value determination, represents a lower bound for the VoI in vehicular networks, as it provides a probabilistic, rather than deterministic, method to assess the VoI. Moreover, for the non-processed case, different characterizations of the quality attribute in urban and highway scenarios (see Fig. 7.10d) result in different overall VoI: for traffic management applications, the gap is as large as 25% when $d > 200$ m, i.e., when the endpoints are in NLOS.

The following results are derived considering processed VoI assessment operations. In Fig. 7.12 we investigate how the VoI evolves as a function of the type of observation. We observe that dynamic



(a) Static observations.

(b) Dynamic observations.

Fig. 7.13: VoI vs. TX-RX distance and AoI, for different types of observation. Processed VoI operations, $r_h = 1080$ px, urban scenarios are considered.

information is likely to have value for safety applications (although the VoI eventually drops to zero at large distance): the gap between static and dynamic, i.e., short-lived, observations is less than 10% for $d < 100$ m. Conversely, the large impact of the timeliness attribute in traffic management operations (the weight is 0.6554) makes the VoI decrease as much as 65% considering dynamic (as opposed to static) observations. This makes sense as time-varying perception records might become obsolete by the time they are actually transmitted.

In Fig. 7.13 we plot the VoI as a function of the AoI and the sensor reading's temporal characteristics. Ideally, we would like information to be received as timely as possible, i.e., at the same instant it was generated by the source. However, real-world constraints, first and foremost the restricted network capacity, put a limit on the frequency at which status updates can be broadcast, thereby making the AoI larger than the inter-transmission time. For safety services, both "new" (i.e., AoI = 100 ms) and "old" (i.e., AoI = 1000 ms) information provide comparable value to the receiver(s), as long as short-range communications are considered. The reason is that even old perception records may still potentially increase the opportunity for vehicles to make object detections, a critical pre-requisite for safety-related operations. For traffic management services, instead, the impact of the AoI is very strong. "Old" information (i.e., AoI = 1000 ms) may, in fact, decrease the VoI by more than 70% considering dynamic observations (Fig. 7.13b). To avoid the decay of information, such systems require very frequent updates to be disseminated through inter-vehicular communications, possibly causing, however, channel access problems. We notice that, for the dynamic case, the VoI considering traffic management applications is below 0.6 even at short distance: at $d = 10$ m, VoI = 0.58 vs. VoI = 0.92 for safety applications, thereby validating the results in Fig. 7.12.

The patterns we observed in the previous plots can be recognized also in Fig. 7.14, which illustrates the VoI evolution for different types of sensors. We see that the VoI increases proportionally to the camera resolution, even though the effect of the sensor quality is not very significant: the gap between high- and low-quality sensor readings is below 15%. This is consistent with the outcomes of the AHP comparative evaluations, which assign very low priority weight to the quality attribute. In fact, although the network requires context information to be reliable, it still needs to prioritize timely dissemination to spatially close neighbors to prevent communication failures. Nevertheless, we observe that sensor quality degradation has a bigger impact on traffic management than on safety services (i.e., $w_3 = 0.2897$ vs. $w_3 = 0.1336$, respectively). Finally, Fig. 7.14 acknowledges the higher

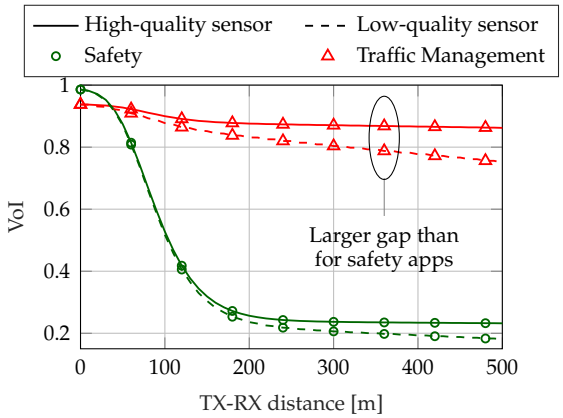


Fig. 7.14: VoI vs. TX-RX distance for different types of sensor. Processed VoI operations, $P_{\text{td}} = 1$, AoI $= 100$ ms, urban scenarios are considered.

VoI for safety operations compared to traffic management at short distance.

7.6 Conclusions and Design Guidelines

Assigning VoI is fundamental to discriminate the importance of the different information sources, in order to prevent the overload of transmission links. In this chapter, we characterized VoI in vehicular networks and investigated data broadacsting methods to tackle capacity issues. In particular, we answered two fundamental questions.

When to Broadcast Information? We studied the trade-off between ensuring accurate position information and preventing congestion of the communication channel in vehicular networks and designed an innovative threshold-based broadcasting algorithm that forces vehicles to distribute state information if the estimated positioning error is above a certain error threshold. We also adopted a new congestion control mechanism that adapts the inter-transmission period according to network topology information. We showed through simulations that the proposed approach outperforms a conventional broadcasting strategy, which relies on a periodic transmission of state information and channel sensing, since it reduces the positioning error with no additional resources.

What Information to Broadcast? We proposed a method that quantifies the expected VoI based on time, space and quality dependencies. The goal is to identify the transmission(s) that maximize the utility for potential receiver(s) while avoiding the overload of the communication channel. We evaluated the impact of the operating distance, the type of observation, the type of sensor, the propagation scenario and the AoI on the value assessment. Moreover, we numerically showed the rate at which VoI decreases considering obsolete, time-varying and inaccurate observations.

Part III

Towards 6G Networks

Chapter 8

Towards 6G Networks: Use Cases and Technologies

8.1 Introduction

From 1G to 5G, passing through Universal Mobile Telecommunication Systems (UMTS) and LTE, each generation of mobile technology has been designed to meet the needs of end users and network operators, as shown in Fig. 8.1. However, nowadays societies are becoming ever more data-centric, data-dependent and automated. Radical automation of industrial manufacturing processes will drive productivity. Autonomous systems are hitting our roads, oceans and air space. Millions of sensors will be embedded into cities, homes and food production environments, and new systems operated by artificial intelligence residing in local 'cloud' and 'fog' environments will create a plethora of new applications. Communications networks will provide the nervous system of these new smart system paradigms. The demands, however, will be daunting. Networks will need to transfer much greater amounts of data, at much higher speeds. Connections will move beyond personalized communication to machine-type communication, connecting not just people, but also data, computing resources, vehicles, devices, wearables, sensors and even robotic agents.

5G made a significant step towards developing a low latency tactile access network, by providing new additional wireless nerve tracts through (i) new frequency bands (e.g., the mmWave spectrum), (ii) advanced spectrum usage and management, in licensed and unlicensed bands, and (iii) a complete redesign of the core network. Yet, the rapid development of data-centric and automated processes may exceed even the capabilities of the emerging 5G systems. The above discussion has recently motivated researchers to look into a new generation of wireless systems, i.e., 6th generation (6G) systems, to meet the demands for a *fully connected, intelligent digital world* [326].

8.1.1 Motivations and Chapter Structure

Along these lines, the broad purpose of this chapter is to understand how future 6G systems can be developed to provide ever more capable and vertical-specific wireless network solutions. Specifically, Sec. 8.2 considers several potential applications for future connected systems and attempts to estimate their key requirements in terms of throughput, latency, connectivity and other factors.

This chapter is based on the contributions presented in **[J8]**.

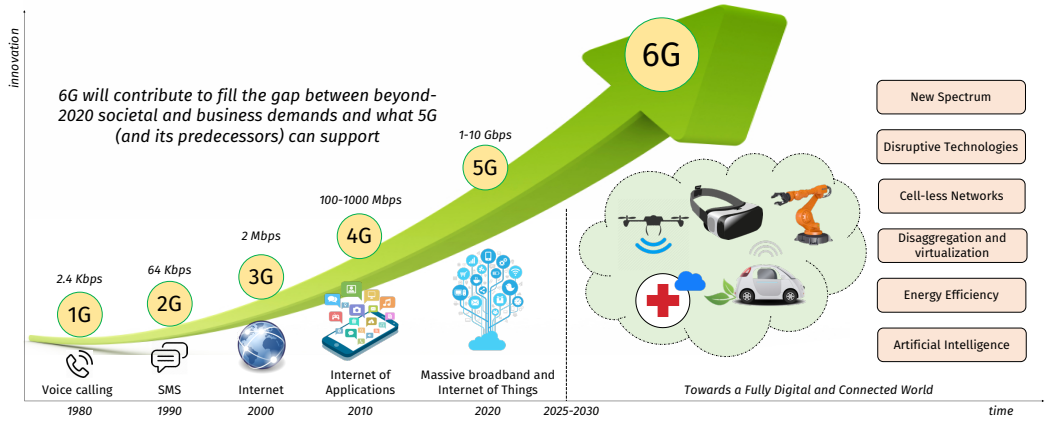


Fig. 8.1: Evolution of cellular networks, from 1G to the disruption foreseen for 6G networks, with a relevant/representative application for each generation.

Importantly, we identify several use cases that go beyond the performance of 5G systems under development today and demonstrate why it is important to think about the long term evolutions of 5G. Our analysis suggests that, in order to meet these demands, radically new underlying communication technologies, network architectures, and deployment models will be needed. In particular, in Sec. 8.3 we envision:

- *Novel disruptive communication technologies:* although 5G networks have already been designed to operate at very high frequencies, e.g., in the mmWave bands in NR, 6G networks could very much benefit from even higher spectrum technologies, e.g., through terahertz and optical communications.

- *Innovative network architectures:* despite 5G advancements towards more efficient network setups, the heterogeneity of the requirements of future network applications calls for new cell-less architectural paradigms based on tight integration among different communication technologies, for both access and backhaul, and the disaggregation and virtualization of the networking equipment.

- *Integrating Intelligence in the Network:* we expect 6G to bring intelligence from centralized computing facilities to end terminals, thereby providing concrete implementation to distributed learning models that have been studied from a theoretical point of view in a 5G context. Unsupervised learning and inter-user inter-operator knowledge sharing will promote real-time network decisions through prediction.

We transfer into our study a multifaceted critical spirit too, having selected, out of several possible breakthrough architectural innovations, the technologies that we believe have significant potential for future 6G systems, including developments at all layers of the protocol stack, from physical layer communication methods to networking design. We also study evolutions of network designs that have not yet been thoroughly addressed in early 5G standards development and will therefore not be fully transposed into commercial 5G deployments. We expect our investigation to promote research efforts towards the definition of new communication and networking technologies to meet the boldest requirements of 6G use cases.

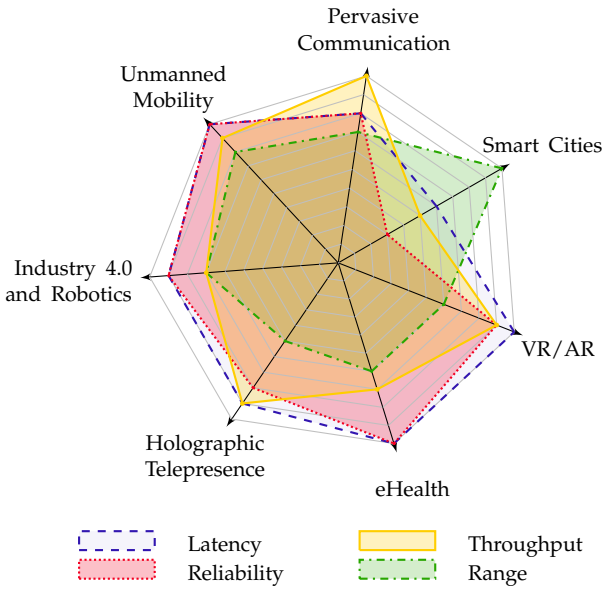


Fig. 8.2: Representation of the requirements (according to multiple KPIs) of different 6G use cases.

8.2 6G Potential Applications

5G technologies are associated with trade-offs on latency, power consumption, costs, hardware complexity, throughput, end-to-end reliability, and communication resilience. 6G innovations, on the contrary, will be developed in such a way that stringent network demands (in terms of ultra-high reliability, capacity, energy efficiency, and low latency) are jointly met in a holistic fashion.

In this section, we review the proprieties, characteristics and foreseen requirements of applications that, for their generality and complementarity, are generally believed to be good representatives of next-generation 6G services. Although some of these applications have already been discussed in 5G, we believe that they will likely not be part of future 5G deployments either due to technological limitations or because the market will not be mature enough to support them (especially within the very short timeframe in which 5G is supposed to be released). Fig. 8.2 provides a comprehensive point of view on application requirements in terms of different KPIs.

Augmented Reality (AR) and Virtual Reality (VR) AR and VR over wireless will enable novel applications including, but not limited to, (i) education and training, (ii) gaming, (iii) workspace communication, (iv) entertainment. VR/AR applications will face unprecedented challenges in terms of increased image quality and per-user capacity, sub-ms latency and uniform quality of experience (also at the cell edge). 6G will bring intelligence to end users, to support efficient data dissemination while fulfilling the network's heterogeneous requirements and backhaul/fronthaul limitations.

Holographic Telepresence (Teleportation) The human tendency to remotely connect with an increasing degree of digital accuracy will pose severe communications challenges in next generation network infrastructures. The authors of [327] explore a 3D holographic display and its data transmission requirement: a raw hologram, without any compression, with colors, full parallax, and 30 fps, would require a daunting 4.32 Tbps data rate. The latency requirement will hit the sub-ms,

and thousands of synchronized view angles will be necessary, as opposed to the few required for VR/AR. Moreover, to fully realize an immersive remote experience, all the 5 human senses are destined to be digitized and transferred across future networks, increasing the overall target data rate.

eHealth 6G will revolutionize the health-care sector, eliminating time and space barriers through remote surgery and guaranteeing health-care workflow optimizations. Besides the high cost, the current major limitation is the lack of real-time tactile feedback [328]. Moreover, QoS expectations for eHealth services (i.e., continuous connection availability, ultra-low latency, and mobility support) will unlikely be jointly fulfilled by 5G systems, due to the increased inherent variability of the mmWave channel and the congestion below 6 GHz. 6G will enable eHealth applications through innovations like mobile edge computing, virtualization and artificial intelligence.

Pervasive Connectivity Although 5G networks support more than 1'000'000 connections per km^2, mobile traffic will grow 3-fold from 2016 to 2021, thereby pushing the number of mobile devices to the extreme, with more than 125 billion connected devices worldwide by 2030 [329]. This will stress already congested networks, which will not manage to provide connectivity to every user and to meet the requirements shown in Fig. 8.2. Moreover, while 80% of the mobile traffic is generated indoor, cellular networks never really targeted indoor coverage. For example, 5G infrastructures, which may be operating in the mmWave spectrum, will hardly provide indoor connectivity as high-frequency radio signals cannot easily penetrate solid material. Furthermore, 5G densification presents scalability issues and high deployment and management costs for operators. 6G networks will instead provide seamless and pervasive connectivity in a variety of different contexts, matching stringent QoS requirements in both outdoor and indoor scenarios with a cost-aware and resilient infrastructure.

Industry 4.0 and Robotics 6G will foster the Industry 4.0 revolution started with 5G, i.e., the digital transformation of manufacturing through cyber physical systems and IoT services. The overcoming of the boundaries between the physical factory dimension and the cyber computational space will enable among other things, Internet-based diagnostics, maintenance, operation, and direct Machine to Machine (M2M) communications in a cost-effective, flexible and efficient way [330]. Automation comes with its own set of requirements in terms of reliable and isochronous communication [331], which 6G is positioned to address through the disruptive set of technologies we will describe in Sec. 8.3.

Smart Cities 6G will accelerate the adoption of solutions for smart cities, targeting life quality improvements, environmental monitoring, traffic control and city management automation [332]. These services build upon data generated by low-cost and low-energy sensors, which efficiently interact with each other and their surrounding environment. Current cellular systems have been mainly developed for broadband applications, with ad hoc configurations for M2M traffic. Conversely, 6G will seamlessly include support for user-centric machine to machine communication, providing native support for smart cities in a cost-effective way. 6G will also promote ultra-long battery lifetime combined with energy harvesting approaches, a research challenge that 5G and its predecessors have, so far, largely disregarded.

Unmanned Mobility The automotive industry is evolving towards fully autonomous transportation systems, offering safer traveling, improved traffic management, and support for infotainment, with market estimates in the order of 7 trillions USD [29]. The design and deployment of

connected and autonomous vehicles is still challenging: with the safety of passengers at stake, unprecedented levels of reliability and low end-to-end latency (i.e., above 99.9999% and below 1 ms, respectively) are expected, even in ultra-high mobility scenarios (up to an impressive 1000 km/h). Moreover, cars will be equipped with an increasing number of sensors (more than 200 per vehicle by 2020) which will demand increasing data rates (in the order of terabytes per driving hour [191]), saturating the capacity of traditional technologies. In addition, flying vehicles (e.g., drones) represent a huge market potential for various use cases such as construction, agriculture, and first responders. Swarms of drones will need improved capacity for expanding Internet connectivity. In this perspective, 6G will pave the way for the coming era of connected vehicles through advances in hardware and software as well as the pioneering connectivity solutions the we will discuss in Sec. 8.3.

8.3 6G Enabling Technologies

In this Section, we will focus on the technologies that were deliberately left out of early 5G standards development (i.e., 3GPP NR Release 15 and 16). These innovations will contribute to overcome the limitations of commercial 5G deployments, to satisfy the KPIs for the 6G applications that we described in Fig. 8.2. We will consider physical layer breakthroughs in Sec. 8.3.1, new architectural and protocol solutions in Sec. 8.3.2, and finally disruptive applications of artificial intelligence in Sec. 8.3.3. Table 8.1 summarizes the main technological innovations that could be introduced in 6G networks, considering their potential, the associated challenges and which of the use cases introduced in Sec. 8.2 they empower.

8.3.1 Disruptive Communication Technologies

A new generation of mobile networks is generally characterized by a set of novel communication technologies that provide unprecedented performance (e.g., in terms of available data rate and latency) and capabilities. For example, massive MIMO and mmWave communications are both key enablers of 5G networks. In order to meet the requirements that we described in Sec. 8.2, 6G networks are expected to rely on conventional spectrum (i.e., sub-6 GHz and mmWaves) but also on frequency bands that have not been considered yet for cellular standards, namely the terahertz band and Visible Light Communications (VLC). Fig. 8.3 represents the pathloss for each of these bands, in typical deployment scenarios, in order to highlight the differences and the opportunities that each portion of the spectrum can exploit. In the following paragraphs, we will focus on the two novel spectrum bands that will be used in 6G.

- **Terahertz communications** exploit the frequency bands between 100 GHz and 10 THz [333]. With respect to the millimeter waves used in 5G, terahertz brings to the extreme the potentials and challenges of high-frequency communications. The main issues that prevented the adoption of terahertz links in commercial systems are propagation loss, molecular absorption, high penetration loss, and challenges in the engineering of antennas and RF circuitry. As for mmWaves, the propagation loss can be compensated using directional antenna arrays, also enabling spatial multiplexing with limited interference. Furthermore, some frequencies in the terahertz spectrum are affected by an additional loss due to atmospheric molecular absorption, as shown in Fig. 8.3. However, it is possible to avoid this loss by choosing deployments in frequency bands not severely affected by molecular absorption, with contiguous chunks of up to 200 GHz of free spectrum [333]. Finally, the usage of such high frequencies enables new kinds of ultra-small-scale electronic packaging solutions for the RF and antenna circuitry.

Table 8.1: Comparison of 6G enabling technologies and relevant use cases.

Enabling Technology	Potential	Challenges	Use cases
New Spectrum			
Terahertz	High bandwidth, small antenna size, focused beams	Circuit design, high propagation loss	Pervasive connectivity, industry 4.0, holographic telepresence
VLC	Low-cost hardware, low interference, unlicensed spectrum	Limited coverage, need for RF uplink	Pervasive connectivity, smart cities
Novel PHY techniques			
Full duplex	Continuous TX/RX and relaying	Management of interference, scheduling	Pervasive connectivity, industry 4.0
Out-of-band channel estimation	Flexible multi-spectrum communications	Need for reliable frequency mapping	Pervasive connectivity, holographic telepresence
Sensing and localization	Novel services and context-based control	Efficient multiplexing of communication and localization	eHealth, unmanned mobility, industry 4.0
Innovative Network Architectures			
Cell-less architecture and multi-connectivity	Seamless mobility and integration of different kinds of links	Scheduling, need for new network design	Pervasive connectivity, unmanned mobility, holographic telepresence, eHealth
Disaggregation and virtualization	Lower costs for operators for massively-dense and edge deployments	High performance for PHY and MAC processing	Pervasive connectivity, holographic telepresence, eHealth, industry 4.0, smart cities, unmanned mobility
Advanced access-backhaul integration	Flexible deployment options, outdoor-to-indoor relaying	Scalability, scheduling and interference	Pervasive connectivity, smart cities, eHealth
Energy-harvesting and low-power operations	Energy-efficient network operations, resiliency	Need to integrate energy source characteristics in protocols	Smart cities, eHealth
Intelligence in the network			
Learning for value of information assessment	Intelligent and autonomous selection of the information to transmit	Complexity, unsupervised learning	Pervasive connectivity, eHealth, holographic telepresence, industry 4.0, unmanned mobility, smart cities
Knowledge sharing	Speed up learning in new scenarios	Need to design novel sharing mechanisms	Pervasive connectivity, smart cities, unmanned mobility
User-centric network architecture	Distributed intelligence to the endpoints of the network	Real-time and energy-efficient processing	Pervasive connectivity, eHealth, smart cities, industry 4.0
Not considered in 5G		With new features/capabilities in 6G	

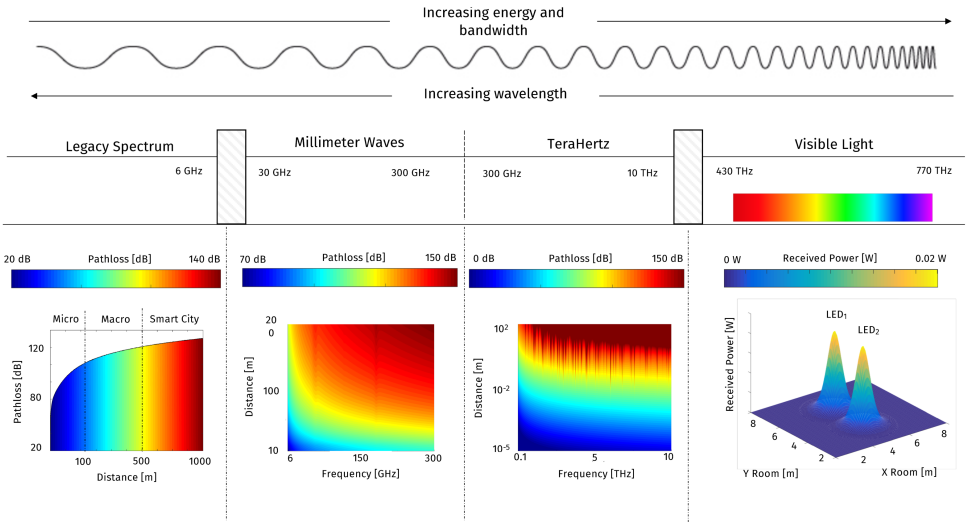


Fig. 8.3: Pathloss for sub-6 GHz, mmWave and terahertz bands, and received power for VLC. The sub-6 GHz and mmWave pathloss follows the 3GPP models and considers both LOS and NLOS conditions, while LOS-only is considered for terahertz [333] and VLC [335].

- **VLC** have been proposed to complement RF communications by piggybacking on the wide adoption of cheap Light Emitting Diode (LED) luminaries. These devices can indeed quickly switch between different light intensities to modulate a signal which can be transmitted to a proper receiver [334]. The research on VLC is more mature than that on terahertz communications, also thanks to a lower cost of experimental platforms. As reported in Fig. 8.3, VLC have limited coverage range, require an illumination source, and suffer from shot noise from other light sources (e.g., the sun), thus can be mostly used indoors [334]. Moreover, they need to be complemented by RF for the uplink. Nonetheless, VLC could be used to introduce cellular coverage in indoor scenarios, which, as mentioned in Sec. 8.2, is a use case that has not been properly addressed by cellular standards.

Although standardization bodies are promoting study items that are oriented towards the investigation of THz and VLC solutions for future wireless systems (i.e., IEEE 802..15.3d and 802.15.7, respectively), these technologies have not been so far considered by the 3GPP for inclusion in a cellular network standard, and will be targeting beyond 5G use cases.

Besides the new spectrum, 6G will also transform wireless networks by leveraging a set of technologies that have been enabled by recent physical layer and circuits research, but are not part of 5G. The following will be key enablers for 6G:

- **Full-duplex communication stack.** With full-duplex communications, the transceiver in base stations and users will be capable of receiving a signal while also transmitting, thanks to self-interference-suppression circuits [336]. Practical full-duplex deployments have been made feasible by breakthrough in the development of the aforementioned circuits only recently, thus have never been included into cellular network standards. These technology advancements can enable continuous downlink and uplink transmission, to increase the multiplexing capabilities and the overall system throughput without using additional bandwidth. Nonetheless, 6G networks will need careful planning for the full-duplex procedures and deployments, to avoid interference, as well as novel resource scheduler designs [336].
- **Novel channel estimation techniques (e.g., out-of-band estimation and compressed sens-**

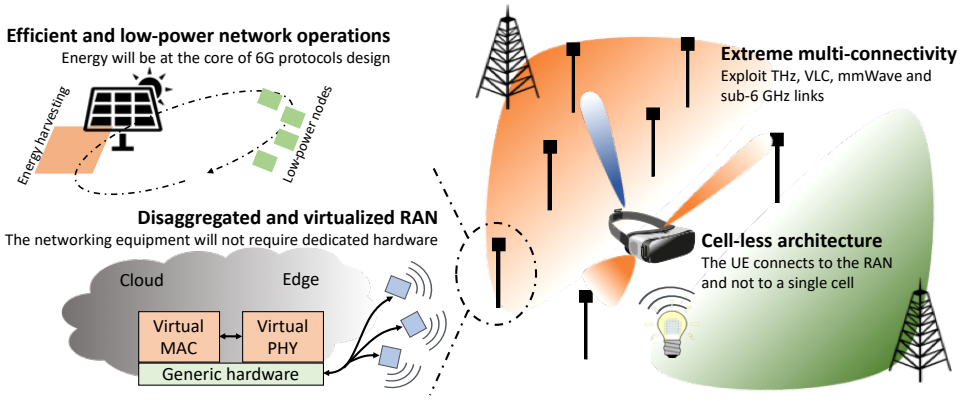


Fig. 8.4: Architectural innovations introduced in 6G networks.

ing). The channel estimation for directional communications will be a key component of communications at ultra-high frequencies in a cellular context, as for mmWaves. However, it is difficult to design efficient procedures for directional communications, considering multiple frequency bands and possibly a very large bandwidth. Therefore, 6G systems will need new channel estimation techniques. For example, out-of-band estimation (e.g., for the angular direction of arrival of the signal) can improve the reactiveness of beam management, by mapping the omnidirectional propagation of sub-6 GHz signals to the channel estimation for mmWave frequencies [337]. Similarly, given the sparsity in terms of angular directions of mmWave and terahertz channels, it is possible to exploit compressive sensing to estimate the channel using a reduced number of samples.

- **Sensing and network-based localization.** The usage of RF signals to enable simultaneous localization and mapping has been widely studied [338], but such capabilities have never been deeply integrated with the operations and protocols of cellular networks. 6G networks will exploit a unified interface for localization and communications to (i) improve control operations, which can rely on context information to shape beamforming patterns, reduce interference, and predict handovers; and (ii) offer innovative user services, e.g., for vehicular and eHealth applications.

8.3.2 Innovative Network Architectures

The disruption brought by the communication technologies described in Sec. 8.3.1 will enable a new 6G network architecture, but also potentially require structural updates with respect to current mobile network designs. For example, the density and the high access data rate of terahertz communications will create constraints on the underlying transport network, which has to provide both more points of access to fiber and a higher capacity than today's backhaul networks. Moreover, the wide range of different communication technologies available will increase the heterogeneity of the network, which will need to be managed.

The main architectural innovations that 6G will introduce are described in Fig. 8.4. In this context, we envision the introduction and/or deployment of the following architectural paradigms:

- **Cell-less architecture and tight integration of multiple frequencies and communication technologies.** 6G will break the current boundaries of cells, with users connected to the network as a whole and not to a single cell. This can be achieved, for example, through multi

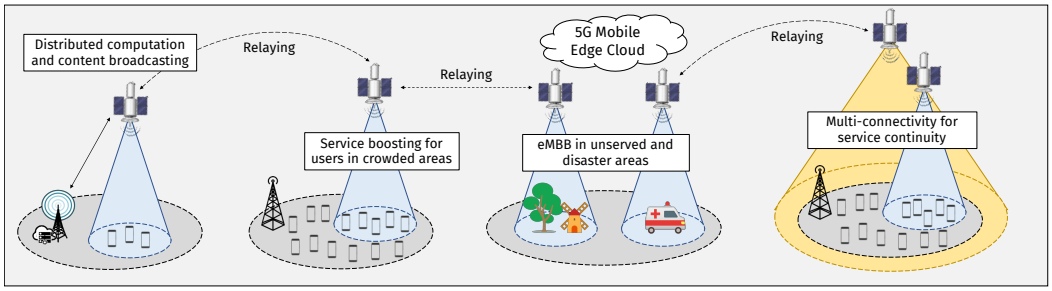


Fig. 8.5: Illustration of use cases supported by integration of terrestrial and non-terrestrial communications.

connectivity techniques, and the support for heterogeneous radios in the devices. The cellless network procedures will guarantee a seamless mobility support, without overhead due to handovers (which might be frequent when considering systems at terahertz frequencies), and will provide QoS guarantees even in challenging mobility scenarios such as vehicular ones. Overcoming the cell concept will also enable a tight integration of the different 6G communication technologies. The devices will be able to seamlessly transition among different heterogeneous links (e.g., sub-6 GHz, mmWave, terahertz or VLC) without manual intervention or configuration. Finally, according to the specific use case, the user may also concurrently use different network interfaces to exploit their complementary characteristics, e.g., the sub-6 GHz layer for control, and a terahertz link for the data plane.

- **3D network architecture.** Traditionally, networks have been designed to provide connectivity for an essentially bi-dimensional space, i.e., network access points are deployed to offer connectivity to devices on the ground. On the contrary, we envision future 6G heterogeneous architectures to provide three-dimensional (3D) coverage, thereby complementing terrestrial infrastructures with non-terrestrial platforms (e.g., drones, balloons, and satellites). In particular, satellite networking has rapidly gained attention to serve various applications, as illustrated in Fig. 8.5, including:
  - *Communication resilience*: Satellites enable on-demand wide geographical connectivity coverage and guarantee seamless service continuity, e.g., in rural areas or when terrestrial infrastructures are not available – for example after natural disasters – avoiding the operational and management costs of always-on fixed infrastructures.
  - *Resource optimization on parallel backhaul links*: Satellites offer an additional and robust channel for backhauling operations, thereby saving terrestrial resources for access traffic requests. This also guarantees that, in the event that terrestrial links are unavailable, the on-the-ground terminals can find an alternate route to preserve the connection.
  - *Efficient data broadcasting and relaying:* Satellites can convey multimedia and entertainment contents to an unlimited number of terminals using broadcast and multicast streams with information-centric networking.
  - *QoS enhancement through edge computing*: Satellites can host Mobile Edge Cloud (MEC) functionalities to support communication, computing, and storage operations, thereby providing on-the-ground users with an execution environment for their mobile edge appliances and evolving cloud services coverage towards 3D.

  In the effort to maximize communication data rates through satellite links, mmWave bands can be considered thanks to the larger bandwidth available compared to the conventional

sub-6 GHz bands. The potential of mmWave technology as a means to support satellite communications has also been recognized by the 3GPP which, in [339], defines non-terrestrial network deployment scenarios and related system parameters including channel modeling at NR frequencies. Despite such promising opportunities, there are various challenges to be solved before flying platforms can effectively be used in wireless networks, e.g., air-to-ground channel modeling, topology and trajectory optimization, resource management and energy efficiency, coverage constraints (the higher the satellite altitude, the larger the coverage, the longer the visibility period, but the higher the latency). These issues will not be fully solved within the relatively limited timeframe in which 5G networks will be commercialized, and will therefore be addressed as part of future 6G research efforts.

- **Disaggregation and virtualization of the networking equipment.** Networks have recently started to transition towards the disaggregation of once-monolithic networking equipments: for example, 5G base stations can be deployed via distributed units with the lower layer of the protocol stack, and centralized units in data centers at the edge. The advances in computing capabilities will bring virtualization and disaggregation to the extreme, with the 6G MAC and PHY layers fully virtualized and simple and low-cost distributed units with just the antennas and minimal processing units. This will decrease the costs of networking equipment, making a massively dense deployment economically feasible. Moreover, in its current state, research on virtualization does not close all the security gaps, which could result in numerous cyber-attacks. Network security concerns will be addressed in the upcoming 6G wireless network standards.
- **Advanced access-backhaul integration.** The massive data rates provided by the new 6G access technologies will require an adequate growth of the backhaul capacity. Moreover, terahertz and VLC deployments will call for a massive increase in the density of access points, which should be provided with backhaul connectivity to their neighbors and the core network. However, the huge capacity of 6G technologies can be exploited for self-backhauling solutions, in which the radios in the base stations provide both access and backhaul services. While a similar option is already being considered for 5G, the scale of 6G deployments will introduce new challenges and opportunities: the networks will need higher autonomous configuration capabilities, but the increase in access capacity will not need to be matched by an increase in fiber points of presence.
- **Energy-harvesting strategies for low-power consumption network operations.** 6G devices will be deployed in a pervasive manner to satisfy the future connectivity requirements. User terminals and networking equipment will need to be powered with energy sources and, given the scale expected in 6G networks, it is necessary to design the system to be more efficient and less energy consuming with respect to current networks. This means that both the circuitry and the communication stack will be developed with energy-awareness in mind. One option is using energy harvesting circuits to allow devices to be self-powered, which could be critical for example to enable off-grid operations, long-lasting IoT devices and sensors, or long stand-by intervals for devices and equipment which are rarely used.

8.3.3 Integrating Intelligence in the Network

The complexity of 6G communication technologies and network deployments will probably prevent closed-form and/or manual optimizations. While intelligent techniques in cellular networks are already being discussed for 5G, we expect 6G deployments to be much denser (i.e., in terms of number of access points and users), heterogeneous (in terms of integration of different technologies), and

with stricter requirements in terms of performance with respect to 5G. Therefore, intelligence will play a more prominent role in the network, going beyond classification and prediction tasks which are being considered for 5G systems. Notice that the standard may not specify the techniques and learning strategies to be deployed in networks, but data-driven approaches can be seen as tools that network vendors and operators can use to meet the 6G requirements [340]. In particular, 6G research will be oriented towards the following aspect:

- **Learning techniques for data selection and feature extraction.** The large volume of data generated by future connected devices (e.g., sensors in autonomous vehicles) will likely put a strain on communication technologies, which may be unable to guarantee the required quality of service. It is therefore fundamental to discriminate the *value of information* to use the (limited) network resources in a way that maximizes the utility for the end users. In this context, machine learning strategies can be developed to evaluate the degree of correlation in observations, or to extract features from input vectors and predict the a posteriori probability of a sequence given its entire history. Given that the amount of data generated will be massive, labeling the data via supervised learning may be infeasible. Unsupervised learning, coupled with reinforcement learning approaches, on the other hand, does not need labeling, and can be used to operate the network in a truly autonomous fashion.
- **Inter-user inter-operator knowledge sharing.** Spectrum and infrastructure sharing has already been proven to be beneficial in cellular networks, to maximize the multiplexing capabilities. With learning-driven networks, operators and users can also share learned/processed representations of specific network deployments and/or use cases, for example to speed up the network configuration in new markets, or to better adapt to new unexpected scenarios which may emerge during the operations of the network. The trade-offs in latency, power consumption, system overhead, and cost will be studied in 6G, for both on-board and edge-cloud-assisted solutions.
- **User-centric network architecture.** The application of machine learning in networks is still in its infancy, but is promising in the context of complex 6G systems that envision artificial intelligence solutions to be implemented in a distributed way, in view of a fully-user-centric network architecture. In such a way, end terminals will be able to make autonomous network decisions based on the outcomes of previous operations, so as not to incur into communication overhead to and from centralized controllers. On-board distributed methods can process machine learning algorithms in real time, thereby yielding more responsive network management and overcoming, for the first time, the physical limits of signal propagation that currently prevents control operations to be performed within sub-ms latency.

Chapter 9

Conclusions

5G is the new generation of wireless telecommunication networks. It is based on a cloud-native, softwarized, end-to-end architecture, and is characterized by ultra-high data rates, ultra-low latency and support for a massive number of connections. The mmWave spectrum represents a key technology for fulfilling the foreseen extreme traffic demands of future 5G applications, although the harsh propagation characteristics of this frequency band pose significant challenges for the whole protocol stack. Along these lines, in this thesis we shed light on the potentials and limitations of mmWaves to support next-generation cellular and vehicular network operations. In particular, we focused on how to design and dimension mobility management strategies in mmWaves cellular systems, on how to accurately deploy mmWave networking architectures, on which radio access technologies provide the best performance in complex vehicular scenarios, and on the most promising options to broadcast vehicular sensory observations in an efficient way.

After an overview of the main 5G requirements and enabling technologies, in the first part of the thesis we focused on the design of cellular control operations for the mobility management of network devices in both idle and connected modes. In particular, we discussed the implications that the directional nature of the mmWave technology has on users accessing the network for the first time, i.e., during initial access, or handing over to a new serving base station, when the quality of the signal becomes unacceptably low. Based on an extensive analysis and simulation campaign, we demonstrated the fundamental role of multi-connectivity to improve the performance of a network in which mmWave access links are established with respect to several metrics, including latency, throughput, radio control signaling overhead, detection accuracy and reactiveness.

We also reviewed the benefits and challenges that an IAB architecture introduces in mmWave deployments. Our investigation was based on the results obtained using a new ns-3 module that accurately models the IAB functionalities proposed in the 3GPP NR standard on both the data and the control planes. We believe that this tool can be used by researchers to understand the main limitations and the performance gains that IAB networks can provide, and to evaluate new integrated scheduling algorithms and multi-hop routing strategies with a realistic, end-to-end protocol stack. We showed that IAB appears as a viable paradigm to efficiently relay cell-edge traffic, even though the benefits reduce in case of congested networks. We also highlighted the limitations of the IAB solution and provided guidelines on how to overcome them.

In the second part of the thesis we validated the feasibility of designing mmWave communications protocols in the context of future vehicular networks. In particular, we made the case that the potential of connected and autonomous vehicles can be greatly magnified by the synergistic exploita-

tion of multiple RATs, including indeed mmWave frequencies (thanks to the availability of massive bandwidth to support multi-Gbps transmission speeds) and sub-6 GHz technologies (thanks to the intrinsic robustness of the legacy channel). We also investigated the impact of the base station and vehicular density, the target application, the propagation scenario, and the vehicular mobility on the overall end-to-end system performance.

We then analyzed the possibility of improving the accuracy of position estimation and prediction in future vehicular networks by implementing a threshold-based broadcasting strategy that considers VoI as a proxy for distributing sensory observations. We showed that VoI can be characterized as a function of the environment in which the vehicles are deployed, and evolves as a function of spatial, temporal, and quality criteria. Additionally, we designed an innovative congestion control strategy that forces vehicles to broadcast state information only when the estimated positioning error is above a critical threshold. We demonstrated that the proposed strategy can reduce the packet collision probability with minimum overhead compared to a baseline solution which relies on periodic dissemination of data.

In the third and last part of the thesis, we envisioned how 5G can evolve into 6G to support use cases and technologies that we believe will characterize the 2030 society. The research in the wireless domain has the potential to disrupt in 6G the traditional cellular networking paradigms that still exist in 5G, introducing for example the support for terahertz and visible light spectra, cell-less and aerial architectures, and distributed intelligence, among others. These technologies, however, are not market-ready: this represents a unique opportunity for the wireless research community to foster innovations that will enable unforeseen use cases towards a fully digital and connected world.

Acknowledgments

The work and results presented in this thesis has been supported in part by the network division group of Toyota Motor North America, Inc., the wireless research group of the New York University (NYU), and the R&D group of InterDigital Communications, Inc.

Bibliography

[1] GSMA Intelligence, "Understanding 5G: Perspectives on future technological advancements in mobile," *White Paper*, 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[3] Cisco, "Cisco Visual Networking Index: Forecast and Trends, 2017–2022," *White Paper*, 2019.

[4] P. Marsch, Ö. Bulakci, O. Queseth, and M. Boldi, *5G system design: architectural and functional considerations and long term research*. John Wiley & Sons, 2018.

[5] ITU-R WP5D, "Draft New Recommendation, "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," Doc. R12-SG05-C-0199, 2015.

[6] NGMN Alliance, "NGMN 5G White Paper," *White Paper*, 2015.

[7] Global mobile Suppliers Association (GSA), "Global Progress to 5G – Trials, Deployments & Launches – July 2019," *Report*, 2019. [Online]. Available: https://gsacom.com/paper/global-progress-to-5g-trials-deployments-launches-july-2019/

[8] 3GPP, "Study on scenarios and requirements for next generation access technologies (Release 14)," *TR 38.913*, 2018.

[9] 3GPP, "NR and NG-RAN Overall Description - Rel. 15," TS 38.300, 2018.

[10] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, March 2014.

[11] 3GPP, "NR - Physical channels and modulation - Release 15," TS 38.211, V15.0.0, 2018.

[12] M. A. Marsan, N. B. Melazzi, and S. Buzzi, *5G Italy White eBook: from Research to Market*, 2018, 5G Italy White Book.

[13] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "Standalone and Non-Standalone Beam Management for 3GPP NR at mmWaves," *IEEE Communications Magazine*, vol. 57, no. 4, pp. 123–129, April 2019.

[14] 3GPP, "System Architecture for the 5G System; Stage 2 - Release 15," TS 23.501, V15.3.0, 2018.

[15] A. Maeder, A. Ali, A. Bedekar, A. F. Cattoni, D. Chandramouli, S. Chandrashekar, L. Du, M. Hesse, C. Sartori, and S. Turtinen, "A scalable and flexible radio access network architecture for fifth generation mobile networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 16–23, November 2016.

[16] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network Slicing in 5G: Survey and Challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, May 2017.

[17] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research: Policy and Practice*, vol. 77, pp. 167–181, Jul 2015.

[18] G. Silberg, R. Wallace, G. Matuszak, J. Plessers, C. Brower, and D. Subramanian, "Self-driving cars: The next revolution," *White Paper, KPMG LLP & Center of Automotive Research*, p. 36, 2012.

[19] Ohio State University, "The Future of Driving," *White Paper*, 2015.

[20] P. Diamandis, "Self-driving cars are coming," *Forbes Magazine*, 2015.

[21] L. M. Clements and K. M. Kockelman, "Economic effects of automated vehicles," *Transportation Research Record*, vol. 2606, pp. 106–114, 2017.

[22] 3GPP, "Service requirements for enhanced V2X scenarios (Release 15)," *TS 22.186*, Sept 2018.

[23] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger, and W. Xu, "Connected Roads of the Future: Use Cases, Requirements, and Design Considerations for Vehicle-to-Everything Communications," *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 110–123, Sept 2018.

[24] S. Kim, B. Qin, Z. J. Chong, X. Shen, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus, "Multivehicle Cooperative Driving Using Cooperative Perception: Design and Experimental Validation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 663–680, April 2015.

[25] Y. J. Li, "An overview of the DSRC/WAVE technology," in *International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*. Springer, 2010, pp. 544–558.

[26] 3GPP, "Study on LTE-based V2X services (Release 14)," TR 36.885, 2016.

[27] T. Zugno, M. Drago, M. Giordani, M. Polese, and M. Zorzi, "Towards Standardization of Millimeter Wave Vehicle-to-Vehicle Networks: Open Challenges and Performance Evaluation," *submitted to IEEE Vehicular Technology Magazine*, 2019.

[28] G. Naik, B. Choudhury, and J. Park, "IEEE 802.11bd 5G NR V2X: Evolution of Radio Access Technologies for V2X Communications," *IEEE Access*, vol. 7, pp. 70 169–70 184, 2019.

[29] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289–299, May 2014.

[30] 3GPP, "Technical Specification Group Services and System Aspects; Study on enhancement of 3GPP Support for 5G V2X Services (Release 15)," *TS 22.886*, 2017.

[31] 3GPP, "Study on evaluation methodology of new Vehicle-to-Everything V2X use cases for LTE and NR (Release 15)," TR 37.885, 2019.

[32] ——, "Study on NR Vehicle-to-Everything (V2X) (Release 16)," TR 38.885, 2019.

[33] 3GPP, "Study on NR beyond 52.6 GHz (Release 16)," *RP-181435*, 2018.

[34] ——, "Draft Report of 3GPP TSG RAN WG1 95 v0.3.0," 2019.

[35] ——, "Feature Lead Summary for NR-V2X Resource Alloc. Mechanism," *R1-1903397*, 2019.

[36] ——, "Discussion on mode 2 resource allocation in NR-V2X)," *R1-1810977*, 2018.

[37] ——, "Summary for NR-V2X Resource Allocation Mechanism)," *R1-1814260*, 2018.

[38] ——, "Evaluation of NR V2X Mode 2 Resource Allocation)," *R1-1901844*, 2019.

[39] R. Cao, H. Zhang, and P. Sharma, "Potential PHY Designs for NGV," *IEEE 802.11-19/0016r0*, 2019.

[40] W. Anwar, N. Franchi, G. Fettweis, and A. Trassl, "On the Reliability of NR-V2X and IEEE 802.11bd," 09 2019.

[41] M. Fisher, A. Filippi, and V. Martinez, "Interoperable NGV PHY Improvements," *IEEE 802.11-18/1186r0*, 2018.

[42] R. Cao, "Doppler Impact on OFDM Numerology for NGV," *IEEE 802.11-18/1553r0*, 2018.

[43] L. Dongguk, "Consideration on Features for 11bd," *IEEE 802.11-19/0009r0*, 2019.

[44] M. Hiroyuki, "mmW for V2X use cases," *IEEE 802.11-18/1187r2*, 2018.

[45] L. Dongguk, "NGV PHY Feasibility Discussions," *IEEE 802.11-18/0860r3*, 2018.

[46] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, May 2013.

[47] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Pearson Education, 2014.

[48] V. Va, J. Choi, and R. W. Heath, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5014–5029, 2016.

[49] M. Rebato, L. Resteghini, C. Mazzucco, and M. Zorzi, "Study of realistic antenna patterns in 5G mmwave cellular scenarios," in *IEEE International Conference on Communications (ICC)*, 2018.

[50] J. S. Lu, D. Steinbach, P. Cabrol, and P. Pietraski, "Modeling human blockers in millimeter wave radio links," *ZTE Communications*, vol. 10, no. 4, pp. 23–28, 2012.

[51] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.

[52] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial Access in 5G mmWave Cellular Networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 40–47, November 2016.

[53] 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, Firstquarter 2019.

[54] M. Giordani, M. Rebato, A. Zanella, and M. Zorzi, "Coverage and connectivity analysis of millimeter wave vehicular networks," *Ad Hoc Networks*, vol. 80, pp. 158–171, Aug 2018.

[55] M. Rebato, M. Mezzavilla, S. Rangan, F. Boccardi, and M. Zorzi, "Understanding Noise and Interference Regimes in 5G Millimeter-Wave Cellular Networks," in *22th European Wireless Conference*, 2016.

[56] M. Zhang, M. Mezzavilla, R. Ford, S. Rangan, S. Panwar, E. Mellios, D. Kong, A. Nix, and M. Zorzi, "Transport layer performance in 5G mmWave cellular," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2016.

[57] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in Cellular Systems: Understanding Ultra-Dense Small Cell Deployments," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2078–2101, Fourthquarter 2015.

[58] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, Apr 2016.

[59] M. Polese, M. Giordani, and M. Zorzi, "3GPP NR: the cellular standard for 5G networks," *5G-ITALY White Book*, Dec 2018.

[60] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[61] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 136–143, December 2014.

[62] Federal Communications Commission (FCC), "Notice of proposed rulemaking: FCC 15-138," *IEEE Wireless Communications*, Oct 2015. [Online]. Available: /https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-138A1.pdf

[63] European Commission (EU), "Commission Implementing Decision (EU) 2019/784," *IEEE Wireless Communications*, May 2019. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019D0784

[64] 3GPP, "Study on New Radio (NR) Access Technology - Physical Layer Aspects - Release 14," TR 38.802, 2017.

[65] V. Raghavan, L. Akhoondzadeh-Asl, V. Podshivalov, J. Hulten, M. A. Tassoudji, O. H. Koymen, A. Sampath, and J. Li, "Statistical Blockage Modeling and Robustness of Beamforming in Millimeter-Wave Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 7, pp. 3010–3024, July 2019.

[66] E. Ben-Dor, T. S. Rappaport, Y. Qiao, and S. J. Lauffenburger, "Millimeter-wave 60 GHz outdoor and vehicle AOA propagation measurements using a broadband channel sounder," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2011.

[67] M. Giordani, M. Mezzavilla, A. Dhananjay, S. Rangan, and M. Zorzi, "Channel dynamics and SNR tracking in millimeter wave cellular systems," in *European Wireless (EW)*, 2016.

[68] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[69] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both?" *IEEE Communications Magazine*, vol. 52, no. 12, pp. 110–121, December 2014.

[70] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: the next wireless revolution?" *IEEE Communications Magazine*, vol. 52, no. 9, pp. 56–62, September 2014.

[71] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[72] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[73] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *IEEE Globecom Workshops*. IEEE, 2015.

[74] M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Multi-connectivity in 5G mmWave cellular networks," in *IEEE Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, 2016.

[75] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-Wave Enhanced Local Area Systems: A High-Data-Rate Approach for Future Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, June 2014.

[76] Z. He, S. Mao, and T. S. Rappaport, "Minimum time length link scheduling under blockage and interference in 60 GHz networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 837–842.

[77] J. G. Rois, B. Lorenzo, F. J. Gonzalez-Castano, and J. C. Burguillo, "Heterogeneous millimeter-wave/microwave architecture for 5G wireless access and backhauling," in *European Conference on Networks and Communications (EuCNC)*, 2016.

[78] S. Chandrashekar, A. Maeder, C. Sartori, T. Höhne, B. Vejlgaard, and D. Chandramouli, "5G multi-RAT multi-connectivity architecture," in *IEEE International Conference on Communications Workshops (ICC)*, May 2016, pp. 180–186.

[79] 3GPP, "Study on Small Cell enhancements for E-UTRA and E-UTRAN - Release 12," TR 36.842, 2013.

[80] ——, "NR - Multi-connectivity - Overall description (Stage 2)," TS 37.340, 2018.

[81] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[82] H. Sun, W. Fang, J. Liu, and Y. Meng, "Performance evaluation of CS/CB for coordinated multipoint transmission in LTE-A downlink," in *IEEE 23rd Int. Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2012.

[83] S. Sesia, M. Baker, and I. Toufik, *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.

[84] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2," TS 36.300, 2018.

[85] C. Barati, S. Hosseini, S. Rangan, P. Liu, T. Korakis, S. Panwar, and T. Rappaport, "Directional Cell Discovery in Millimeter Wave Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6664–6678, Dec 2015.

[86] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter Wave Cellular Networks: A MAC Layer Perspective," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3437–3458, Oct 2015.

[87] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, Apr 2013.

[88] M. Giordani, M. Mezzavilla, C. N. Barati, S. Rangan, and M. Zorzi, "Comparative analysis of initial access techniques in 5G mmWave cellular networks," in *Information Science and Systems (CISS), 2016 Annual Conference on*. IEEE, 2016, pp. 268–273.

[89] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "Initial access frameworks for 3GPP NR at mmWave frequencies," in *17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, June 2018.

[90] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: issues and approaches," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 180–185, 2015.

[91] C. N. Barati, S. A. Hosseini, M. Mezzavilla, P. Amiri-Eliasi, S. Rangan, T. Korakis, S. S. Panwar, and M. Zorzi, "Directional initial access for millimeter wave cellular systems," in *49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 307–311.

[92] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmWave communications," in *48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 1926–1930.

[93] L. Wei, Q. Li, and G. Wu, "Exhaustive, Iterative and Hybrid Initial Access Techniques in mmWave Communications," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017.

[94] J. Choi, "Beam selection in mm-Wave multiuser MIMO systems using compressive sensing," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2936–2947, August 2015.

[95] A. Capone, I. Filippini, and V. Sciancalepore, "Context information for fast cell discovery in mm-wave 5G networks," in *21th European Wireless Conference; Proceedings of European Wireless*, 2015.

[96] A. Capone, I. Filippini, V. Sciancalepore, and D. Tremolada, "Obstacle avoidance cell discovery using mm-waves directive antennas in 5G networks," in *IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2015, pp. 2349–2353.

[97] Q. C. Li, H. Niu, G. Wu, and R. Q. Hu, "Anchor-booster based heterogeneous networks with mmWave capable booster cells," in *IEEE Globecom Workshops*. IEEE, 2013, pp. 93–98.

[98] W. B. Abbas and M. Zorzi, "Context information based initial cell search for millimeter wave 5G cellular networks," in *European Conference on Networks and Communications (EuCNC)*. IEEE, 2016, pp. 111–116.

[99] A. Alkhateeb, Y. H. Nam, M. S. Rahman, J. Zhang, and R. W. Heath, "Initial Beam Association in Millimeter Wave Cellular Systems: Analysis and Design Insights," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2807–2821, May 2017.

[100] Y. Li, J. Luo, M. Castaneda, R. Stirling-Gallacher, W. Xu, and G. Caire, "On the Beamformed Broadcast Signaling for Millimeter Wave Cell Discovery: Performance Analysis and Design Insight," *arXiv preprint arXiv:1709.08483*, 2017.

[101] L. Wei, Q. C. Li, and G. Wu, "Initial Access Techniques for 5G NR: Omni/Beam SYNC and RACH designs," in *International Conference on Computing, Networking and Communications (ICNC)*, March 2018, pp. 249–253.

[102] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1018–1044, Second Quarter 2016.

[103] T. Nitsche, C. Cordeiro, A. Flores, E. Knightly, E. Perahia, and J. Widmer, "IEEE 802.11ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi [Invited Paper]," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, December 2014.

[104] J. Wang, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, October 2009.

[105] R. Santosa, B.-S. Lee, C. K. Yeo, and T. M. Lim, "Distributed Neighbor Discovery in Ad Hoc Networks Using Directional Antennas," in *The Sixth IEEE International Conference on Computer and Information Technology*, Sept 2006, pp. 97–97.

[106] K. Chandra, R. V. Prasad, I. G. Niemegeers, and A. R. Biswas, "Adaptive beamwidth selection for contention based access periods in millimeter wave WLANs," in *IEEE 11th Consumer Communications and Networking Conference (CCNC)*. IEEE, 2014, pp. 458–464.

[107] J. Liu, K. Au, A. Maaref, J. Luo, H. Baligh, H. Tong, A. Chassaigne, and J. Lorca, "Initial Access, Mobility, and User-Centric Multi-Beam Operation in 5G New Radio," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 35–41, March 2018.

[108] E. Onggosanusi, M. S. Rahman, L. Guo, Y. Kwak, H. Noh, Y. Kim, S. Faxer, M. Harrison, M. Frenne, S. Grant, R. Chen, R. Tamrakar, and a. Q. Gao, "Modular and High-Resolution Channel State Information and Beam Management for 5G New Radio," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 48–55, March 2018.

[109] 3GPP, "NR - Radio Resource Control (RRC) protocol specification - Release 15," TS 38.331, 2018.

[110] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2069–2084, Sept 2017.

[111] 3GPP, "NR - Physical layer measurements - Rel. 15," TS 38.215, 2018.

[112] ——, "NR PRACH preamble resource allocation," Ericsson - Tdoc R1-1611905, 2016.

[113] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[114] Ericsson, "5G New Radio: designing for the future," Ericsson Technology Review, 2017.

[115] 3GPP, "SS burst periodicity for initial cell selection in NR," Nokia, Alcatel-Lucent Shanghai Bell - Tdoc R4-1705123, 2017.

[116] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, February 2015.

[117] S. Dutta, C. N. Barati, A. Dhananjay, and S. Rangan, "5G Millimeter Wave Cellular System Capacity with Fully Digital Beamforming," in *51st Asilomar Conference on Signals, Systems and Computers*, Nov 2017.

[118] 3GPP, "Discussion on remaining issues of SS block and SS burst set," Motorola Mobility, Lenovo - Tdoc R1-1714212, 2017.

[119] ——, "Discussion on NR 4-Step Random Access Procedure," Ericsson - Tdoc R1-1718052, 2017.

[120] K. Takeda, L. H. Wang, and S. Nagata, "Latency Reduction toward 5G," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 2–4, June 2017.

[121] 3GPP, "LS on NR PRACH BW Aspects," Tdoc R1-1716814, 2017.

[122] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter Wave Receiver Efficiency: A Comprehensive Comparison of Beamforming Schemes With Low Resolution ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8131–8146, Dec 2017.

[123] M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "An Efficient Uplink Multi-Connectivity Scheme for 5G Millimeter-Wave Control Plane Applications," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6806–6821, Oct 2018.

[124] M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Mobility management for TCP in mmWave networks," in *Proceedings of the 1st ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, 2017.

[125] S. Schwarz, C. Mehlführer, and M. Rupp, "Calculation of the spatial preprocessing and link adaption feedback for 3GPP UMTS/LTE," in *6th conference on Wireless advanced (WiAD)*. IEEE, 2010.

[126] A. S. Cacciapuoti, "Mobility-Aware User Association for 5G mmWave Networks," *IEEE Access*, vol. 5, pp. 21 497–21 507, 2017.

[127] J. Palacios, D. De Donno, and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," in *IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2017.

[128] S. Jayaprakasam, X. Ma, J. W. Choi, and S. Kim, "Robust Beam-Tracking for mmWave Mobile Communications," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2654–2657, Dec 2017.

[129] M. Polese, M. Mezzavilla, and M. Zorzi, "Performance Comparison of Dual Connectivity and Hard Handover for LTE-5G Tight Integration," in *Proceedings of the 9th International Conference on Simulation Tools and Techniques*, 2016.

[130] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Caching Meets Millimeter Wave Communications for Enhanced Mobility Management in 5G Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 779–793, Feb 2018.

[131] N. Gonzalez-Prelcic, A. Ali, V. Va, and R. W. Heath, "Millimeter-Wave Communication with Out-of-Band Information," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 140–146, Dec. 2017.

[132] X. Yan, Y. A. Sekercioglu, and S. Narayanan, "A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks," *Computer Networks*, vol. 54, no. 11, pp. 1848 – 1863, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128610000502

[133] F. Guidolin, I. Pappalardo, A. Zanella, and M. Zorzi, "Context-aware handover policies in HetNets," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1895–1906, March 2016.

[134] A. Talukdar, M. Cudak, and A. Ghosh, "Handoff rates for millimeterwave 5G systems," in *IEEE 79th Vehicular Technology Conference (VTC Spring)*, May 2014.

[135] H. Song, X. Fang, and L. Yan, "Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4633–4646, Nov 2014.

[136] V. Yazici, U. C. Kozat, and M. O. Sunay, "A new control plane for 5G network architecture with a case study on unified handoff, mobility, and routing management," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 76–85, Nov 2014.

[137] S. Sadr and R. S. Adve, "Handoff rate and coverage analysis in multi-tier heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2626–2638, May 2015.

[138] S. Zang, W. Bao, P. L. Yeoh, B. Vucetic, and Y. Li, "Managing vertical handovers in millimeter wave heterogeneous networks," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1629–1644, Feb 2019.

[139] M. Joud, M. Garcia-Lozano, and S. Ruiz, "On the mobility of moderate speed users in ultra dense small cell deployments with mmw," in *IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5.

[140] P. Coucheney, E. Hyon, and J. M. Kelif, "Mobile association problem in heterogenous wireless networks with mobility," in *IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 3129–3133.

[141] 3GPP, "Measurement configuration for CSI-RS," Ericsson - Tdoc R2-1704103, 2017.

[142] ——, "NR - Physical layer procedures for data - Release 15," TS 38.214, 2018.

[143] ——, "NR CSI-RS configuration for RRM measurement," Samsung - Tdoc R2-1709593, 2017.

[144] ——, "Discussion on CSI-RS Design," Qualcomm - Tdoc R1-1718546, 2017.

[145] ——, "Measurement reporting for NR SS and CSI-RS ," Huawei - Tdoc R2-1708703, 2017.

[146] ——, "Consideration on CSI RS for beam management," ZTE Corporation - Tdoc R2-1708123, 2017.

[147] A. Elnashar and M. A. El-Saidny, "Looking at LTE in Practice: A Performance Analysis of the LTE System Based on Field Test Results," *IEEE Vehicular Technology Magazine*, vol. 8, no. 3, pp. 81–92, Sep. 2013.

[148] 3GPP, "RACH-less HO in NR when UE is in CA or DC," Ericsson - Tdoc R2-1706626, 2017.

[149] R. Ford, M. Zhang, S. Dutta, M. Mezzavilla, S. Rangan, and M. Zorzi, "A Framework for End-to-End Evaluation of 5G mmWave Cellular Networks in Ns-3," in *Proceedings of the Workshop on Ns-3*, June 2016, pp. 85–92.

[150] M. Mezzavilla, S. Dutta, M. Zhang, M. R. Akdeniz, and S. Rangan, "5G MmWave Module for the Ns-3 Network Simulator," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Nov 2015, pp. 283–290.

[151] T. R. Henderson, M. Lacage, G. F. Riley, C. Dowell, and J. Kopena, "Network simulations with the ns-3 simulator," *SIGCOMM demonstration*, vol. 14, no. 14, p. 527, 2008.

[152] J. Oueis and E. C. Strinati, "Uplink traffic in future mobile networks: Pulling the alarm," in *Int. Conf. on Cognitive Radio Oriented Wireless Networks*. Springer, 2016, pp. 583–593.

[153] 3GPP, "Study on new radio access technology: Radio access architecture and interfaces," TR 38.801, 2017.

[154] ——, "Remaining issues on SRS," InterDigital, Inc. - Tdoc R1-1716472, 2017.

[155] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero, "An open source product-oriented LTE network simulator based on ns-3," in *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Nov. 2011, pp. 293–298.

[156] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-End Simulation of 5G mmWave Networks," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2237–2263, thirdquarter 2018.

[157] Next Generation Mobile Networks Alliance, "Optimised backhaul requirements," Tech. Rep., 2008.

[158] 3GPP, "Relation between radio link failure and beam failure," Ericsson - Tdoc R1-1705917, 2017.

[159] N. Moraitis and P. Constantinou, "Indoor channel measurements and characterization at 60 GHz for wireless local area network applications," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 12, pp. 3180–3189, Dec 2004.

[160] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer, "Steering with eyes closed: mm-wave beam steering without in-band measurement," in *IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 2416–2424.

[161] A. Patra, L. Simic, and P. Mahonen, "Smart mm-wave beam steering algorithm for fast link re-establishment under node mobility in 60 GHz indoor WLANs." in *ACM International Symposium on Mobility Management and Wireless Access (MobiWac)*, 2015.

[162] K. Takeda, L. H. Wang, and S. Nagata, "Latency Reduction toward 5G," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 2–4, June 2017.

[163] 3GPP, "Beam failure detection and beam recovery actions," Ericsson - Tdoc R1-1705893, 2017.

[164] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in Cellular Systems: Understanding Ultra-Dense Small Cell Deployments," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 2078–2101, Fourth quarter 2015.

[165] N. Makris, C. Zarafetas, P. Basaras, T. Korakis, N. Nikaein, and L. Tassiulas, "Cloud-based Convergence of Heterogeneous RANs in 5G Disaggregated Architectures," in *IEEE International Conference on Communications (ICC)*. IEEE, 2018.

[166] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 236–262, Sept 2016.

[167] H. S. Dhillon and G. Caire, "Wireless backhaul networks: Capacity bound, scalability analysis and design guidelines," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6043–6056, Nov 2015.

[168] 3GPP, "NR; Study on integrated access and backhaul; Release 15," TR 38.874, 2018.

[169] M. Polese, M. Giordani, A. Roy, D. Castor, and M. Zorzi, "Distributed path selection strategies for integrated access and backhaul at mmWaves," in *IEEE Global Communications Conference (GLOBECOM)*, 2018.

[170] M. Polese, M. Giordani, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "End-to-end simulation of integrated access and backhaul at mmWaves," in *IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2018.

[171] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated Access and Backhaul in 5G mmWave Networks: Potentials and Challenges," *arXiv preprint arXiv:1906.01099*, 2019.

[172] V. Gambiroza, B. Sadeghi, and E. W. Knightly, "End-to-end performance and fairness in multihop wireless backhaul networks," in *10th Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2004.

[173] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Relay radio transmission and reception," TR 36.826 (Rel. 11), 2012.

[174] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. E. Fuja, "Bandwidth-and power-efficient routing in linear wireless networks," *IEEE Trans. on Inf. Theory*, vol. 52, no. 6, pp. 2624–2633, June 2006.

[175] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar, and A. Yener, "Rethinking information theory for mobile ad hoc networks," *IEEE Commun. Mag.*, vol. 46, no. 12, pp. 94–101, December 2008.

[176] S. Singh, M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Tractable Model for Rate in Self-Backhauled Millimeter Wave Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2196–2211, May 2015.

[177] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: challenges and research advances," *IEEE Network*, vol. 28, no. 6, pp. 6–11, Nov 2014.

[178] A. Mesodiakaki, A. Kassler, E. Zola, M. Ferndahl, and T. Cai, "Energy efficient line-of-sight millimeter wave small cell backhaul: 60, 70, 80 or 140 GHz?" in *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2016.

[179] C. Saha, M. Afshang, and H. S. Dhillon, "Integrated mmWave Access and Backhaul in 5G: Bandwidth Partitioning and Downlink Analysis," in *IEEE International Conference on Communications (ICC)*, 2018.

[180] A. Ometov, D. Moltchanov, M. Komarov, S. V. Volvenko, and Y. Koucheryavy, "Packet Level Performance Assessment of mmWave Backhauling Technology for 3GPP NR Systems," *IEEE Access*, vol. 7, pp. 9860–9871, 2019.

[181] T. Tian, Y. Dou, G. Ren, L. Gu, J. Chen, Y. Cui, T. Takada, M. Iwabuchi, J. Tsuboi, and Y. Kishiyama, "Field Trial on Millimeter Wave Integrated Access and Backhaul," in *IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, April 2019.

[182] Q. Hu, Y. Liu, Y. Yan, and D. M. Blough, "End-to-end Simulation of mmWave Out-of-band Backhaul Networks in Ns-3," in *Proceedings of the 2019 Workshop on Next-Generation Wireless with Ns-3 (WNGW)*, 2019.

[183] Y. Liu, A. Tang, and X. Wang, "Joint incentive and resource allocation design for user provided network under 5g integrated access and backhaul networks," *IEEE Transactions on Network Science and Engineering*, Apr 2019.

[184] A. Fouda, A. S. Ibrahim, I. Guvenc, and M. Ghosh, "Interference Management in UAV-assisted Integrated Access and Backhaul Networks," *arXiv preprint arXiv:1907.02585*, 2019.

[185] 3GPP, "Study on Integrated Access and Backhaul," TR 38.874 V0.2.1 (Rel. 15), 2018.

[186] ——, "Way Forward - IAB Architecture for L2/3 relaying," Qualcomm Inc, KDDI, AT&T, Nokia, Nokia Shangai Bell, Huawi, Ericsson, Intel, LG Electronics, CMCC, Samsung - Tdoc R3-181502, 2018.

[187] 3GPP, "NG-RAN; Architecture description," TS 38.401 (Rel. 15), 2018.

[188] T. Stockhammer, "Dynamic Adaptive Streaming over HTTP: Standards and Design Principles," in *Proceedings of the Second Annual ACM Conference on Multimedia Systems (MMSys)*, 2011.

[189] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, Oct 2014.

[190] G. Araniti, C. Campolo, M. Condoluci, A. Iera, and A. Molinaro, "LTE for vehicular networking: a survey," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 148–157, May 2013.

[191] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-Wave Vehicular Communication to Support Massive Automotive Sensing," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 160–167, Dec 2016.

[192] Automotive Edge Computing Consortium, "General principle and vision," *White Paper*, 2018.

[193] M. Giordani, A. Zanella, and M. Zorzi, "Millimeter wave communication in vehicular networks: Challenges and opportunities," in *6th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 2017.

[194] J. Bu, G. Tan, N. Ding, M. Liu, and C. Son, "Implementation and evaluation of wave 1609.4/802.11 p in ns-3," in *Proceedings of the 2014 Workshop on ns-3*, 2014.

[195] M. Boban, X. Gong, and W. Xu, "Modeling the Evolution of Line-of-Sight Blockage for V2V Channels," in *IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sept 2016.

[196] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-Wave Technology for Automotive Radar Sensors in the 77 GHz Frequency Band," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 845–860, March 2012.

[197] A. Kato, K. Sato, and M. Fujise, "ITS wireless transmission technology. Technologies of millimeter-wave inter-vehicle communications: Propagation characteristics," *Journal of the Communications Research Laboratory*, vol. 48, pp. 99–110, March 2001.

[198] R. C. Daniels, E. R. Yeh, and R. W. Heath, "Forward Collision Vehicular Radar With IEEE 802.11: Feasibility Demonstration Through Measurements," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1404–1416, Feb 2018.

[199] P. Kumari, J. Choi, N. González-Prelcic, and R. W. Heath, "IEEE 802.11ad-Based Radar: An Approach to Joint Vehicular Communication-Radar System," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3012–3027, April 2018.

[200] E. Grossi, M. Lops, L. Venturino, and A. Zappone, "Opportunistic Radar in IEEE 802.11ad Networks," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2441–2454, May 2018.

[201] F. Baccelli and B. Błaszczyszyn, "Stochastic Geometry and Wireless Networks: Volume I Theory," *Foundations and Trends® in Networking*, vol. 3, no. 3–4, pp. 249–449, 2010.

[202] A. Tassi, M. Egan, R. J. Piechocki, and A. Nix, "Modeling and Design of Millimeter-Wave Networks for Highway Vehicular Communication," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10 676–10 691, Dec 2017.

[203] K. Mase, J. Inoue, and M. Kizu, "Performance evaluation of a roadside-to-vehicle communication system using narrow antenna beam switching based on traffic flow model," in *IEEE GLOBECOM Workshops*, 2008.

[204] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Beam design for beam switching based millimeter wave vehicle-to-infrastructure communications," in *IEEE International Conference on Communications (ICC)*, May 2016.

[205] L. Liang, H. Peng, G. Y. Li, and X. Shen, "Vehicular Communications: A Physical Layer Perspective," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10 647–10 659, Dec 2017.

[206] S. A. Busari, "Millimetre-wave massive mimo for cellular vehicle-to-infrastructure communication," *IET Intelligent Transport Systems*, vol. 13, pp. 983–990(7), June 2019.

[207] M. Giordani, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "Performance Study of LTE and mmWave in Vehicle-to-Network Communications," *IEEE 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, 2018.

[208] S. Kato, M. Hiltunen, K. Joshi, and R. Schlichting, "Enabling vehicular safety applications over LTE networks," in *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, Dec 2013, pp. 747–752.

[209] Z. Hameed Mir and F. Filali, "LTE and IEEE 802.11p for vehicular networking: a performance evaluation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 1, p. 89, May 2014.

[210] K. C. Dey, A. Rayamajhi, M. Chowdhury, P. Bhavsar, and J. Martin, "Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network–Performance evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 168–184, Apr 2016.

[211] M. Giordani, A. Zanella, and M. Zorzi, "LTE and Millimeter Waves for V2I Communications: an End-to-End Performance Comparison," *IEEE 89th Vehicular Technology Conference (VTC-Spring), Workshops*, 2019.

[212] M. Giordani, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "Unleashing the Potential of Millimeter Waves for Future Vehicular Networks: a Performance Evaluation," *submitted to the IEEE Transactions on Vehicular Technology*, 2019.

[213] C. Perfecto, J. D. Ser, and M. Bennis, "Millimeter-Wave V2V Communications: Distributed Association and Beam Alignment," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2148–2162, Sept 2017.

[214] Q. Hu, C. Wu, X. Zhao, X. Chen, Y. Ji, and T. Yoshinaga, "Vehicular Multi-Access Edge Computing With Licensed Sub-6 GHz, IEEE 802.11p and mmWave," *IEEE Access*, vol. 6, pp. 1995–2004, 2018.

[215] Z. Sheng, A. Pressas, V. Ocheri, F. Ali, R. Rudd, and M. Nekovee, "Intelligent 5G Vehicular Networks: An Integration of DSRC and mmWave Communications," in *International Conference on Information and Communication Technology Convergence (ICTC)*, Oct 2018.

[216] M. Giordani, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "On the Feasibility of Integrating mmWave and IEEE 802.11p for V2V Communications," in *IEEE Connected and Automated Vehicles Symposium (CAVS)*, 2018.

[217] V. Petrov, J. Kokkoniemi, D. Moltchanov, J. Lehtomäki, M. Juntti, and Y. Koucheryavy, "The Impact of Interference From the Side Lanes on mmWave/THz Band V2V Communication Systems With Directional Antennas," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5028–5041, June 2018.

[218] B. Coll-Perales, M. Gruteser, and J. Gozalvez, "Evaluation of ieee 802.11ad for mmwave v2v communications," in *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, April 2018, pp. 290–295.

[219] M. G. Sánchez, M. P. Táboas, and E. L. Cid, "Millimeter wave radio channel characterization for 5G vehicle-to-vehicle communications," *Measurement*, vol. 95, pp. 223–229, 2017.

[220] M. Giordani, T. Shimizu, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "Path Loss Models for V2V mmWave Communication: Performance Evaluation and Open Challenges," in *IEEE Connected and Automated Vehicles Symposium (CAVS)*, 2019.

[221] J. Park, J. Lee, K. Kim, K. Lee, and M. Kim, "Vehicle Antenna Position Dependent Path Loss for Millimeter-Wave V2V Communication," in *11th Global Symposium on Millimeter Waves (GSMM)*, May 2018.

[222] C. S. Chen, F. Baccelli, and L. Roullet, "Joint optimization of radio resources in small and macro cell networks," in *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011.

[223] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *IEEE International Conference on Communications (ICC)*, 2012, pp. 2457–2461.

[224] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3484–3495, Aug 2012.

[225] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "User association and load balancing for cellular massive mimo," in *Information Theory and Applications Workshop (ITA)*, 2014.

[226] Y. Xu and S. Mao, "User association in massive MIMO HetNets," *IEEE Systems Journal*, vol. 11, no. 1, pp. 7–19, Sep 2017.

[227] D. Liu, Y. Chen, K. K. Chai, T. Zhang, and M. Elkashlan, "Opportunistic User Association for Multi-Service HetNets Using Nash Bargaining Solution," *IEEE Comm. Letters*, vol. 18, no. 3, pp. 463–466, March 2014.

[228] H. Liu, Z. Gao, X. Shao, and W. Zhou, "A centralized user association scheme for load balancing and UE energy efficiency in HetNets," in *2nd IEEE International Conference on Computer and Communications (ICCC)*, Oct 2016, pp. 2965–2969.

[229] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, Jun 2013.

[230] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "Load-aware modeling and analysis of heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1666–1677, April 2013.

[231] W. C. Cheung, T. Q. S. Quek, and M. Kountouris, "Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 561–574, April 2012.

[232] L. Liang, H. Ye, and G. Y. Li, "Toward intelligent vehicular networks: A machine learning framework," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 124–135, Feb 2019.

[233] Z. Li, C. Wang, and C. Jiang, "User Association for Load Balancing in Vehicular Networks: An Online Reinforcement Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 2217–2228, Aug 2017.

[234] D. Peron, M. Giordani, and M. Zorzi, "An Efficient Requirement-Aware Attachment Policy for Future Millimeter Wave Vehicular Networks," *30th IEEE Intelligent Vehicles Symposium (IV)*, 2019.

[235] European Commission, "Cooperative, connected and automated mobility (CCAM)," *White Paper*, 2016.

[236] S. Lasaulce and H. Tembine, *Game theory and learning for wireless networks: fundamentals and applications*. Academic Press, 2011.

[237] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1998.

[238] 3GPP, "Study on LTE support for Vehicle to Everything (V2X) services (Rel. 14)," *TR 22.885*, 2015.

[239] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) (Release 8)," *TR 36.300*, 2018.

[240] J. Gozalvez, M. Sepulcre, and R. Bauza, "Ieee 802.11p vehicle to infrastructure communications in urban environments," *IEEE Communications Magazine*, vol. 50, no. 5, pp. 176–183, May 2012.

[241] G. Bansal, J. B. Kenney, and C. E. Rohrs, "LIMERIC: A linear adaptive message rate algorithm for DSRC congestion control," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4182–4197, 2013.

[242] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Millimeter wave vehicular communications: A survey," *Foundations and Trends® in Networking*, vol. 10, no. 1, pp. 1–113, 2016.

[243] N. González-Prelcic, R. Méndez-Rial, and R. W. Heath, "Radar aided beam alignment in mmwave v2i communications supporting antenna diversity," in *Information Theory and Applications Workshop (ITA)*, 2016.

[244] N. Garcia, H. Wymeersch, E. G. Ström, and D. Slock, "Location-aided mm-wave channel estimation for vehicular communication," in *IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2016.

[245] L. X. Cai, H. Hwang, X. Shen, J. W. Mark, and L. Cai, "Optimizing geographic routing for millimeter-wave wireless networks with directional antenna," in *6th Conference on Broadband Communications, Networks, and Systems*. IEEE, 2009.

[246] J. Qiao, L. X. Cai, X. S. Shen, and J. W. Mark, "Enabling multi-hop concurrent transmissions in 60 ghz wireless personal area networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3824–3833, November 2011.

[247] M. Zhang, M. Mezzavilla, J. Zhu, S. Rangan, and S. Panwar, "The bufferbloat problem over intermittent multi-Gbps mmwave links," *arXiv preprint arXiv:1611.02117*, 2016.

[248] Y. Ni, J. He, L. Cai, and Y. Bo, "Delay analysis and message delivery strategy in hybrid V2I/V2V networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2016.

[249] M. Polese, R. Jana, and M. Zorzi, "TCP and MP-TCP in 5G mmWave Networks," *IEEE Internet Computing*, vol. 21, no. 5, pp. 12–19, 2017.

[250] G. Piro, N. Baldo, and M. Miozzo, "An LTE module for the ns-3 network simulator," in *Proceedings of the 4th International Conference on Simulation Tools and Techniques*, 2011, pp. 415–422.

[251] 3GPP, "Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects (Release 12) ," *TR 36.842*, 2014.

[252] ——, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 14)," TR 38.901, 2018.

[253] A. Ahmed, Z. Shafiq, H. Bedi, and A. Khakpour, "Suffering from buffering? detecting QoE impairments in live video streams," in *IEEE 25th International Conference on Network Protocols (ICNP)*, 2017.

[254] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 128–138, December 2012.

[255] 3GPP, "New SID on NR V2X (Release 16)," *RP-181480*, 2018.

[256] ——, "Study on NR beyond 52.6 GHz (Release 16)," *RP-181435*, 2018.

[257] C. Sommer, R. German, and F. Dressler, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis," *IEEE Transactions on Mobile Computing*, vol. 10, no. 1, pp. 3–15, 1 2011.

[258] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017.

[259] B. Schünemann, "V2X Simulation Runtime Infrastructure VSimRTI: An Assessment Tool to Design Smart Traffic Management Systems," *Comput. Netw.*, vol. 55, no. 14, pp. 3189–3198, 2011.

[260] L. Cheng, B. E. Henty, D. D. Stancil, F. Bai, and P. Mudalige, "Mobile Vehicle-to-Vehicle Narrow-Band Channel Measurement and Characterization of the 5.9 GHz Dedicated Short Range Communication (DSRC) Frequency Band," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1501–1516, Oct 2007.

[261] M. Boban, T. T. V. Vinhoza, M. Ferreira, J. Barros, and O. K. Tonguz, "Impact of Vehicles as Obstacles in Vehicular Ad Hoc Networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 1, pp. 15–28, January 2011.

[262] T. Driver, "Long-term prediction of GPS accuracy: Understanding the fundamentals," in *Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS)*, 2007.

[263] N. Alam, A. T. Balaei, and A. G. Dempster, "Relative positioning enhancement in VANETs: A tight integration approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 47–55, June 2013.

[264] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Feb 2014.

[265] M. Rebato, F. Boccardi, M. Mezzavilla, S. Rangan, and M. Zorzi, "Hybrid spectrum sharing in mmwave cellular networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 2, pp. 155–168, June 2017.

[266] A. Boukerche, H. A. Oliveira, E. F. Nakamura, and A. A. Loureiro, "Vehicular ad hoc networks: A new challenge for localization-based systems," *Computer communications*, vol. 31, no. 12, pp. 2838–2849, Jul. 2008.

[267] 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, Dec. 2013.

[268] L. N. Balico, A. A. F. Loureiro, E. F. Nakamura, R. S. Barreto, R. W. Pazzi, and H. A. B. F. Oliveira, "Localization Prediction in Vehicular Ad Hoc Networks," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2784–2803, May 2018.

[269] S. Yousefi, M. S. Mousavi, and M. Fathy, "Vehicular Ad Hoc Networks (VANETs): Challenges and Perspectives," in *6th International Conference on ITS Telecommunications*, Jun. 2006, pp. 761–766.

[270] C. Joo and A. Eryilmaz, "Wireless scheduling for information freshness and synchrony: Drift-based design and heavy-traffic analysis," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2556–2568, Dec. 2018.

[271] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *IEEE International Conference on Computer Communications (INFOCOM)*, Mar. 2012, pp. 2731–2735.

[272] R. A. Howard, "Information Value Theory," *IEEE Transactions on Systems Science and Cybernetics*, vol. 2, no. 1, pp. 22–26, Aug 1966.

[273] M. Giordani, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "Investigating Value of Information in Future Vehicular Communications," *IEEE 2nd Connected and Automated Vehicles Symposium (CAVS)*, 2019.

[274] ——, "A Framework to Assess Value of Information in Future Vehicular Networks," *1st ACM Mobihoc Workshop on Technologies, mOdels, and Protocols for Cooperative Connected Cars (TOP-Cars)*, 2019.

[275] L. Wischhof and H. Rohling, "Congestion control in vehicular ad hoc networks," in *IEEE International Conference on Vehicular Electronics and Safety, 2005.*, Oct. 2005, pp. 58–63.

[276] G. Caizzone, P. Giacomazzi, L. Musumeci, and G. Verticale, "A power control algorithm with high channel availability for vehicular ad hoc networks," in *IEEE International Conference on Communications (ICC)*, vol. 5. IEEE, May 2005, pp. 3171–3176.

[277] C.-L. Huang, Y. P. Fallah, and R. Sengupta, "Analysis of aggregated power level and rate-power control designs for status update messages in VANETs," in *IEEE 6th International Conference on Mobile Adhoc and Sensor Systems (MASS)*. IEEE, Oct. 2009, pp. 615–620.

[278] H. S. Ramos, A. Boukerche, R. W. Pazzi, A. C. Frery, and A. A. F. Loureiro, "Cooperative target tracking in vehicular sensor networks," *IEEE Wireless Communications*, vol. 19, no. 5, pp. 66–73, Oct. 2012.

[279] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *11th International Conference on Information Fusion*, Jun. 2008.

[280] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME - Journal of basic Engineering*, vol. 82, pp. 35–45, Jan. 1960.

[281] E. A. Wan and R. V. D. Merwe, "The unscented Kalman filter for nonlinear estimation," in *IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, Oct. 2000, pp. 153–158.

[282] P. Del Moral, "Non-linear filtering: interacting particle resolution," *Markov processes and related fields*, vol. 2, no. 4, pp. 555–581, Jan. 1996.

[283] A. U. Peker, O. Tosun, and T. Acarman, "Particle filter vehicle localization and map-matching using map topology," in *IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2011, pp. 248–253.

[284] A. T. Akabane, R. W. Pazzi, E. R. M. Madeira, and L. A. Villas, "Modeling and Prediction of Vehicle Routes Based on Hidden Markov Model," in *IEEE 86th Vehicular Technology Conference*, Sep. 2017.

[285] A. B. Poritz, "Hidden Markov models: a guided tour," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 1988, pp. 7–13.

[286] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[287] M. Roth, G. Hendeby, and F. Gustafsson, "EKF/UKF maneuvering target tracking using coordinated turn models with polar/Cartesian velocity," in *17th International Conference on Information Fusion (FUSION)*, Jul. 2014, pp. 1–8.

[288] C. M. Kang, S. J. Jeon, S. Lee, and C. C. Chung, "Parametric trajectory prediction of surrounding vehicles," in *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, Jun. 2017, pp. 26–31.

[289] T. King, H. Füßler, M. Transier, and W. Effelsberg, "Dead-reckoning for position-based forwarding on highways," in *International Workshop on Intelligent Transportation (WIT)*, Mar. 2006, pp. 199–204.

[290] D. Wang, J. Liao, Z. Xiao, X. Li, and V. Havyarimana, "Online-SVR for vehicular position prediction during GPS outages using low-cost INS," in *IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Aug. 2015, pp. 1945–1950.

[291] J. Park, D. Li, Y. L. Murphey, J. Kristinsson, R. McGee, M. Kuang, and T. Phillips, "Real time vehicle speed prediction using a Neural Network Traffic Model," in *The 2011 International Joint Conference on Neural Networks*, Jul. 2011, pp. 2991–2996.

[292] N. Deo, A. Rangesh, and M. M. Trivedi, "How Would Surround Vehicles Move? A Unified Framework for Maneuver Classification and Motion Prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 2, pp. 129–140, Jun. 2018.

[293] C. Hermes, C. Wohler, K. Schenk, and F. Kummert, "Long-term vehicle motion prediction," in *IEEE Intelligent Vehicles Symposium*, Jun. 2009.

[294] A. Houenou, P. Bonnifait, V. Cherfaoui, and W. Yao, "Vehicle trajectory prediction based on motion model and maneuver recognition," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov. 2013, pp. 4363–4369.

[295] "Tracking the position of neighboring vehicles using wireless communications," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 3, pp. 335 – 350, Jun. 2010.

[296] C. Huang, Y. P. Fallah, R. Sengupta, and H. Krishnan, "Adaptive intervehicle communication control for cooperative safety systems," *IEEE Network*, vol. 24, no. 1, pp. 6–13, Jan. 2010.

[297] Y. P. Fallah, C. Huang, R. Sengupta, and H. Krishnan, "Analysis of information dissemination in vehicular ad-hoc networks with application to cooperative vehicle safety systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 233–247, Jan. 2011.

[298] H. A. Omar, W. Zhuang, and L. Li, "VeMAC: A TDMA-Based MAC Protocol for Reliable Broadcast in VANETs," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1724–1736, Sep. 2013.

[299] N. Taherkhani and S. Pierre, "Centralized and Localized Data Congestion Control Strategy for Vehicular Ad Hoc Networks Using a Machine Learning Clustering Algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11, pp. 3275–3285, Nov. 2016.

[300] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of management information systems*, vol. 12, no. 4, pp. 5–33, Dec. 1996.

[301] S. Basagni, L. Boloni, P. Gjanci, C. Petrioli, C. A. Phillips, and D. Turgut, "Maximizing the value of sensed information in underwater wireless sensor networks via an autonomous underwater vehicle," in *IEEE INFOCOM*, 2014.

[302] N. Suri, G. Benincasa, R. Lenzi, M. Tortonesi, C. Stefanelli, and L. Sadler, "Exploring value-of-information-based approaches to support effective communications in tactical networks," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 39–45, Oct. 2015.

[303] C. Bisdikian, L. M. Kaplan, and M. B. Srivastava, "On the quality and value of information in sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 9, no. 4, p. 48, Jul. 2013.

[304] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-network aggregation techniques for wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 14, no. 2, May 2007.

[305] D. Heckerman, E. Horvitz, and B. Middleton, "An approximate nonmyopic computation for value of information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 3, pp. 292–298, Mar. 1993.

[306] E. Kamar and E. Horvitz, "Light at the end of the tunnel: A Monte Carlo approach to computing value of information," in *ACM International conference on Autonomous agents and multi-agent systems (AAMAS)*, 2013.

[307] L. Boloni, D. Turgut, S. Basagni, and C. Petrioli, "Scheduling data transmissions of underwater sensor nodes for maximizing value of information," in *IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 438–443.

[308] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Mar. 2013.

[309] M. Sepulcre, J. Gozalvez, O. Altintas, and H. Kremo, "Exploiting context information for estimating the performance of vehicular communications," in *Vehicular Networking Conference (VNC)*. IEEE, 2013, pp. 39–46.

[310] T. Higuchi, M. Giordani, A. Zanella, M. Zorzi, and O. Altintas, "Value-Anticipating V2V Communications for Cooperative Perception," *30th IEEE Intelligent Vehicles Symposium (IV)*, 2019.

[311] E. Mu and M. Pereyra-Rojas, "Understanding the Analytic Hierarchy Process," in *Practical Decision Making using Super Decisions*. Springer, 2018.

[312] P. Lytrivis, G. Thomaidis, M. Tsogas, and A. Amditis, "An Advanced Cooperative Path Prediction Algorithm for Safety Applications in Vehicular Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 669–679, Sep. 2011.

[313] T. Eiter, H. Füreder, F. Kasslatter, J. X. Parreira, and P. Schneider, "Towards a semantically enriched local dynamic map," *International Journal of Intelligent Transportation Systems Research*, vol. 17, no. 1, pp. 32–48, Jan 2019.

[314] S. El-Tawab, M. Abuelela, and Y. Gongjun, "Real-time weather notification system using intelligent vehicles and smart sensors," in *6th International Conference on Mobile Adhoc and Sensor Systems*, 2009, pp. 627–632.

[315] M. Tsogas, A. Polychronopoulos, and A. Amditis, "Unscented Kalman filter design for curvilinear motion models suitable for automotive safety applications," in *7th International Conference on Information Fusion*, vol. 2, Jul. 2005.

[316] S. J. Julier, "The scaled unscented transformation," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, vol. 6, May 2002, pp. 4555–4559 vol.6.

[317] D. Jiang and L. Delgrossi, "IEEE 802.11 p: Towards an international standard for wireless access in vehicular environments," in *Vehicular Technology Conference (VTC-Spring)*. IEEE, May 2008, pp. 2036–2040.

[318] F. Mason, M. Giordani, F. Chiariotti, A. Zanella, and M. Zorzi, "To Broadcast or Not To Broadcast? A Quality-Aware Tracking Strategy for Future Vehicular Networks," *submitted to the IEEE Transactions on Wireless Communications*, 2019.

[319] ——, "Quality-Aware Broadcasting Strategies for Position Estimation in VANETs," *European Wireless (EW)*, 2019.

[320] J. Benin, M. Nowatkowski, and H. Owen, "Vehicular network simulation propagation loss model parameter standardization in ns-3 and beyond," in *2012 Proceedings of IEEE Southeastcon*, Mar. 2012, pp. 1–5.

[321] B. Kim and K. Yi, "Probabilistic and Holistic Prediction of Vehicle States Using Sensor Fusion for Application to Integrated Vehicle Safety Systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2178–2190, Oct. 2014.

[322] G. Falco, M. Pini, and G. Marucco, "Loose and Tight GNSS/INS Integrations: Comparison of Performance Assessed in Real Urban Scenarios," *Sensors*, vol. 2017, p. 27, Jan. 2017.

[323] T. L. Saaty, *Decision making for leaders: the analytic hierarchy process for decisions in a complex world*. RWS publications, 1990.

[324] G. P. Stein, O. Mano, and A. Shashua, "Vision-based ACC with a single camera: bounds on range and range rate accuracy," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2003, pp. 120–125.

[325] New York State Department of Motor Vehicles, "NYS DMV – Driver's Manual – Chapter 8: Defensive Driving," 2011.

[326] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Towards 6G Networks: Use Cases and Technologies," *submitted to the IEEE Communications Magazine*,2019.

[327] X. Xu, Y. Pan, P. P. M. Y. Lwin, and X. Liang, "3D holographic display and its data transmission requirement," in *International Conference on Information Photonics and Optical Communications*, Oct 2011, pp. 1–4.

[328] Q. Zhang, J. Liu, and G. Zhao, "Towards 5G Enabled Tactile Robotic Telesurgery," *arXiv preprint arXiv:1803.03586*, 2018.

[329] I. Markit, "The Internet of Things: a movement, not a market," *White Paper*, 2017.

[330] J. Lee, B. Bagheri, and H.-A. Kao, "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18 – 23, 2015.

[331] M. Wollschlaeger, T. Sauter, and J. Jasperneite, "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0," *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 17–27, March 2017.

[332] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[333] J. M. Jornet and I. F. Akyildiz, "Channel Modeling and Capacity Analysis for Electromagnetic Wireless Nanonetworks in the Terahertz Band," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3211–3221, October 2011.

[334] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2047–2077, Fourthquarter 2015.

[335] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb 2004.

[336] S. Goyal, P. Liu, S. S. Panwar, R. A. Difazio, R. Yang, and E. Bala, "Full duplex cellular systems: will doubling interference prevent doubling capacity?" *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 121–127, May 2015.

[337] A. Ali, N. González-Prelcic, and R. W. Heath, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1038–1052, Feb 2018.

[338] J. Schloemann, H. S. Dhillon, and R. M. Buehrer, "Toward a Tractable Analysis of Localization Fundamentals in Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1768–1782, March 2016.

[339] 3GPP, "Study on New Radio (NR) to support non terrestrial networks," TR 38.811 (Release 15), 2018.

[340] M. Wang, Y. Cui, X. Wang, S. Xiao, and J. Jiang, "Machine Learning for Networking: Workflow, Advances and Opportunities," *IEEE Network*, vol. 32, no. 2, pp. 92–99, March 2018.

List of Publications

Publications on international journals

[J1] **M. Giordani**, M. Mezzavilla and M. Zorzi, "Initial Access in 5G mmWave Cellular Networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 40-47, November 2016.

[J2] M. Polese, **M. Giordani**, M. Mezzavilla, S. Rangan and M. Zorzi, "Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks", *in IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2069-2084, Sept. 2017.

[J3] **M. Giordani**, M. Mezzavilla, S. Rangan, and M. Zorzi, "An Efficient Uplink Multi- Connectivity Scheme for 5G mmWave Control Plane Applications", *in IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6806-6821, Oct. 2018.

[J4] **M. Giordani**, M. Rebato, A. Zanella, and M. Zorzi, "Coverage and Connectivity Analysis of mmWave Vehicular Networks: A Stochastic Geometry Approach," *Ad Hoc Networks*, vol. 80, pp. 158-171, Aug 2018.

[J5] **M. Giordani**, M. Polese, A. Roy, D. Castor, M. Zorzi, "A Tutorial on Beam Management for 3GPP NR at mmWave Frequencies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 173-196, 1st quarter 2019.

[J6] **M. Giordani**, M. Polese, A. Roy, D. Castor, M. Zorzi, "Standalone and Non-Standalone Beam Management for 3GPP NR at mmWaves," *IEEE Communications Magazine*, vol. 57, no. 4, pp. 123-129, Apr. 2019.

[J7] M. Polese, **M. Giordani**, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated Access and Backhaul in 5G mmWave Networks: Potentials and Challenges," *submitted to IEEE Communications Magazine*, May 2019.

[J8] **M. Giordani**, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Towards 6G Networks: Use Cases and Technologies" *submitted to IEEE Communications Magazine*, Jul. 2019.

[J9] **M. Giordani**, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "Unleashing the Potential of Millimeter Waves for Future Vehicular Networks: A Performance Evaluation", *submitted to IEEE Transactions on Vehicular Technology*, 2019.

[J10] F. Mason, **M. Giordani**, F. Chiariotti, A. Zanella, M. Zorzi, "To Broadcast or Not To Broadcast? A Quality-Aware Tracking Strategy for Future Vehicular Networks", *submitted to IEEE Transactions on Wireless Communications*, 2019.

[J11] T. Zugno, M. Drago, **M. Giordani**, M. Polese, M. Zorzi, "Towards Standardization of Millimeter Wave Vehicle-to-Vehicle Networks: Open Challenges and Performance Evaluation", *submitted to IEEE Vehicular Technology Magazine*, 2019.

Publications on conference and workshop proceedings

[C1] **M. Giordani**, A. Zanella and M. Zorzi, "Millimeter wave communication in vehicular networks: Challenges and opportunities," *6th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, May 2017.

[C2] **M. Giordani** and M. Zorzi, "Improved user tracking in 5G millimeter wave mobile networks via refinement operations," *16th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, Jun. 2017.

[C3] **M. Giordani**, M. Rebato, A. Zanella and M. Zorzi, "Poster: Connectivity analysis of millimeter wave vehicular networks," *IEEE Vehicular Networking Conference (VNC)*, Dec. 2017.

[C4] **M. Giordani**, M. Polese, A. Roy, D. Castor, and M. Zorzi, "Initial Access Frameworks for 3GPP NR at mmWave Frequencies", *17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, Jun. 2018.

[C5] **M. Giordani**, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "Performance Study of LTE and mmWave in Vehicle-to-Network Communications", *17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, Jun. 2018.

[C6] **M. Giordani**, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "On the Feasibility of Integrating mmWave and IEEE 802.11p for V2V Communications", *IEEE Connected and Automated Vehicles Symposium (CAVS)*, Aug. 2018.

[C7] M. Polese, **M. Giordani**, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "End- to-End Simulation of Integrated Access and Backahul at mmWaves," *IEEE International Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks (CAMAD)*, Sep. 2018.

[C8] M. Polese, **M. Giordani**, A. Roy, D. Castor, and M. Zorzi, "Distributed Path Selection Strategies for Integrated Access and Backhaul at mmWaves", *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018.

[C9] **M. Giordani**, A. Zanella, M. Zorzi, "LTE and Millimeter Waves for V2I Communications: an End-to-End Performance Comparison", *IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, Apr. 2019.

[C10] F. Mason, **M. Giordani**, F. Chiariotti, A. Zanella, M. Zorzi, "Quality-Aware Broadcasting Strategies for Position Estimation in VANETs", *European Wireless (EW)*, May 2019.

[C11] T. Higuchi, **M. Giordani**, A. Zanella, M. Zorzi, Onur Altintas, "Value-Anticipating V2V Communications for Cooperative Perception", *30th IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2019.

[C12] D. Peron, **M. Giordani**, M. Zorzi, "An Efficient Requirement-Aware Attachment Policy for Future Millimeter Wave Vehicular Networks", *30th IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2019.

[C13] **M. Giordani**, A. Zanella, T. Higuchi, O. Altintas, M. Zorzi, "A Framework to Assess Value of Information in Future Vehicular Networks ", *20th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, Jul. 2019.

[C14] **M. Giordani**, T. Shimizu, A. Zanella, T. Higuchi, O. Altintas, M. Zorzi, "Path Loss Models for V2V mmWave Communication: Performance Evaluation and Open Challenges", *2nd IEEE Connected and Automated Vehicles Symposium (CAVS)*, Sep. 2019.

[C15] **M. Giordani**, A. Zanella, T. Higuchi, O. Altintas, M. Zorzi, "Investigating Value of Information in Future Vehicular Communications", *2nd IEEE Connected and Automated Vehicles Symposium (CAVS)*, Sep. 2019.

[C16] **M. Giordani**, A. Zanella, T. Higuchi, O. Altintas, M. Zorzi, S. Nakamura, "Enabling Technologies for Future Transportation Systems: an End-to-End Performance Evaluation", *26th Intelligent Transport Systems World Congress (C-ITS)*, Oct. 2019.

[C17] A. Signori, C. Pielli, F. Chiariotti, **M. Giordani**, F. Campagnaro, N. Laurenti, M. Zorzi, "Jamming the Underwater: a Game-Theoretic Analysis of Energy-Depleting Jamming Attacks", *14th ACM International Conference on Underwater Networks and Systems (WUWNet)*, Oct. 2019.

Other publications (book chapters and patents)

[B1] M. Ericson, P. Spapis, M. Saily, K. Pedersen, Y. Qi, N. Barati, T. Svensson, M. Shariat,**M. Giordani**, M. Mezzavilla, M. Doll, H. Miao, and C. Zhou, "Mobility Management via UL-based Measurements" in *5G System Design – Architectural and Functional Considerations and Long Term Research*, Ed. Wiley, 2018.

[B2] **M. Giordani**, A. Zanella, T. Higuchi, O. Altintas, and M. Zorzi, "Emerging Trends in Vehicular Communication Networks," in *Emerging Wireless Communication & Network Technologies*, Ed. Springer, 2018.

[B3] M. Polese, **M. Giordani**, M. Zorzi, "3GPP NR: the cellular standard for 5G networks", *5G- ITALY White Book*, 2018.

[P1] T. Higuchi, O. Altintas, M. Zorzi, A. Zanella, **M. Giordani**, "Wireless Communication Assurance for Connected Vehicles in High Network Load Scenarios", *Patent Application Number 16/181,479*, Application Date: Nov. 06, 2018. Pending.